

MASTER'S THESIS

Governance mechanisms to realise openness in data platform ecosystems

van der Krieken, Floris

Award date:
2024

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

Take down policy

If you believe that this document breaches copyright please contact us at:

pure-support@ou.nl

providing details and we will investigate your claim.

Downloaded from <https://research.ou.nl/> on date: 16. May. 2025

Open Universiteit
www.ou.nl



Governance mechanisms to realise openness in data platform ecosystems

Degree programme: Open University of the Netherlands, Faculty Science
Master of Science Business Process Management & IT

Course: IM1803 BPMIT Graduation Assignment Preparation
IM9806 Business Process Management and IT Graduation Assignment

Student: Floris van der Krieken

Identification number:

Date: 04-02-2024

Thesis supervisor Samaneh Bagheri

Second reader Khoi Nguyen

Version number: 1.0

Status: Final version

Abstract

Data platforms, facilitating the trade of data, face challenges in balancing openness and control. This challenge arises for platform providers, given data owners' preference for control and consumers' need for safeguards. Governance mechanisms that reconcile these tensions are crucial for enabling openness in data platforms. This study aims to identify the Governance Mechanisms that enable openness in data platform ecosystems and reconcile the tension between openness and control. Drawing on a framework for the governance of digital platform ecosystems, it categorizes crucial mechanisms into Platform Boundary Resources, Platform Rules, and Ecosystem Identity. The framework has been validated and refined through 10 semi-structured interviews with subject matter experts to its applicability for Data Platform Ecosystems. The results show that while most governance mechanisms are validated, this does not hold true for every dimension and mechanism. In addition, new governance mechanisms and data-specific mechanisms are introduced by the participants. The research underscores the significance of customised approaches in overseeing openness within data platforms, considering their unique context and ecosystem-specific requirements.

Key terms

Governance Mechanisms, Openness, Data Platform Ecosystem, Platform Governance

Summary

In the dynamic landscape of data platform ecosystems, the tension between openness and control poses a unique challenge, and this thesis delves into the exploration of governance mechanisms that can reconcile this tension. Employing a comprehensive mixed-methods approach, the research combines a rigorous literature review with expert interviews, aiming to not only identify existing governance mechanisms but to also validate their practical relevance in the specific context of data platform ecosystems.

The initial phase of the research unfolds with a literature review process to identify existing governance mechanisms used to enable openness in digital platforms. A total of 908 articles were identified and subsequently screened based on titles, resulting in 91 potentially relevant articles. A refined selection of 20 articles was subjected to a detailed abstract review. Subsequently, 12 articles were found relevant for thematic analysis in Atlas.ti, based on their full text analysis. Building upon this literature, a conceptual framework was developed, categorising the governance mechanisms Application boundary resources, Development boundary resources, Social boundary resources, Intellectual property sharing, Decision rights, Gate keeping, Pricing, Revenue sharing, and Relational control into three pillars: Platform Boundary Resources, Platform Rules, and Ecosystem Identity.

Subsequently, this study aimed to assess what governance mechanisms, identified in the literature of platforms, are effective in enabling openness in data platform ecosystems. This process included conducting expert interviews to validate and refine the framework. Ten semi-structured interviews were conducted with subject matter experts in the data platform domain including a technical background, over five years of work experience, or key roles like CEO or CTO.

Based on the results of the research, the dimensions Platform Boundary Resources and Platform Rules and their underlying mechanisms also be apply to the governance framework of Data Platform Ecosystems. Participants consistently underscored the importance of Platform Boundary Resources and Platform Rules in enabling openness within the data platform domain. Platform boundary resources, as APIs, create connectivity to the data platform, enhancing its functionality and receiving support for this through tools and documentation. Platform rules, as Decision rights and Gate keeping practices, provide oversight by determining access privileges, specifying authorized actions, clarifying ownership of innovations, defining pricing structures, and outlining revenue-sharing allocations. However, Ecosystem Identity, a pillar within the conceptual framework, displays varying levels of recognition, emphasizing a potential context-specific nature, according to participant insights.

Additionally, new governance mechanisms, such as Privacy, Data Sharing Agreements, Roles & Responsibilities, Common Terminology, Data Lineage, and Data Catalog, emerged from the interviews, suggesting the need to include Data Governance as a foundational pillar in the framework. Furthermore, the study identified ten data-specific mechanisms that contribute to improved findability, accessibility, and usability of data within platforms. Notably, "Standardization" emerged as a critical mechanism highlighted by multiple respondents.

The findings of this research pave the way for future exploration. The non-uniform recognition of Relational Control prompts further investigation into its applicability in specific business contexts. Additionally, the newly suggested mechanisms and data-specific dimensions require in-depth exploration and validation, offering avenues for continuous refinement and evolution of governance strategies within data platform ecosystems.

Contents

Abstract	ii
Key terms.....	ii
Summary	iii
Contents	iv
1. Introduction.....	1
1.1. Background	1
1.2. Exploration of the topic	1
1.3. Problem statement	3
1.4. Research objective and questions	3
1.5. Motivation/relevance	3
1.6. Main lines of approach	4
2. Theoretical framework.....	5
2.1. Research approach.....	5
2.1.1. Identification of research.....	5
2.1.2. Selection of primary studies	6
2.1.3. Data extraction	6
2.1.4. Data synthesis.....	6
2.2. Implementation	7
2.3. Results and conclusions	7
2.4. Objective of the follow-up research	10
3. Methodology.....	11
3.1. Conceptual design: select the research method(s).....	11
3.2. Technical design: elaboration of the method	11
3.3. Data analysis	13
3.4. Reflection with regards to validity, reliability and ethical aspects.....	13
3.4.1. Validity	13
3.4.2. Reliability	14
3.4.3. Ethical aspects	14
4. Results	15
4.1. Research implementation.....	15
4.1.1. Feedback from pilot interview.....	15
4.2. Data analysis	16
4.3. Validated Framework.....	19
5. Discussion, conclusions and recommendations	20

5.1.	Discussion – reflection	20
5.2.	Conclusions	21
5.3.	Recommendations for practice.....	22
5.4.	Recommendations for further research	22
6.	References.....	23
	Appendix 1: Systematic Literature Review process.....	26
	Appendix 2: Final list of selected papers	63
	Appendix 3: Thematic analysis of Systematic Literature Review	64
	Appendix 4: Invitation letter expert interviews	85
	Appendix 5: Interview protocol.....	86
	Appendix 6: Overview of participants	87
	Appendix 7: Interview transcriptions	89
	Transcript Interview 1	89
	Transcript Interview 2	103
	Transcript Interview 3	123
	Transcript Interview 4	137
	Transcript Interview 5	154
	Transcript Interview 6	163
	Transcript Interview 7	179
	Transcript Interview 8	192
	Transcript Interview 9	205
	Transcript Interview 10	216
	Appendix 8: Data extraction and analysis	233
	Analysis interview 1.....	233
	Analysis interview 2.....	236
	Analysis Interview 3.....	239
	Analysis Interview 4.....	243
	Analysis interview 5.....	248
	Analysis interview 6.....	251
	Analysis interview 7.....	257
	Analysis interview 8.....	262
	Analysis interview 9.....	268
	Analysis interview 10.....	272

1. Introduction

1.1. Background

Platform businesses have become a recent focus of research in various management fields (Chen et al., 2022). In the information sciences, digital platforms have emerged as a popular research topic within the sociotechnical field, which intersects with areas such as information systems, media studies, and telecommunications. In the information sciences, digital platforms are a popular research topic, connecting with areas like information systems, media studies, and telecommunications. These studies recognise that the digital nature of platforms sets them apart from traditional literature on business and economics (De Reuver, Sørensen, et al., 2018, p. 127). A digital platform can be defined as technical elements (of software and hardware) and associated organisational processes and standards (De Reuver, Sørensen, et al., 2018, p. 127).

Digital platforms and data platforms have many similar characteristics, but also have significant differences. While any digital platform creates data as a side-product of platform usage (Van der Vlist & Helmond, 2021), the main offerings of data platforms revolve exclusively around data (de Reuver et al., 2022). Data platforms facilitate actors to trade data. There is a growing trend in the market for the emergence of data platforms that not only facilitate the sharing of data but also enable additional complementary services and analytics capabilities (Van den Broek & van Veenstra, 2015). The term platform ecosystem refers to the platform and all stakeholders interacting on the platform (Gawer & Cusumano, 2014). Data platform ecosystems can be considered a specific subset where the focal value proposition is enabled by data sharing (De Prieëlle et al., 2020). Multiple stakeholders, consisting of data owners, data consumers, and solution providers, converge on a data platform, facilitated by the platform owner (de Reuver et al., 2022).

Proponents of the data economy typically advocate for complete openness of data platforms to third parties (Niculescu et al., 2018). This would enable data exchange and access to third-party analytics or AI modules by data consumers and owners (Niculescu et al., 2018). However, data owners generally prefer to maintain control and sovereignty over their data, and data consumers require safeguards against low-quality or malicious third-party offerings (Jarke et al., 2019); Lauf et al. (2022). Therefore, determining the appropriate level of openness of a data platform poses a significant challenge for data platform providers (de Reuver et al., 2022).

1.2. Exploration of the topic

Data Platform Ecosystem

According to Gawer (2009), platforms are the fundamental components that function as a base for a cluster of stakeholders to design supplementary products, technologies, or services. Data platforms share several characteristics with digital platforms. Digital platforms have been highlighted for their multi-sidedness (Rochet & Tirole, 2003), which is also true for data platforms as they cater to data owners, data consumers, and third parties (de Reuver et al., 2022). Additionally, digital platforms are known to facilitate transactions and allow for recombinant innovation (Gawer, 2014), and data platforms serve the same purpose by enabling transactions between data owners and consumers while also facilitating third-party innovation (de Reuver et al., 2022). Technically, digital platforms are considered extensible systems (Tiwana et al., 2010), and data platforms can be extended with features like analytics. However, data platforms differ from other digital platforms in that their primary offerings revolve around data. On digital platforms, users generate data traces that owners can monetise or use to enhance the platform (Van der Vlist & Helmond, 2021). In contrast, data is not a by-product of platform usage for data platforms but instead forms the core value element that is

exchanged and built upon (de Reuver et al., 2022.) Data platforms are critical enablers of a data economy where businesses can freely exchange and monetise data (de Reuver et al., 2022).

The concept of platform ecosystems encompasses the platform and all stakeholders interacting on the platform (Gawer & Cusumano, 2014). Platform ecosystems consist of technological components that are shared by the platform owner and partners, including the platform core, which provides basic functionality for modular services (Tiwana et al., 2010), boundary resources that support the interface between ecosystem actors (Teece and Linden, 2017), and complements, which are individual components within the ecosystem (Hein et al., 2020). These ecosystems are networks in which platform owners enable third parties to create complementary innovations and generate new products and services (Cusumano et al., 2019; Cozzolino et al., 2021; Parida et al., 2019;). Examples of Data platform ecosystems are a data marketplace where data owners can sell their data to other businesses, or a data space, which provides an ecosystem for parties to share data (Beverungen et al., 2022; Koutroumpis et al., 2020).

Tension between platform Openness and control

Platform governance refers to who makes what decisions about a platform (ecosystem) (Tiwana et al., 2010). One of the factors on which platform governance is based is openness (Tan et al., 2019). Openness, as a governance-related concept, refers to reducing restrictions on who can participate on a platform and contributes to platform adoption and potential network effects (De Reuver, Nederstigt, et al., 2018).

In the digital platform research, openness is understood as the provision of access to platform resources by third-party entities through open source or specific interfaces (Ghazawneh & Henfridsson, 2013). However, the concept of digital platform openness cannot be directly applied to data platforms (de Reuver et al., 2022). In existing conceptualisations of platform openness, the object of openness typically entails platform modules, such as software and hardware modules that developers can build upon (Eisenmann et al., 2009). In the case of data platforms, the object of openness could be similar, such as analytics modules (Mucha & Seppala, 2020). However, the object of openness may also be specific to the context of data, such as datasets, data products, or data-driven insights (Bergman et al., 2022).

For the reason mentioned earlier, the relationship between openness and control is often characterised by tension. Some research suggests that the tension between openness and control can be addressed through proper governance (Tilson et al., 2010; Wareham et al., 2014). Two mechanisms that facilitate openness are giving up control or the selective granting of access to "boundary resources" (de Reuver et al., 2022). In addition, new mechanisms for data-driven insights have emerged that aim to reconcile this tension. Collaborative computing is an example of such a mechanism, allowing multiple parties to derive insights from data without divulging the underlying information. By utilising privacy-enhancing technologies, like Multi-Party Computation (MPC), collaborative computing enables openness while ensuring the protection of societal values such as privacy, security, and sovereignty (Agahari & de Reuver, 2022). Other similar mechanisms also exist, such as algorithms that move to data, rather than vice versa (Van Alstyne & Lenart, 2020). These innovative approaches have the potential to disrupt the traditional trade-off between openness and control, enabling data accessibility while maintaining control over how data consumers utilise the data (de Reuver et al., 2022). However, these mechanisms may also have implications that are not adequately considered in existing literature, such as the need for an additional layer of trust to verify computed insights (Bruun et al., 2020).

1.3. Problem statement

Data platforms are becoming increasingly popular in the market, enabling actors to trade data and the use of additional complementary services and analytics capabilities. However, determining the appropriate level of openness of a data platform poses a significant challenge for data platform providers, as data owners generally prefer to maintain control and sovereignty over their data, while data consumers require safeguards. The tension between openness and control can potentially be relieved by governance mechanisms that enable openness while safeguarding the needs of data owners and consumers. However, governance mechanisms that could facilitate openness of data platform ecosystems are not adequately considered in existing literature.

1.4. Research objective and questions

The objective of this research is to explore the tension between openness and control in data platform ecosystems and investigate what and how mechanisms can reconcile this tension.

The main research questions is:

What are the governance mechanisms that can enable openness and reconcile the tension between openness and control in data platform ecosystems?

The objective of the research can be addressed in the following sub-questions:

Theoretical research:

1. What are the existing governance mechanisms used to enable openness in platforms?

In order to address the initial sub-question, a literature review will be conducted. To this end, peer-reviewed articles that examine the mechanisms for achieving platform openness in general will be sought. Through this literature review, suitable mechanisms for opening data platforms will be determined based on the identified mechanisms for platform openness.

Empirical research:

2. What governance mechanisms, identified in the literature of platforms, are effective in enabling openness in data platforms specifically?
3. Why are the identified mechanisms suitable to enable openness in data platform ecosystems?
4. What are data-specific mechanisms used to manage data platform openness?

To answer the second and third sub-questions, interviews will be conducted with platform provider practitioners and subject-matter experts.

1.5. Motivation/relevance

Data platforms are critical components of the data economy, and their importance is increasing with the growing demand for data and insights. The tension between openness and control in data platforms is a significant challenge for platform providers. Mechanisms have emerged to address this issue. However, the implications of these mechanisms on existing tensions and potential new tensions in data platforms are not adequately explored in existing literature. Therefore, this research seeks to contribute to the academic discourse by exploring the potential of mechanisms to break the tension between openness and control in data platforms and identifying new tensions that may arise from their implementation. In addition, this research may also be of practical interest to platform providers

caught up in the tension between openness and control, where the research can provide insight into how data can be made accessible without losing control over data consumers.

1.6. Main lines of approach

This chapter commenced with an overview of the subject, outlining the research question and goal. Chapter 2 delved into the extraction of scientific literature concerning governance mechanisms facilitating openness in digital platforms. Subsequently, chapter 3 outlines the methodology employed in this study. Chapter 4 presents the results, while chapter 5 encompasses the discussion, conclusions, and recommendations derived from this research.

2. Theoretical framework

2.1. Research approach

To answer the first research question, "What are the existing governance mechanisms used to enable openness in platforms?", a Systematic Literature review will be conducted. The Systematic Literature Review (SLR) method is selected to investigate this research question as it provides a structured and comprehensive approach to gather, evaluate, and synthesise existing research literature on the governance mechanisms used to enable openness in platforms (Saunders et al., 2019). The SLR method is particularly appropriate as it enables the researcher to systematically identify and analyse relevant studies, reducing the risk of bias and enhancing the rigor of the research (Dyba et al., 2007). The protocol proposed by Dyba et al. (2007), will be followed to conduct the review, which consist of Identification of research, Selection of primary studies, Study quality assessment, Data extraction and Data synthesis.

2.1.1. Identification of research

Identification of Keywords and Search Terms

To construct the search string for the systematic literature review, the building block method will be used. This involves breaking down the research question into individual elements, identifying synonyms and related terms, and combining them to create a comprehensive search string (VU, 2022). The following key words were identified from the research question "What are the existing governance mechanisms used to enable openness in platforms?": "governance mechanisms", "openness", and "platforms". The Boolean operators (AND, OR) are used to connect these key words. The search string used to search for relevant literature is:

((“mechanisms”) OR (“governance mechanisms”)) AND ((“openness”) OR (“platform openness”))
AND ((“data platforms”) OR (“digital platforms”) OR (“platform ecosystem”))

Through the application of the search string across multiple databases, the systematic literature review will be able to identify relevant studies that explore the governance mechanisms used to enable openness in platforms.

Sources and Databases

To search for the defined search string multiple databases and sources will be used, including ACM Digital Library, AIS, IEEE, Scimedirect and Springerlink, which are made available by the Open University. These databases were chosen because they are used in alike studies and considered authoritative sources for academic literature in computer science and related fields (Dyba et al., 2007; Soto Setzke et al., 2019).

Inclusion and exclusion criteria

To ensure that the literature review is comprehensive and relevant, inclusion and exclusion criteria will be established. The search will be limited to articles published in English, and between 2013 and the present day, in order to ensure that the review is up-to-date and relevant. All articles included in the review will be peer-reviewed journal and conference papers with the focus on governance mechanisms that enable openness in platforms. The search will exclude editorials, prefaces, article summaries, interviews, news, reviews, correspondence, discussions, comments, reader’s letters and summaries of tutorials, workshops, panels and poster sessions.

2.1.2. Selection of primary studies

The study selection process is divided into four distinct phases. In the first phase, relevant studies are identified through defined search queries on selected databases based on specific abovementioned inclusion and exclusion criteria. The second and third phase involves an evaluation of the publications found during the initial search, where their titles and abstracts are analysed for suitability. Literature that is clearly irrelevant is excluded at this stage. In the fourth phase, publications that have been selected from the foregoing phases undergo a more comprehensive analysis. This involves reading the full text to ascertain whether the publication contains relevant information for the study being performed and data that can be extracted for future analysis. The purpose of this thorough evaluation is to ensure that the selected publications provide accurate and appropriate data for the research, thereby increasing the validity and reliability of the study.

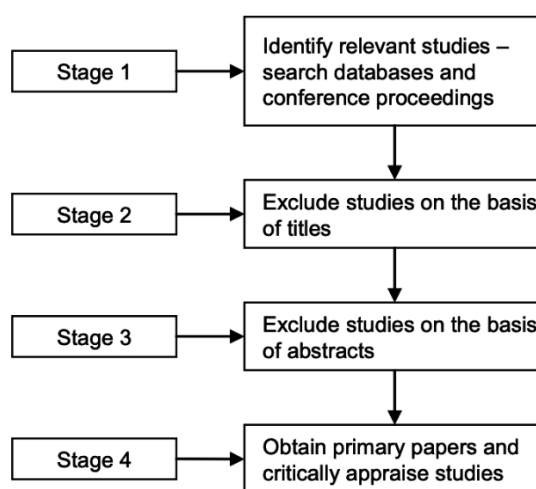


Figure 1: Stages of the study selection process (Dyba et al., 2007)

2.1.3. Data extraction

In this review, data extraction will be performed using the software Atlas.ti, which is a widely used tool for qualitative data analysis. The aim of this stage is to identify and extract data from each publication that is relevant to the research question. The data to be extracted will include the following:

- Publication details (e.g. author, title, year of publication, and publisher)
- Governance mechanisms identified in the study
- Benefits and limitations of the governance mechanisms
- Relationship between governance mechanisms and platform openness
- Key findings

2.1.4. Data synthesis

The final stage of a systematic literature review involves data synthesis, which aims to identify patterns and themes that emerge from the extracted data. This stage entails organising and synthesising the extracted data, which will be facilitated through the use of Atlas.ti software. The software will aid in categorising the data, identifying patterns, and extracting meaningful insights. Thematic analysis, a widely used method of data synthesis, will be employed to conduct the synthesis. This approach involves six recommended steps proposed by Braun and Clarke (2006): becoming

familiar with the data, generating initial codes, searching for themes, reviewing themes, defining themes, and writing up the results.

2.2. Implementation

The systematic literature review process involved searching academic databases including SpringerLink, IEEE, AIS, ScienceDirect, and ACM. A total of 908 articles were identified and subsequently screened based on titles, resulting in 91 potentially relevant articles. These articles underwent abstract review, reducing the number to 20. The full text of these 20 articles was analysed to determine the presence of Governance mechanisms for enabling openness. Ultimately, 12 relevant articles were chosen for thematic analysis. The complete list of articles can be found in Appendix 1, and the final list in Appendix 2.

For the data extraction phase, firstly the 12 relevant articles were documented into Atlas.ti. In order to become familiar with the data, all articles were thoroughly read. During this familiarising process, initial codes were applied by highlighting relevant pieces of information related to the research question. Subsequently, these codes were mapped and underlying relationships were determined.

2.3. Results and conclusions

In this section, the thematic analysis results of the literature review are presented. A detailed overview from the thematic analysis can be found in Appendix. 3. First, as the term platform openness is a multidimensional construct, its dimensions are clarified and defined. Second, governance mechanisms that enable platform openness are identified in the literature. In addition, their specific effect on the openness of platforms is described.

Platform openness dimensions

The literature review reveals that the concept of platform openness encompasses various dimensions. Van Alstyne and Lenart (2020) and West and O'mahony (2008) categorise openness, as either providing accessibility or transparency. Accessibility refers to the extent to which different roles, such as providers, third-party developers, or end users, are allowed to join and access the platform (Van Alstyne & Lenart, 2020; West & O'mahony, 2008). Transparency, on the other hand, relates to the comprehensibility of platform-related governance decisions, emphasising the understanding of what is happening and why (Van Alstyne & Lenart, 2020; West & O'mahony, 2008).

Benlian et al. (2015) take the perspective of complementors and developed an instrument for measuring complementors' perceived platform openness. They identify four dimensions: technical platform, distribution channel, accessibility, and transparency. Schlagwein et al. (2010) proposed a matrix-based framework for measuring the openness of platform resources along the dimensions of access and control. Access can take various forms, including exclusive access for specific groups or open access, while control can reside with either the platform owner, a specific group, or an external actor (Soto Setzke et al., 2019).

Previous research by Boudreau and Jeppesen (2015) and Eisenmann et al. (2009) defines actor openness based on two dimensions: access and authority. Access refers to who is allowed to access the platform, while authority pertains to the extent of actions actors are permitted to perform on the platform. An "open" platform, according to these studies, imposes no restrictions on participation and places reasonable and non-discriminatory restrictions on entry requirements, technical standards conformity, and licensing fees (Eisenmann et al., 2009).

Broekhuizen et al. (2021) makes a distinction between actor-based dimensions and non-actor-based dimensions. They argue that, in addition to making decisions about who is allowed on the platform and how much authority is granted to these users, platforms also need to decide on categories

("what") of assortments and brands and the channels ("where") of communication and distribution. The research discusses that platforms should define their openness along five dimensions: three actor-based dimensions (suppliers, customers, complementary service providers) and two non-actor-based dimensions (categories and channels). For actor-based dimensions, openness is a two-dimensional construct comprising both access and authority, while for non-actor-based dimensions, it relates to access alone (Broekhuizen et al., 2021).

While the literature reveals numerous studies that have explored the dimensions of platform openness, a comprehensive and consistent conceptualisation of these dimensions remains absent. Nevertheless, the literature review reveals that the terms "access" and "accessibility" are used interchangeably, denoting the same concept of who is allowed to join and access the platform. Transparency, authority, and control are separate concepts that relate to the set of rules governing the actions actors are allowed to take on the platform and the underlying justifications for these permissions. Based on these studies, this paper contends that the concept of platform openness encompasses the allowance of actors to the platform and the range of actions that various actors are permitted to carry out on the platform.

Identified governance mechanisms.

Concerning governance mechanisms, our objective was to identify governance mechanisms in the literature review articles that facilitate platform openness. From the literature can be conducted that openness is implemented through boundary resources (Boudreau, 2010; Schrieck et al., 2017). Boundary resources encompass the rules and tools that serve as the interface to govern the arm's-length relationship between the platform owner and different members of the platform ecosystem (Bonina & Eaton, 2020; Ghazawneh & Henfridsson, 2013). This encompasses both contributors on the supply side and developers on the demand side (Bonina & Eaton, 2020).

Staub et al. (2022) argues that governance mechanisms enabling openness can be categorised into three pillars: Platform boundary resources, Platform rules, and Ecosystem identity. We chose to build upon these three pillars identified by Staub et al. due to their ability to address the issue of diverse terminology. This choice is supported by their recent literature review, which offers a comprehensive overview of the governance mechanisms identified to achieve openness in digital platforms. In their research, they successfully reconcile the use of heterogeneous terminology, such as autonomy vs. control, openness vs. control, and generativity vs. stability. Consequently, we will focus on these three pillars in our research, rather than the often used two pillars of tools and rules highlighted in the other identified articles.

Platform boundary resources are resources provided by platform owners to complementors in order to support the development of complementary add-ons (Ghazawneh & Henfridsson, 2013; Staub et al., 2022). These boundary resources can be categorised as either technical, such as Application Programming Interfaces (APIs) and Software Development Kits (SDKs), or non-technical, such as documentation and support, as mentioned by Dal Bianco et al. (2014) and Tiwana (2013). In addition, Dal Bianco et al. (2014) differentiate between boundary resources between application boundary resources, development boundary resources and social boundary resources.

Platform rules, on the other hand, are established by the platform owner to define the scope of activities for each actor within the platform ecosystem, ensuring quality control for third-party apps and services developed for the platform (Bonina & Eaton, 2020; Staub et al., 2022). Ecosystem identity refers to the shared identity among the actors as a sense of belonging to the platform ecosystem (Staub et al., 2022).

The identified governance mechanisms are presented in table 2 below and are categorised into the three pillars Platform boundary resources, Platform rules and Ecosystem identity.

Categories	Sub- categories		Description	Effect on platform openness	References
Pillars	Governance mechanisms				
Platform boundary resources	Technical	<p>Application boundary resources</p> <ul style="list-style-type: none"> Application programming interfaces 	Application programming interfaces constitutes a means to allow third parties to access an organisation's capabilities or data.	Interfaces provide standardised access and connection to the platform and facilitate data exchange.	(Vesselkov et al., 2019) (Engert et al., 2022) (Soto Setzke et al., 2019) (Staub et al., 2022) (Farshchian & Thomassen, 2019) (Wulfert & Schütte, 2022) (Bonina & Eaton, 2020) (Dal Bianco et al., 2014)
	Technical	<p>Development boundary resources</p> <p>Include:</p> <ul style="list-style-type: none"> Software tools Software development kits Libraries Templates 	Technical developments boundary resources or tools are provided by the platform owner to complementors to help them develop add-ons.	Technical and non-technical programming resources support complementors in identifying possible ideas, lower the threshold to develop on the platform and influence complementors in innovating services.	(Engert et al., 2022) (Soto Setzke et al., 2019) (Staub et al., 2022) (Wulfert & Schütte, 2022) (Bonina & Eaton, 2020) (Dal Bianco et al., 2014)
	Non-technical	<p>Social boundary resources</p> <ul style="list-style-type: none"> Documentation Training Support Guidelines 	Non-technical boundary resources that enable transferring knowledge about and coordinating the thirdparty application development.		
Platform rules	<p>Intellectual property sharing</p> <ul style="list-style-type: none"> Licences 		Sharing of Intellectual property can happen via direct inter-firm agreements (licensing) or by making a technology accessible to the general public (open-source project)	Attract more complementors by expanding their intellectual property rights	(Staub et al., 2022) (Boudreau & Hagiu, 2009) (Bonina & Eaton, 2020)
	<p>Decision rights</p> <ul style="list-style-type: none"> Terms and conditions 		Decision rights can be defined as the division of authority and responsibilities between the platform owner and complementors. In general, it must become clear who decides about the strategic objectives of the platform or individual apps and how these objectives should be implemented	Decision rights ensure decision-making autonomy of complementors to increase their innovation output	(Staub et al., 2022) (Bonina & Eaton, 2020) (Engert et al., 2022)
	Gate keeping		Gatekeeping refers to the degree to which the	Gatekeeping means eases access for	(Staub et al., 2022)

	<ul style="list-style-type: none"> Access rights 	platform owner uses predefined objective acceptance criteria for judging what apps and app developers are allowed into a platform ecosystem	complementors for a more open platform and increases diversity in offered complements.	(Urovi et al., 2022) (Soto Setzke et al., 2019)
	Pricing	Pricing policies are used by the platform owner to create incentives for complementors to make personal investments to ensure the prosperity of their own offerings and in turn the whole ecosystem.	Pricing policies accelerate the adoption of the platform among complementors and users. By employing subsidies for one group of platform participants the critical mass of complementors and users will be achieved.	(Staub et al., 2022) (Vesselkov et al., 2019) (Soto Setzke et al., 2019)
	Revenue sharing	Revenue sharing represents the degree to which the platform extracts revenue that is co-created with the complementors.	Ensures complementors are motivated by sharing revenue.	(Staub et al., 2022) (Vesselkov et al., 2019)
Ecosystem identity	Relational control	Relational control refers to the degree to which the platform owner relies on norms and values that it shares with app developers to influence their behaviour and can be divided into self-control and clan control.	Relational control enhances complementor performance through fewer mistakes and less rework. A shared vision and similar ambitions bring complementors closer together and attracts new complementors.	(Staub et al., 2022)

Table 1: Conceptual framework of identified Governance Mechanisms

2.4. Objective of the follow-up research

The objective of the empirical part of the follow-up research is to validate the identified governance mechanisms identified in Chapter 2. Building upon the findings from the previous section, the research aims to investigate the practical relevance of the identified mechanisms in the context of data platforms and why they are relevant. In addition, the aim is to identify data-specific mechanisms and their underlying reasons.

3. Methodology

3.1. Conceptual design: select the research method(s)

In chapter 2, a conceptual framework of the identified governance mechanisms that enable platform openness was drawn. The objective of the empirical part of the research is to validate and investigate the practical relevance of the identified mechanisms in the specific context of data platforms and the reasoning for their relevance. Furthermore, the aim is to identify data platform-specific mechanisms and the underlying reasons. By addressing these objectives, the empirical research will address the following sub-questions:

Sub-question 2: What governance mechanisms, identified in the literature of platforms, are effective in enabling openness in data platforms specifically?

Sub-question 3: Why are the identified mechanisms suitable to enable openness in data platform ecosystems?

Sub-question 4: What are data-specific mechanisms used to manage data platform openness?

Several research methods have been examined to address these research objectives. A **survey** is usually associated with a deductive research approach and is often used for exploratory and descriptive research (Saunders et al., 2019). An advantage of the use of surveys is that they enable the collection of large amounts of data to validate our framework of governance mechanisms, increasing the generalisability of the findings. A disadvantage is its limitations in asking follow-up questions, what makes it difficult to develop an in-depth understanding.

Expert interviews are employed in exploratory and is useful when one wishes to understand something or wants to assess phenomena in a bright light (Saunders et al., 2019). Expert interviews are a valuable way to ask open questions to discover what is going and allow for detailed exploration of the subject matter, providing rich and comprehensive insights into the research topic. A disadvantage of expert interviews is that you rely on the quality of the contributions from those who participate.

A **case study** is commonly employed in explanatory, descriptive and exploratory research (Saunders et al., 2019). A Case study is defined as an in-depth inquiry into a topic or phenomenon within its real-life setting (Yin, 2018). According to Saunders et al. (2019), case studies are particularly effective in evaluating conceptual frameworks. A disadvantage of a case study could be limitation of the generalisability to other data platforms.

To validate and refine our governance mechanisms framework for data platforms, we opted for expert interviews as an appropriate research method due to the unavailability of a suitable case study. Expert interviews provide the advantage of engaging diverse experts across various organisations, specialising in different data types and sectors. This broader scope enhances the generalisability of the research.

3.2. Technical design: elaboration of the method

To conduct the expert interviews, semi-structured interviews will be organised with experts in the platform domain. Semi-structured interviews involve following a predetermined list of themes, in our case, the governance mechanisms that have been identified, along with relevant key questions pertaining to these mechanisms (Saunders et al., 2019). The choice was made to conduct semi-structured interviews to ensure a systematic and consistent examination of the identified governance

mechanisms in a structured format. Moreover, it provides an avenue to inquire further and gather comprehensive insights (Saunders et al., 2019).

Selection of participants

To ensure a thorough understanding of the platform domain and governance mechanisms, experts with relevant experience will be chosen based on their involvement in platform design, development, or expertise in openness. Selection criteria include a strong technical background, over five years of work experience, or key roles like CEO or CTO. Candidates are identified through Google searches for data platforms and their competitors, along with a review of organisation LinkedIn pages. Suitable candidates will be contacted via LinkedIn by private message, attached in appendix 4.

The target is to conduct 10 interviews with experts in the platform domain, which is expected to provide a sufficient range of perspectives and insights to address the research objectives effectively. The research aims to reach saturation point, which is the stage at which new information or insights from additional interviews no longer emerge (Saunders et al., 2019). Before conducting the full set of interviews, a pilot interview will be conducted with a knowledgeable individual who meets the criteria of an expert in the platform domain. The pilot interview aims to test the interview protocol, identify any potential issues or improvements, and ensure that the questions effectively address the research objectives (Saunders et al., 2019). Moreover, feedback from the pilot interviewee will be asked to improve the interview protocol when necessary. The duration of each interview will depend on the depth of the discussions and the insights shared but is estimated to last approximately 60 minutes. Qualitative research interviews are normally audio-recorded and then transcribed (Saunders et al., 2019). Permission for this will be requested from the interviewee at the start of the interview.

Interview protocol

The interview protocol, which can be found in Appendix 5, consists of three main parts: Introduction, Main Body, and Closing. The purpose of the introduction is to establish rapport with the interviewee and provide a brief overview of the research objectives. This part may include general questions to familiarise the interviewee with the interview process and to give the interviewee the opportunity to introduce him or herself. In addition, open questions are asked regarding platform openness and governance mechanisms that enable openness, before seeing the conceptual framework. Based on the answers of the interviewees new follow-up questions might arise.

The main body of the interview focuses on investigating the practical relevance of the identified governance mechanisms to platform openness and their reasoning, this includes open-ended questions. These questions aim to gather detailed insights, perspectives, and examples from the experts. In addition, practices of data-specific mechanisms are questioned and their reason for implementation regarding openness.

The closing part of the interview allows for questions, comments, or additional insights that the interviewee may want to share. Moreover, In the closing part the participants are asked to point out any missing elements that aren't discussed yet.

3.3. Data analysis

In accordance with the recommendations of Saunders et al. (2019), a complete transcription of the interviews conducted with the experts will be carried out, encompassing not only the spoken words but also the nuances and expressions. Furthermore, detailed notes will be taken to capture any non-verbal communication. The transcriptions will subsequently be shared with the interviewees to ensure factual accuracy (Saunders et al., 2019). Subsequently, the answers of the interviews are processed anonymously.

Thematic Analysis, as proposed by Braun and Clarke (2006), will be employed to analyse the translated transcriptions. The software Atlas.ti will be utilised for conducting the thematic analysis. The first objective of the empirical part of this research is to validate and assess the relevance of the governance mechanisms identified in Chapter 2. For this reason, the initial framework for coding will be based on the predefined codes and themes established in Chapter 2. However, new codes may also be introduced to accommodate the open-ended and follow-up questions in the interview protocol. The purpose of this analysis is to validate and potentially refine the previously identified overview of governance mechanisms. Each mechanism in the conceptual framework will be evaluated for its relevance based on the responses provided by the interviewees. This evaluation will determine whether the aspect receives a positive or negative validation, as outlined in Table 2.

Relevance	Validation	Clarification
Yes	Positive	The interviewee fully agrees
Yes, with constraints	Positive	The interviewee partially agrees
Possibly	Negative	The interviewee possibly agrees
No	Negative	The interviewee does not agree

Table 2: Codes of validation

Regarding the second objective, which is to identify data-specific mechanisms used to manage data platform openness and evaluate their effectiveness, new codes may be introduced, and themes will be identified. These newly identified mechanisms will be incorporated into the framework, along with their effect on the openness of data platforms.

3.4. Reflection with regards to validity, reliability and ethical aspects

3.4.1. Validity

Validity refers to the extent to which data collection method or methods accurately measure what they were intend to measure (Saunders et al., 2019). In the context of the research methodology described, several aspects contribute to ensuring validity.

Internal validity refers to the extent to which findings can be attributed to interventions rather than any flaws in the research design. To enhance internal validity, the interview was established based on the outcomes of the Systematic Literature Review and thematic analysis in chapter two. Additionally, a pilot interview will be conducted to identify and rectify any issues in the protocol before data collection. These measures increase the validity of subsequent interviews by minimising potential biases and errors. A threat to the internal validity is the absence of a quality assessment on the selected articles from the SLR due to time constraints.

External validity refers to how well the results of a study can be applied to other relevant contexts. The selection of interviewees with expertise in the platform domain enhances external validity. However, it's essential to recognise that the findings may be limited to the specific context of data platforms and participants' knowledge. The study focused on providing insights within this context,

but the absence of a concrete case study is a potential threat to external validity. Conducting expert interviews addressed the research questions but may limit generalisability.

3.4.2. Reliability

Reliability is defined as the extent to which the data collection technique or techniques will yield consistent findings, similar observations would be made or conclusions reached by other researchers or there is transparency in how sense was made from the raw data (Saunders et al., 2019). The use of a standardised interview protocol ensures consistency and minimises variations in data collection, enhancing reliability. Moreover, by transcribing the interviews in their entirety, including non-verbal communication, and subsequently sharing these transcriptions with the interviewees for verification, promotes accuracy and reliability. Additionally, the adoption of thematic analysis introduces a rigorous and structured framework for the data analysis process, while the use of Atlas.ti further reinforces the reliability by providing a systematic and transparent platform for coding and analysing the data. A potential threat to reliability is the subjectivity involved in the interpretation and coding of the data during the thematic analysis. To mitigate this, multiple researchers can independently code a subset of the data and establish inter-coder reliability to ensure consistency in the coding process. However, this will not be done in this research due to the lack of time and resources.

3.4.3. Ethical aspects

Research ethics are standards of the researcher's behaviour in relation to the rights of those who become the subject of the research project, or who are affected by it (Saunders et al., 2019). To uphold these ethical standards, we will obtain explicit permission from the interviewees at the beginning of each interview for audio-recording and transcription. This demonstrates respect for the participants' autonomy and ensures their informed consent regarding data collection and analysis. To protect the participants' privacy, all transcriptions and subsequent analysis will be treated with confidentiality and anonymity. We will not collect any personally identifiable information from the participants, as it is unnecessary for our research objectives. Furthermore, any data collected will solely be used for the purposes explicitly discussed with and consented to by the participants. We will not employ the collected data for secondary purposes and only utilise the data for the purposes discussed and consented to by the participants.

4. Results

4.1. Research implementation

To validate and refine the theoretical framework, ten semi-structured interviews were conducted with selected participants. These experts were selected based on specific criteria outlined in Chapter 2, reflecting their knowledge and experience with respect to the topic. The participants fulfilled the criteria outlined in the methodology section, except for one person who had four years of experience instead of the specified five years. Additionally, one interview was conducted in Dutch due to the participant's preference.

Table 3 provides an overview of the interviewees, detailing their respective roles within their organisations, their experience with data platform ecosystems and years of experience with data platforms. The complete overview of the respondents and their determining answers to the introduction questions can be found in Appendix 6.

#	Role within organisation	Experience with data platforms	Years of experience
Int_1	CEO of a financial data platform	Engineering	11 years
Int_2	Principal Consultant at a consultancy firm	Research	10 years
Int_3	Former CEO of a Data marketplace	General- and Product management	30+ years
Int_4	Assistant Professor at a Dutch university	Research	5 years
Int_5	Chief Science Officer at a Dutch ministry	Research	30+ years
Int_6	CEO/CTO of multiple tech companies	Architecture and Engineering	14 years
Int_7	CEO of a data engineering company	Architecture and Engineering	10 years
Int_8	Partner at a consultancy firm	Engineering	20+ years
Int_9	Data platform lead at a maritime contracting company	Engineering	4 years
Int_10	Head of tech data platforms of a supermarket chain	Architecture and Engineering	14 years

Table 3: Overview of participants

To maintain the specified level of anonymity, company names have been substituted, and interviews have been assigned unique codes. These interviews will be referenced by their corresponding codes as outlined in table 3 throughout this thesis. Both the codes and company names have been replaced consistently within the interview transcriptions (see appendix 7) and the interview analysis (refer to appendix 8).

4.1.1. Feedback from pilot interview

Two pilot interviews were conducted with two experts in data management. The pilot interviews revealed that it was necessary to share additional information about requested Governance Mechanisms when these were not clear to the interviewee. Consequently, a list has been made, encompassing the identified governance mechanisms along with their descriptions extracted from the initial framework. This list served as a reference point to address instances where questions may not be explicitly clear to the interviewees but were not shared beforehand to prevent bias. Furthermore, a lack of consistency was observed between questions from 13 to 15 compared to the rest of the questions, which included additional context acknowledging the constructs as mechanisms to facilitate openness. Consequently, the decision was made to omit this supplementary information in those questions.

4.2. Data analysis

This section provides the outcomes derived from the conducted interviews. Using Atlas.ti and Excel, every response underwent systematic coding to facilitate data extraction and analysis. Full transcripts of the interviews are available in Appendix 7, while the data extraction and analysis per interview is provided in Appendix 8.

Interview results (introduction)

During the introduction phase, the interviewees were asked to name observed Governance mechanisms being used in data platform ecosystems to facilitate openness. To ensure interviewees gave unbiased answers, this question was asked before discussing the validation question of the theoretical framework. The answers to these questions have been clustered and can be found in Appendix 9.

Multiple participants had observed licensing, access mechanisms, and APIs as mechanisms that enable openness in data platform ecosystems. These governance mechanisms align with the previously identified governance mechanisms in the conceptual framework. Additionally, interviewees introduced new mechanisms. Participants indicated that they observed “Data Sharing Agreements”, “Standardisation”, “Privacy”, “Data Products”, “Education”, “Encryption”, “Anonymisation”, “Data Governance”, “Data Lineage”, “Roles & Responsibilities” and “Common Terminology” as mechanisms that promote openness in data platform ecosystems.

Interview results (main body)

During the main body, interviewees were asked about the role of the identified governance mechanisms in achieving openness in data platforms. This approach allowed for the validation of whether the previously identified governance mechanisms could be deemed relevant. A summary of the validation questions can be found in Table 4. Two participants revealed a lack of definitive responses to questions regarding Revenue Sharing and three participants to questions regarding Relational Control.

Mechanism	Total				Aggregated		Score
	Yes	Yes, with constraints	Possibly	No	Positive	Negative	
Application programming interfaces	7	3	0	0	10	0	100%
Technical development boundary resources	7	2	1	0	9	1	90%
Non-technical social boundary resources	8	2	0	0	10	0	100%
Intellectual property sharing	7	2	1	0	9	1	90%
Decision rights	6	3	1	0	9	1	90%
Gatekeeping	7	3	0	0	10	0	100%
Pricing	3	5	1	1	8	2	80%
Revenue sharing	4	2	2	0	6	2	75%
Relational control	3	1	1	2	4	3	57%

Table 4: Summary of validation questions

Application programming interfaces

All the interviewees unanimously considered APIs as a governance mechanism that enables openness in data platforms. As one participant emphasised, 'APIs are the best way to distribute your data, because you can change the data underneath it without breaking the connection to your third parties or the people who you are distributing the data' (Int_8). Additionally, interviewees highlighted the exceptional ease of use associated with APIs, with one noting, 'The API is the easiest possible way. I

mean when you come to our website, you get an account, you get API keys, and within a minute and a half, you can be pulling data down from our databases' (Int_1)."

Technical development boundary resources

Regarding Technical Development Boundary Resources, nine out of ten participants viewed them as crucial for enabling openness. Emphasising usability, one interviewee mentioned, 'API is the first step. You have to have the delivery mechanism, but the education and usability around it are crucial. If that's lacking, it's not really open. That's where SDKs come in' (Int_1). Another participant highlighted the importance of ensuring consumer privacy in software development kits, saying, 'One critical aspect in SDKs is ensuring consumer privacy' (Int_4). However, another participant expressed uncertainty, stating, 'It helps intermediaries, but it's not something we've seen. Openness of data relies on proper governance and mechanisms to find and access data' (Int_3).

Non-technical social boundary resources

Participants unanimously acknowledged documentation and training as crucial for enabling openness in data platforms. One interviewee emphasised the need for clear documentation and training to support user understanding (Int_8), while another participant highlighted the necessity of these resources for users to know what they can access (Int_3). However, challenges such as potential confusion from poorly executed documentation and inadequate communication were recognised as potentially hindering openness (Int_6, Int_7).

Intellectual property sharing

Intellectual property sharing was recognised by 90% of the participants as a crucial governance mechanism for openness in data platforms. One participant (Int_2) highlighted its role in addressing companies' hesitancy to engage in open platforms, citing the Creative Commons licensing scheme as an example. The participant noted variations in openness levels between commercial and open licenses, emphasising that commercial licenses often impose restrictive access, while open licenses allow more freedom. However, concerns were raised about the quality of open data and the limited applications built on it (Int_2, Int_4).

Decision rights

90% of the participants recognised the relevance of Decision Rights. This was notably acknowledged by Int_1 and Int_7. They emphasise that while most companies generally accept the standard terms and conditions, they do engage in signing additional data sharing agreements or contracts. These documents provide more detailed specifications regarding their specific usage of the data, contingent upon the type of data they acquire or utilise. As mentioned by Int_4, 'In the agreement, it should be clearly described who is going to make the decisions, who is going to analyse data, who is going to use the data, and where the data is coming from.'

Gate keeping

Interviewees unanimously recognised gatekeeping as a crucial governance mechanism for enabling openness in data platforms. Int_8 highlighted the importance of understanding the type of users connecting to the platform, while Int_9 emphasised the necessity of access rights. However, Int_4 noted that gatekeeping poses a major challenge for organisations, mainly due to security reasons. Int_6 acknowledged the complexities of implementing gatekeeping but mentioned effective ways to handle it. Int_2 discussed the possibilities of EOSC Authentication and Authorisation Infrastructure (AAI) services in gatekeeping practices, highlighting multi-sided platforms.

Pricing

80% of the participants recognised the relevance of pricing, acknowledging its dual impact. Int_2 highlighted how pricing increased awareness of data value but noted potential benefits in certain

cases of free accessibility. Int_4 emphasised the need to treat data as a monetisable resource, separating concerns about consent from the monetisation process. Int_4 stressed that pricing should be part of governance mechanisms to enable transparent data trading and price discovery. In contrast, Int_1 expressed concerns about pricing inhibiting openness, citing potential barriers like exchange fees and technology delivery fees.

Revenue sharing

Revenue Sharing is recognised by 60% of the interviewees as relevant within data platforms. Int_2 emphasises the pivotal role of revenue streams in fostering openness, stating, 'In these data platforms, understanding that you often cannot do this alone is crucial. Hiring a third party, paying per service or per data amount to have a service built on your data is interesting. You provide the data for free, they build a service, and you receive a percentage.' Int_8 echoes this view, underlining the integration of financial models with operational plans, saying, 'Typically you work with plans, and each plan can be attached to some financial model.'

Relational control

In the study, 40% of the interviewees recognised Relational Control as an enabling governance mechanism for openness. Int_8 expressed caution about its applicability in their business context, stating, 'I don't think it can be applied for things used in our business. We deliver platforms for specific customers, and they decide on openness for their system.' In contrast, Int_2 highlighted evolving trends in openness within the German industry, especially in manufacturing and machinery. Int_2 noted, 'Now it's fashionable to be open and share data, helping the German industry catch up on digitisation. Companies like Bosch and Siemens fund start-ups to explore their data, fostering openness and brand enhancement.'

Data-specific mechanisms

After discussing the governance mechanisms of the initial framework, the participants were asked what data-specific mechanisms are used to manage data platform openness. The findings, including the discussed data-specific mechanisms, are detailed in Table 5.

Mechanism	Description	Contribution to openness	Observed in
Standardisation	The process of converting data into a standard format	It makes the data structured so it's accessible.	Int_1, Int_2, Int_3, Int_7
Data Quality	The planning, implementation, and control of activities that apply quality management techniques to data.	It increases the use of the data.	Int_1
Data Principles	Guidelines or rules that govern the management, usage, and handling of data	It makes the data more easily accessed, understood, exchanged, and reused.	Int_2
Metadata management	Planning, Implementation, and control activities to enable access to high quality, integrated metadata.	It describes and gives meaning to the data, so its findable and usable.	Int_2
Sticky policies	Represent one approach to improve owners' control over their data.	It outlines rules governing automatic permission or restriction of specific actions on the data.	Int_2
Data Lineage	Identifies the sources of data and the transformations through which it has passed	It allows you to trace the origin of the data.	Int_4
Data Classification	Systematic arrangement in groups or categories according to established criteria	It makes data findable by the use of tags.	Int_4
Data Utility	Measures the usefulness of the data for statistical analyses	It provides users insights into data purposes and its overall usefulness.	Int_5
Event Buses	Enables to route events from multiple sources to multiple destinations, or targets	It allows subscription on the bus and collect data based on system events.	Int_6
Anonymisation	The process of rendering personal data anonymous.	It enables the sharing of data that should not be disclosed publicly.	Int_7

Table 5: Data-specific mechanisms

Interview results closing

At the end of the interview, each interviewee was asked whether they would like to mention any additional covered mechanism and if any element was missing. The mentioned aspects included Data Catalog (Int_1), Privacy (Int_6 & 8), Rules and Regulations (Int_6), and Availability (Int_10).

4.3. Validated Framework

Following the insights gathered from the interviews, the refinement of the proposed framework was completed, detailed in Table 6. Validated mechanisms are coloured green and not sufficiently validated mechanisms red. Newly suggested mechanisms by respondents, not initially included in the theoretical framework, are highlighted in orange. The noticed data-specific mechanisms are coloured yellow.

Categories	Sub- categories	Positively validated/suggested in interview
Pillars	Governance mechanisms	
Platform boundary resources	Technical Application boundary resources	Int_1, Int_2, Int_3, Int_4, Int_5, Int_6, Int_7, Int_8, Int_9, Int_10
	Technical Development boundary resources	Int_1, Int_2, Int_4, Int_5, Int_6, Int_7, Int_8, Int_9, Int_10
	Non-technical Social boundary resources	Int_1, Int_2, Int_3, Int_4, Int_5, Int_6, Int_7, Int_8, Int_9, Int_10
Platform rules	Intellectual property sharing	Int_1, Int_2, Int_3, Int_4, Int_6, Int_7, Int_8, Int_9, Int_10
	Decision rights	Int_1, Int_2, Int_3, Int_4, Int_5, Int_6, Int_7, Int_8, Int_10
	Gate keeping	Int_1, Int_2, Int_3, Int_4, Int_5, Int_6, Int_7, Int_8, Int_9, Int_10
	Pricing	Int_2, Int_3, Int_4, Int_6, Int_7, Int_8, Int_9, Int_10
	Revenue sharing	Int_1, Int_2, Int_4, Int_5, Int_6, Int_7, Int_8, Int_9, Int_10
Ecosystem identity	Relational control	Int_1, Int_2, Int_4, Int_10
Newly suggested mechanisms	Privacy	Int_5, Int_8
	Roles & Responsibilities	Int_5, Int_6, Int_7, Int_8
	Common Terminology	Int_1, Int_2, Int_6
	Data Lineage	Int_4, Int_7
	Data Sharing Agreements	Int_2, Int_3, Int_4, Int_7, Int_9
Data-specific mechanisms	Standardisation	Int_1, Int_2, Int_3, Int_7
	Data Quality	Int_1
	Data principles	Int_2
	Metadata management	Int_2
	Sticky policies	Int_2
	Data Classification	Int_4
	Utility	Int_4
	Anonymisation	Int_5, Int_7
Event buses	Int_6	

Table 6: final framework of governance mechanisms

5. Discussion, conclusions and recommendations

5.1. Discussion – reflection

Reflection on the literature research and framework development

For the literature review, a specific search query was formed with the key words "governance mechanisms", "openness", and "platforms" with Boolean operators (AND, OR) employed to establish connections between these terms. The research involved an extensive literature review process, including searches on academic databases such as SpringerLink, IEEE, AIS, ScienceDirect, and ACM. The query resulted in 908 articles from which twelve articles were found relevant for in-depth thematic analysis. Due to time limitations, alternative search queries were not employed, limiting the inclusion of articles with heterogeneous terminologies, and potentially expanding the scope of the review. The decision was taken to assess the applicability of the Governance Mechanisms in Digital Platform Ecosystems: Addressing the Generativity-Control Tension framework proposed by Staub et al. (2022) to data platform ecosystems. This choice was reinforced by the alignment of governance mechanisms identified in the other relevant articles.

Reflection on the interviews

The identified governance mechanisms, grouped into three pillars (Platform Boundary Resources, Platform Rules, and Ecosystem Identity) based on the framework of Staub et al. (2022), were thoroughly discussed and validated through the use of interviewing platform provider practitioners and subject matter experts. The interview results provided valuable insights into the real-world applicability and effectiveness of the proposed governance mechanisms. While the targeted number of interviews was achieved, challenges arose in extracting detailed insights into the specific operational workings of the identified mechanisms. Follow-up questions, particularly pertaining to the practical aspects of these mechanisms, could have been more effective. Moreover, a subset of respondents encountered challenges in providing definitive responses regarding the application for mechanisms like Relational Control and Revenue Sharing indicating complexities in implementing certain governance mechanisms. Despite these limitations, it is essential to highlight that all governance mechanisms, except for Relational Control, were successfully validated.

Dal Bianco et al. (2014) assert that APIs are essential for a viable software ecosystem, a viewpoint confirmed unanimously by all participants. Tiwana (2013) adds that APIs enable developers to seamlessly utilise platform capabilities, a sentiment echoed by participant Int_8. Staub et al. (2022) highlight the significance of Development and Social Boundary Resources in assisting complementors. Participants generally support this view, though Int_6 and Int_7 note potential confusion from inadequate documentation and communication.

Niculescu et al. (2018) contends that Intellectual Property Sharing can occur through either direct inter-firm agreements, such as licensing, or by making a technology accessible to the general public through open source projects. This distinction between commercial and open licenses is acknowledged by participant of Int_2. Regarding Decision Rights, Ye and Kankanhalli (2018) recommend that platform owners grant complementors sufficient decision rights for decision-making autonomy. While the respondents agree to this stance, they highlighted that although companies usually accept standard terms and conditions, they also engage in signing additional data sharing agreements or contracts. Goldbach and Benlian (2015) state that gatekeeping practices foster an open platform and diverse complements. Participants recognise their relevance, but add that gatekeeping poses a significant challenge for organisations, mainly due to security reasons. Tiwana (2013) highlights how platform owners use pricing policies to drive complementors' personal investments, ensuring prosperity for their offerings and the overall ecosystem. However, this perspective diverges from the viewpoints expressed in interviews focused on data platforms. Participants provided mixed

responses regarding the role of Pricing in facilitating openness. While acknowledging that pricing raises awareness of data value, participants in this study also recognise potential advantages in certain cases of offering free accessibility. Additionally, Int_4 underscores the importance of treating data as a monetisable resource, advocating for the separation of concerns related to consent and the monetisation process (Int_2). However, in contrast, Int_1 expresses concerns that pricing may pose obstacles to openness, citing potential barriers such as exchange fees and technology delivery fees. Karhu et al. (2018) highlight that with Revenue Sharing, a platform owner extracts a portion of co-created value, typically a percentage from sales or service use. Participants, including Int_2, emphasise the collaborative nature in data platforms, noting the importance of hiring third parties to build services on provided data, where the platform owner receives a percentage.

Staub et al. (2022) highlight that Relational Control enhances complementor motivation and commitment through community building. However, mixed responses were observed, with Int_8 expressing caution that this might be business context specific while Int_2 note that it helps traditional German industry catch up on digitisation by fostering openness and brand enhancement.

In addition to the validation process, respondents proposed new mechanisms that were not initially part of the study framework, including Privacy, Data Sharing Agreements, Roles & Responsibilities, Common Terminology, Data Lineage, and Data Catalog. The latter four fall under classic data governance activities. Additionally, the introduction of 10 data-specific mechanisms surfaced valuable insights. Among these, "Standardisation" emerged as a noteworthy mechanism emphasised by multiple respondents for its role in fostering openness within data platform ecosystems.

5.2. Conclusions

The objective of this study was to identify the governance mechanisms that can enable openness and reconcile the tension between openness and control in data platform ecosystems. By thoroughly reviewing existing literature and conducting insightful interviews, the research identified essential governance mechanisms crucial to realise openness in data platform ecosystems. The proposed framework has proven effective in offering a structured way to validate these mechanisms.

What are the existing governance mechanisms used to enable openness in platforms?

Conducting a systematic literature review aimed at identifying governance mechanisms fostering openness in digital platforms, 12 relevant articles were found. Among these, Staub et al. (2022) introduced an extensive framework named "Governance Mechanisms in Digital Platform Ecosystems: Addressing the Generativity-Control Tension" substantiated by insights from the other articles. The framework categorises crucial governance mechanisms, like Application Programming Interfaces, Development Boundary Resources, Social Boundary Resources, Intellectual Property Sharing, Decision Rights, Gatekeeping, Pricing, Revenue Sharing, and Relational Control, into three pillars—Platform Boundary Resources, Platform Rules, and Ecosystem Identity. This synthesis provides a clear understanding and practical foundation for implementing governance mechanisms that foster openness in digital platform ecosystems.

What governance mechanisms, identified in the literature of platforms, are effective in enabling openness in data platforms specifically and why are the identified mechanisms suitable to enable openness in data platform ecosystems?

Subsequently, interviews with data platform practitioners were performed to validate the relevance of the identified governance mechanisms and the applicability of the framework for Data Platform Ecosystems. All Governance mechanisms were validated as relevant to enable openness in Data Platform Ecosystems except for Relational Control, that this might be due specific business context. Platform Boundary Resources turned out to be essential in fostering openness by establishing connectivity to the platform, complementing its functions, and receiving support for this through tools

and documentation. Additionally, Governance Mechanisms categorised under Platform Rules were deemed crucial for facilitating openness. These mechanisms offer oversight, determining access privileges, specifying authorised actions, clarifying ownership of innovations, defining pricing structures, and outlining revenue-sharing allocations. Participants highlighted that such oversight, achieved through the establishment of rules, is necessary to facilitate openness within the platform.

Our study also uncovered newly emerging mechanisms, including Privacy, Data Sharing Agreements, Roles & Responsibilities, Common Terminology, Data Lineage, and Data Catalog. The latter four fall under general Data Governance practices that is not included as a pillar in the proposed framework by Staub et al. (2022). Nevertheless, these governance mechanisms oversee the data on the platform to guarantee proper management and have been recognised as crucial for fostering openness in data platform ecosystems by multiple participants. This suggests the need to include Data Governance as a foundational pillar in the framework.

What are data-specific mechanisms used to manage data platform openness?

This study uncovered ten data-specific mechanisms that enhance openness and contribute to an improved comprehension of findability, accessibility, and increased usability of the data. "Standardisation" stood out among these mechanisms, while it emphasises the importance of structuring data for improved accessibility, as highlighted by multiple respondents in its significant role in fostering openness within data platform ecosystems.

5.3. Recommendations for practice

Based on the confirmed governance mechanisms, practical guidance can be provided to data platform practitioners. It is vitally important to prioritise the adoption of validated mechanisms like Application Boundary Resources, Decision Rights, and Gatekeeping for adequate governance. Moreover, the study underscores the significance of data-specific mechanisms, indicating the need for tailored strategies to manage openness in data platforms. Practical implementation of these mechanisms should be accompanied by robust documentation and communication to avoid potential confusion.

5.4. Recommendations for further research

The findings of the study open avenues for further research. The non-validated mechanism, Relational Control, warrants additional investigation to determine its applicability in specific business contexts. Subsequent research can delve into verifying this assertion and its practical implications. Newly suggested mechanisms, including Privacy, Data Sharing Agreements, Roles & Responsibilities, Common Terminology, Data Lineage, and Data Catalog, require further exploration. Validating these mechanisms will contribute to refining the proposed framework and the understanding of governance in data platforms. In addition, the identified data-specific mechanisms, highlighted by respondents, should be subject to a more extensive examination and validation if the identified mechanisms can be added to the framework. Furthermore, subsequent research can refine these mechanisms, providing practical insights for platform providers and contributing to the continuous advancement of our understanding and improvement of data platform ecosystems. Also, conducting future research with organisations already operating in an ecosystem with multiple organisations interacting on the data platform in a collaborative manner could provide additional viewpoints.

6. References

- Agahari, W., & de Reuver, M. (2022). Rethinking consumers' data sharing decisions with the emergence of multi-party computation: an experimental design for evaluation.
- Benlian, A., Hilkert, D., & Hess, T. (2015). How open is this platform? The meaning and measurement of platform openness from the complementers' perspective. *Journal of Information technology*, 30(3), 209-228.
- Bergman, R., Abbas, A. E., Jung, S., Werker, C., & de Reuver, M. (2022). Business model archetypes for data marketplaces in the automotive industry: Contrasting business models of data marketplaces with varying ownership and orientation structures. *Electronic Markets*, 32(2), 747-765.
- Beverungen, D., Hess, T., Köster, A., & Lehrer, C. (2022). From private digital platforms to public data spaces: implications for the digital transformation. *Electronic Markets*, 32(2), 493-501.
- Bonina, C., & Eaton, B. (2020). Cultivating open government data platform ecosystems through governance: Lessons from Buenos Aires, Mexico City and Montevideo. *Government Information Quarterly*, 37(3), 101479.
<https://doi.org/https://doi.org/10.1016/j.giq.2020.101479>
- Boudreau, K. (2010). Open platform strategies and innovation: Granting access vs. devolving control. *Management science*, 56(10), 1849-1872.
- Boudreau, K. J., & Hagiu, A. (2009). Platform rules: Multi-sided platforms as regulators. *Platforms, markets and innovation*, 1, 163-191.
- Boudreau, K. J., & Jeppesen, L. B. (2015). Unpaid crowd complementors: The platform network effect mirage. *Strategic management journal*, 36(12), 1761-1777.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2), 77-101.
- Broekhuizen, T. L. J., Emrich, O., Gijzenberg, M. J., Broekhuis, M., Donkers, B., & Slood, L. M. (2021). Digital platform openness: Drivers, dimensions and outcomes. *Journal of Business Research*, 122, 902-914. <https://doi.org/https://doi.org/10.1016/j.jbusres.2019.07.001>
- Bruun, M. H., Andersen, A. O., & Mannov, A. (2020). Infrastructures of trust and distrust: The politics and ethics of emerging cryptographic technologies. *Anthropology Today*, 36(2), 13-17.
- Chen, L., Tong, T. W., Tang, S., & Han, N. (2022). Governance and design of digital platforms: A review and future research directions on a meta-organization. *Journal of Management*, 48(1), 147-184.
- Dal Bianco, V., Myllärniemi, V., Komssi, M., & Raatikainen, M. (2014). The role of platform boundary resources in software ecosystems: a case study. 2014 IEEE/IFIP Conference on Software Architecture,
- De Prieëlle, F., De Reuver, M., & Rezaei, J. (2020). The role of ecosystem data governance in adoption of data platforms by Internet-of-Things data providers: Case of Dutch horticulture industry. *IEEE Transactions on Engineering Management*, 69(4), 940-950.
- De Reuver, M., Nederstigt, B., & Janssen, M. (2018). Launch strategies for multi-sided data analytics platforms.
- de Reuver, M., Ofe, H., Agahari, W., Abbas, A. E., & Zuiderwijk, A. (2022). The openness of data platforms: a research agenda. Proceedings of the 1st International Workshop on Data Economy,
- De Reuver, M., Sørensen, C., & Basole, R. C. (2018). The digital platform: a research agenda. *Journal of Information technology*, 33(2), 124-135.
- Dyba, T., Dingsoyr, T., & Hanssen, G. K. (2007). Applying systematic reviews to diverse study types: An experience report. First international symposium on empirical software engineering and measurement (ESEM 2007),
- Eisenmann, T. R., Parker, G., & Van Alstyne, M. (2009). Opening platforms: How, when and why. *Platforms, markets and innovation*, 6, 131-162.

- Engert, M., Evers, J., Hein, A., & Krcmar, H. (2022). The Engagement of Complementors and the Role of Platform Boundary Resources in e-Commerce Platform Ecosystems. *Information systems frontiers*, 24, 1-19. <https://doi.org/10.1007/s10796-021-10236-3>
- Farshchian, B. A., & Thomassen, H. E. (2019). Co-Creating Platform Governance Models Using Boundary Resources: a Case Study from Dementia Care Services. *Computer Supported Cooperative Work (CSCW)*, 28(3), 549-589. <https://doi.org/10.1007/s10606-019-09353-0>
- Gawer, A. (2009). Platform dynamics and strategies: from products to services. *Platforms, markets and innovation*, 45, 57.
- Gawer, A. (2014). Bridging differing perspectives on technological platforms: Toward an integrative framework. *Research policy*, 43(7), 1239-1249.
- Gawer, A., & Cusumano, M. A. (2014). Industry platforms and ecosystem innovation. *Journal of product innovation management*, 31(3), 417-433.
- Ghazawneh, A., & Henfridsson, O. (2013). Balancing platform control and external contribution in third-party development: the boundary resources model. *Information systems journal*, 23(2), 173-192.
- Goldbach, T., & Benlian, A. (2015). How social capital facilitates clan control on software platforms to enhance App-developers' performance and success.
- Jarke, M., Otto, B., & Ram, S. (2019). Data sovereignty and data space ecosystems. In (Vol. 61, pp. 549-550): Springer.
- Karhu, K., Gustafsson, R., & Lyytinen, K. (2018). Exploiting and defending open digital platforms with boundary resources. *Information systems research*, 29(2), 479-497.
- Koutroumpis, P., Leiponen, A., & Thomas, L. D. (2020). Markets for data. *Industrial and Corporate Change*, 29(3), 645-660.
- Lauf, F., Scheider, S., Bartsch, J., Herrmann, P., Radic, M., Rebbert, M., Nemat, A. T., Schlueter Langdon, C., Konrad, R., & Sunyaev, A. (2022). Linking data sovereignty and data economy: arising areas of tension.
- Mucha, T., & Seppala, T. (2020). Artificial intelligence platforms—a new research agenda for digital platform economy.
- Niculescu, M. F., Wu, D., & Xu, L. (2018). Strategic intellectual property sharing: Competition on an open technology platform under network effects. *Information systems research*, 29(2), 498-519.
- Rochet, J.-C., & Tirole, J. (2003). Platform competition in two-sided markets. *Journal of the european economic association*, 1(4), 990-1029.
- Saunders, M., Lewis, P., & Thornhill, A. (2019). *Research methods for business students*. Pearson.
- Schlagwein, D., Schoder, D., & Fischbach, K. (2010). Openness of Information Resources—A framework-based comparison of mobile platforms.
- Schreieck, M., Wiesche, M., & Krcmar, H. (2017). The Platform Owner's Challenge to Capture Value—Insights from a Business-to-Business IT Platform. ICIS,
- Soto Setzke, D., Böhm, M., & Krcmar, H. (2019). Platform openness: A systematic literature review and avenues for future research.
- Staub, N., Haki, K., Aier, S., & Winter, R. (2022). *Governance Mechanisms in Digital Platform Ecosystems: Addressing the Generativity-Control Tension* (Vol. 51). <https://aisel.aisnet.org/cais/vol51/iss1/43>
- Tan, F. T. C., Pan, S. L., & Zuo, M. (2019). Realising platform operational agility through information technology-enabled capabilities: A resource-interdependence perspective. *Information systems journal*, 29(3), 582-608.
- Tilson, D., Lyytinen, K., & Sørensen, C. (2010). Digital infrastructures: The missing IS research agenda. Research commentary. *Information systems research*, 21(4), 748-759.
- Tiwana, A. (2013). *Platform ecosystems: Aligning architecture, governance, and strategy*. Newnes.

- Tiwana, A., Konsynski, B., & Bush, A. A. (2010). Research commentary—Platform evolution: Coevolution of platform architecture, governance, and environmental dynamics. *Information systems research*, 21(4), 675-687.
- Urovi, V., Jaiman, V., Angerer, A., & Dumontier, M. (2022). LUCE: A blockchain-based data sharing platform for monitoring data License accountability and Compliance. *Blockchain: Research and Applications*, 3(4), 100102. <https://doi.org/10.1016/j.bcra.2022.100102>
- Van Alstyne, M. W., & Lenart, A. (2020). Using data and respecting users. *Communications of the ACM*, 63(11), 28-30.
- Van den Broek, T., & van Veenstra, A. F. (2015). Modes of governance in inter-organizational data collaborations.
- Van der Vlist, F. N., & Helmond, A. (2021). How partners mediate platform power: Mapping business and data partnerships in the social media ecosystem. *Big Data & Society*, 8(1), 20539517211025061.
- Vesselkov, A., Hämmäinen, H., & Töyli, J. (2019). Design and Governance of mHealth Data Sharing. *Communications of the association for information systems*, 45(1). <https://aisel.aisnet.org/cais/vol45/iss1/18>
- VU. (2022). *Humanities - Research Skills - Advanced: Smart and systematic literature searches*. Vrije Universiteit Amsterdam. <https://libguides.vu.nl/c.php?g=440243&p=3000961>
- Wareham, J., Fox, P. B., & Cano Giner, J. L. (2014). Technology ecosystem governance. *Organization science*, 25(4), 1195-1215.
- West, J., & O'mahony, S. (2008). The role of participation architecture in growing sponsored open source communities. *Industry and innovation*, 15(2), 145-168.
- Wulfert, T., & Schütte, R. (2022). Retailer's Dual Role in Digital Marketplaces. *SN Computer Science*, 3(3), 206. <https://doi.org/10.1007/s42979-022-01098-w>
- Ye, H., & Kankanhalli, A. (2018). User Service Innovation on Mobile Phone Platforms. *MIS quarterly*, 42(1), 165-A169.
- Yin, R. K. (2018). *Case study research and applications*. Sage.

Appendix 1: Systematic Literature Review process

Title	Author	Year	Stage 1	Stage 2	Stage 3	Stage 4
Open Data Governance in the Time of the Pandemic	E. A. Alampay	2022	x	x		
Open government data: a value chain model proposal	C. S. Albano	2013	x	x		
A proposed checklist for the technical maturity of open government data: an application on GCC countries	N. Alromaih; H. Albassam; H. Al-Khalifa	2016	x	x		
Big Data and Open Government Data in Public Services	M. Anshari; M. N. Almunawar; S. A. Lim	2018	x	x		
Re-Public: workflow to publish and reuse Linked Open Government Data	J. R. Beluzo; B. Dias; G. S. Craveiro; R. Araújo	2021	x	x		
Brazilian government open data: implementation, challenges, and potential opportunities	K. d. S. Brito; M. A. d. S. Costa; V. C. Garcia; S. R. d. L. Meira	2014	x	x		
Assessing the benefits of open government data: the case of Meu Congresso Nacional in Brazilian elections 2014	K. d. S. Brito; M. A. d. S. Costa; V. C. Garcia; S. R. d. L. Meira	2015	x			
A Scalable Framework for Creating Open Government Data Services from Open Government Data Catalog	M. Buranarach; P. Krataithong; S. Hinsheranan; S. Ruengittinun; T. Supnithi	2017	x	x		
A Chatbot for Searching and Exploring Open Data: Implementation and Evaluation in E-Government	I. Cantador; J. Viejo-Tardío; M. E. Cortés-Cediel; M. P. R. Bolívar	2021	x			
Open Government Data for Machine Learning Tax Recommendation	T. Cha	2020	x			
Open Data Policy Innovation Diffusion: An Analysis of Australian Federal and State Governments	A. T. Chatfield; C. G. Reddick	2016	x			
Towards an expanded and integrated open government data agenda for India	S. Chattapadhyay	2013	x			
Access and use of government data by research and advocacy organisations in India: a survey of (potential) open data ecosystem	S. Chattapadhyay	2014	x	x		
The Construction of Government Information Open Platform: Research and Practice of the Ministry of Natural Resources	H. Chen; X. Ye; L. Zhang; Q. Bu	2023	x	x		
Research on Government Regulatory of Sharing Platform Based on New Regulatory Economics	L. Chen; J. Yan	2021	x	x		
Engaging Citizens with Open Government Data: The Value of Dashboards Compared to Individual Visualizations	A. P. Chokki; A. Simonofski; B. Frénay; B. Vanderose	2022	x			
A theoretical framework for evaluating government open data platform	P.-Y. Chu; H.-L. Tseng	2016	x	x	x	x
Open Data in Support of E-governance Evaluation: A Public Value Framework	P.-Y. Chu; H.-L. Tseng	2018	x			
Transparency portals versus open government data: an assessment of openness in Brazilian municipalities	A. S. Corrêa; P. L. P. Corrêa; F. S. C. d. Silva	2014	x	x		
A collaborative-oriented middleware for structuring information to open government data	A. S. Corrêa; P. L. P. Corrêa; F. S. C. d. Silva	2015	x	x		
The Use of Open Government Data to Citizen Empowerment	G. S. Craveiro; J. A. S. Machado; J. S. Machado	2016	x			
Exploring dimensions influencing the usage of Open Government Data portals	K. Y. Dahbi; H. Lamharhar; D. Chiadmi	2018	x	x		

Approaches to assessing open government data programs: comparison of common traits and differences at global context	S. Dawes; L. Vidasova; D. Trutnev	2015	x	x		
Intelligent geovisualizations for open government data (vision paper)	A. Degbelo; C. Kray	2018	x			
Spatial search strategies for open government data: a systematic comparison	A. Degbelo; B. B. Teka	2019	x			
Empowering Cities through Open Data - Open Government Data Initiatives in India	G. Doctor; P. Joshi	2022	x	x		
Show me the data! a systematic mapping on open government data visualization	A. Eberhardt; M. S. Silveira	2018	x			
Open Government Data: Towards a comparison of Data Lifecycle models	H. Elmekki; D. Chiadmi; H. Lamharhar	2019	x	x		
Digital Juries: A Civics-Oriented Approach to Platform Governance	J. Fan; A. X. Zhang	2020	x	x	x	
Open government data as a tool for cooperation between people and government: a case study of open data and e-governance resources in the Eurasian Economic Union	O. Filatova; S. Balabanova; V. Golubev; I. Ibragimov	2017	x			
Implementing open licensing in government open data initiatives: a review of Australian government practice	A. Fitzgerald; N. Hooper; J. S. Cook	2013	x	x	x	
Civil Servants on Open Data: Perceptions of Azorean civil servants before an imminent Open Government Data initiative	L. V. Garcia	2022	x			
Open government data initiatives in Europe: northern versus southern countries analysis	Á. Gomes; D. Soares	2014	x			
Outsource or Invest? A Multiple Case Study of Digital Government Platform Strategies	Y. Gong; X. Yang	2022	x			
Government 2.0: a conceptual framework and a case study using Mexican data for assessing the evolution towards open governments	J. C. González; J. Garcia; F. Cortés; D. Carpy	2014	x			
Co-creating visual overviews for open government data	A. Graves; J. Bustos-Jiménez	2015	x			
Visualization tools for open government data	A. Graves; J. Hendler	2013	x			
Research on Security Mechanism of Hadoop Big Data Platform	W. Gu; S. Jia	2022	x			
Open Government and Open Data in Times of COVID-19	L. Hagen; R. Sandoval-Almazan; S. Okhuijsen; S. Cabaco; E. A. Ruvalcaba-Gomez; J. Villodre; W. Sung; D. Valle-Cruz	2021	x	x		
Adoption of Open Government Data in Local Government Context: Conceptual Model Development	S. I. Haini; N. Z. A. Rahim; N. M. M. Zainuddin	2019	x	x		
A rights-based approach to open government data	R. D. Hanbal; A. Prakash	2019	x	x		
Electoral Competition, Transparency, and Open Government Data	S. Hong	2020	x	x		
Open Data Standard and Analysis Framework: Towards Response Equity in Local Governments	J. Hsu; R. Ravichandran; E. Zhang; C. Keung	2021	x			
Readiness Assessment of Open Government Data Programs: A Case of Shenzhen	Y. Hu; X. Bai; S. Sun	2016	x			
Why Do Public Servants Intend to Open Government Data? The Case of Taiwan	Y.-C. Huang; T.-Y. Huang	2022	x			
A Model and Architecture for Building a Sustainable National Open Government Data (OGD) Portal	L. L. Idowu; I. I. Ali; U. G. Abdullahi	2018	x	x	x	
Is the GaaP wider than we think? Applying a sociotechnical lens to Government-as-a-Platform	D. Jamieson; R. Wilson; M. Martin	2020	x			
Towards generalized process patterns for detecting corruption within the government using open data	M. Janssen; J. Ubacht	2018	x			

Following OCDE recommendations on digital government: open innovation and data science: digital government editor's introduction	C. E. Jimenez-Gomez	2018	x			
Mobile e-services and open data in e-government processes: transforming citizen involvement	D. Johansson; J. Lassinantti; M. Wiberg	2015	x			
Exploring the Quality of Dynamic Open Government Data for Developing Data Intelligence Applications: The Case of Attica Traffic Data	A. Karamanou; P. Brimos; E. Kalampokis; K. Tarabanis	2023	x			
Assessing Information and System Quality of Open Government Data Portals in Croatia	K. Kević; C. Alexopoulos; A. K. Divjak	2022	x	x		
Sustainability Implications of Open Government Data: A Cross-Regional Study	A. Koczanski; M. Sabou	2015	x			
Quality Assessment for Open Government Data in China	X.-T. Li; J. Zhai; G.-F. Zheng; C.-F. Yuan	2018	x			
Three Types of Data Exchange in the Open Government Information Projects	Y. P. Lipuntsov	2014	x	x		
Value Research on Open Government Data: Cases of Shanghai Open Data Apps in China	W. Lyu; L. Zheng	2017	x			
Exploring the Barriers in the Commercial Use of Open Government Data	G. Magalhães; C. Roseira	2016	x			
Business models for open government data	G. Magalhaes; C. Roseira; L. Manley	2014	x			
Open government data intermediaries: a terminology framework	G. Magalhaes; C. Roseira; S. Strover	2013	x			
Digital Ecosystems to Support the Open and Collaborative Government Systems	A. M. Magdaleno; R. M. d. Araujo	2015	x			
Open government data and the data usage for improvement of public services in the Rio de Janeiro City	R. Matheus; J. C. Vaz; M. M. Ribeiro	2014	x			
CivilServant: Community-Led Experiments in Platform Governance	J. N. Matias; M. Mou	2018	x			
Digital Government Interoperability and Data Exchange Platforms: Insights from a Twenty Country Comparative Study	K. McBride; S. Kamalanathan; S.-M. Valdma; T. Toomere; M. Freudenthal	2022	x			
Open Government Data as a panacea against corruption and mismanagement: An analysis of Open Public Data on COVID-19-related procurement at the sub-national level	M. Mendes; J. Voigt	2022	x			
Visualizing aedes aegypti infestation in urban areas: a case study on open government data mashups	P. G. A. d. Mendonça; C. Maciel; J. V. Filho	2014	x			
Open government and data intermediaries: the case of AidData	E. Mercado-Lara; J. R. Gil-Garcia	2014	x			
Open Government Data Policy and Indian Ecosystems	D. Misra; A. Mishra; S. Babbar; V. Gupta	2017	x			
Translating Human Values to Design Requirements: The Case of Developing Digital Government Collaborative Platform (DGCP) for Environmental Sustainability in Sri Lanka	S. Mohamed; S. Han	2022	x			
ADACOP: A Big Data Platform for Open Government Data	A. Moreno; J. Molano-Pulido; J. E. Gomez-Morantes; R. A. Gonzalez	2022	x			
Open Government Data Usage Overview: A Systematic Literature Mapping	A. J. A. Neto; D. F. Neves; L. C. Santos; M. C. R. Junior; R. P. C. d. Nascimento	2018	x			
Towards enrichment of the open government data: a stakeholder-centered determination of High-Value Data sets for Latvia	A. Nikiforova	2022	x			
Barriers to Openly Sharing Government Data: Towards an Open Data-adapted Innovation Resistance Theory	A. Nikiforova; A. Zuiderwijk	2022	x			
Crowdsourcing platform acting as an intermediary role in Collaborative governance	I. H. Noh	2022	x			

Open Government Data as a Right for Effective Citizen Participation	A. O. Odongo; G. C. Rono	2016	x				
Open government data in Brazil a systematic review of its uses and issues	E. F. d. Oliveira; M. S. Silveira	2018	x				
Open Government Data Portals Analysis: The Brazilian Case	M. I. S. Oliveira; H. R. d. Oliveira; L. A. Oliveira; B. F. Lóscio	2016	x				
Developing a model to readiness assessment of open government data in public institutions in Colombia	M. A. Osorio-Sanabria; F. Amaya-Fernández; M. P. González-Zabala	2020	x				
An Investigation on the Application Developer Attitude Towards Greek Open Government Data	A. Plessas; I. Dionysiou	2016	x				
Open government data programs: the political system matters? revisiting the dawes' ecosystem model	E. Przybilowicz; M. A. Cunha	2018	x				
Citizens' Trust in Open Government Data: A Quantitative Study about the Effects of Data Quality, System Quality and Service Quality	A. Purwanto; A. Zuiderwijk; M. Janssen	2020	x				
Secure Mechanism of Intelligent Urban Railway Cloud Platform Based on Zero-trust Security Architecture	Y. Qlu	2022	x				
Open government data usage: a brief overview	A. Quarati; M. D. Martino	2019	x				
And Data for All: On the Validity and Usefulness of Open Government Data	W. Radl; J. Skopek; A. Komendera; S. Jäger; F. Mödritscher	2013	x				
Enhancing Open Government Data With Data Provenance	C. P. d. Reis; W. M. C. d. Silva; L. C. B. Martins; R. Pinheiro; M. C. Victorino; M. Holanda	2020	x				
A rationale for data governance as an approach to tackle recurrent drawbacks in open data portals	J. R. Reis; J. Viterbo; F. Bernardini	2018	x				
Colombian Case Study for the Analysis of Open Data Government: a Data Quality Approach	M. A. O. Sanabria; F. O. A. Fernández; M. P. G. Zabala	2018	x				
Towards an Open Government Data Comparative Model	R. Sandoval-Almazan; E. Styryn	2018	x				
Towards the open government ecosystem: open government based on artificial intelligence for the development of public policies	L. G. d. M. Santos	2018	x				
Supply-side variants in the supply of open data in university governance	F. v. Schalkwyk	2013	x				
Approaching an optimizing open linked government data portal	L. Sinif; B. Bounabat	2018	x	x			
A general framework of smart Open Linked Government Data: Application in E-health	L. Sinif; B. Bounabat	2019	x				
Public Services Redesign and Provision: Government Platform Approach	E. Styryn; K. N. Andersen	2020	x				
Open Data and Open Government: From Abstract Principles to Institutionalized Practices	E. Styryn; L. F. Luna-Reyes; T. M. Harrison	2016	x	x	x		
On the openness of digital platforms/ecosystems	J. Teixeira	2015	x	x	x	x	
Open government data: beyond policy & portal, a study in Indian context	N. Verma; M. P. Gupta	2013	x				
Challenges in publishing Open Government Data: A study in Indian context	N. Verma; M. P. Gupta	2015	x				
Proposal of a Brazilian Database Government Open Linked Data: DBgoldbr: Invited Paper	M. Victorino; M. T. d. Holanda; E. Ishikawa; E. C. Oliveira; G. Ghinea; S. Chhetri	2017	x				
A Centralized Platform of Open Government Data as Support to Applications in the Smart Cities Context	I. Vieira; A. Alvaro	2018	x				
Technological Path of Platform Governance of Central Government Departments in the Digital Era: A Case Study of the Ministry of Natural Resources	X. Wang; J. ZHOU; W. ZHOU; K. WU; Q. TU; J. WU	2023	x				
Barriers to Using Open Government Data	J. Wiczorkowski	2019	x				

Toward a Conceptual Model for Users' Online Open Government Data Interaction	F. Xiao	2021	x	x		
Challenges and Supports for Accessing Open Government Datasets: Data Guide for Better Open Data Access and Uses	F. Xiao; D. He; Y. Chi; W. Jeng; C. Tomer	2019	x	x		
Categorizing Open Government Data Users by Exploring their Challenges and Proficiency	F. Xiao; K. Thaker; D. He	2022	x			
Quality evaluation of government epidemic data openness based on cloud model and PSR theory	L. Yang; J. Wang; Z. Zhao	2022	x			
An Exploration of Factors Influencing Taiwan Government Agencies' Open Data Participation: A Multi-Group Analysis Perspective	T.-M. Yang; M.-C. Wu	2020	x			
Exploring the Information Behaviors of Government Officials in Open Data Initiatives	T.-M. Yang; Y.-J. Wu	2016	x			
Assessing the Open Data Maturity of Government Agencies Having Different Years of Implementation Experience: A Study in Taiwan	T.-M. Yang; Y.-J. Wu	2020	x			
The survey of the factors influencing the use of open government data in Taiwan: The preliminary findings	T.-M. Yang; Y.-J. Wu	2022	x			
Construction Mechanism and Path of Innovative Entrepreneurship and Education Practice Platform based on Maker Education	Y. Yang	2022	x			
An Ontology for Open Government Data Business Model	F. A. Zeleti; A. Ojo	2017	x			
Agile Mechanisms for Open Data Process Innovation in Public Sector Organizations: Towards Theory Building	F. A. Zeleti; A. Ojo	2019	x	x		
Provenance Metadata of Open Government Data Based on PROV-JSON	J. Zhai; H. Chen; C. Yuan	2017	x			
Research on Open Government Data in China: A Critical Assessment of 587 Papers	H. Zhang; L. Zheng	2022	x			
Business Ecosystem Governance and Openness in the New Digital Age: An Exploratory Model	Q. Zhang; Y. Wang	2017	x	x	x	
Coordination between Governmental Resources and Citizen Engagement with Open Government Data: A Coupling Coordinated Model	Y. Zhang; M. Janssen	2022	x			
Assessment on China's Open Government Data Platforms: Framework, Status and Problems	L. Zheng; F. Gao	2016	x	x		
Evaluating global open government data: methods and status	L. Zheng; W.-M. Kwok; V. Aquaro; X. Qi; W. Lyu	2020	x	x		
Research on Big Data Open Intelligent Platform of Guizhou Province E-government Service	G. Zhou; K. Chen; J. Tu	2019	x			
Linked Data in Business	W. Abramowicz; S. Auer; T. Heath	2016	x			
Artificial Intelligence in Information Systems: State of the Art and Research Roadmap	P. J. Ågerfalk; K. Conboy; K. Crowston; J. S. Z. Eriksson Lundström; S. Jarvenpaa; S. Ram; P. Mikalef	2022	x			
Adverse Selection in B2B Secondary Market Online Auctions for IT Equipment: An Empirical Analysis	A. Alhaili; W. J. Elmaghraby; A. Gopal	2022	x			
Software-Defined Business – Implications for IT Management	R. Alt; J. M. Leimeister; T. Priemuth; S. Sachse; N. Urbach; N. Wunderlich	2020	x			
Applying Seven Images of Science in Exploring whether Information Systems Is a Science	S. Alter	2018	x			
Facets of Work: Enriching the Description, Analysis, Design, and Evaluation of Systems in Organizations	S. Alter	2021	x			
Self-Organizing in Blockchain Infrastructures: Generativity Through Shifting Objectives and Forking	J. V. Andersen; C. I. Bogusz	2019	x			
Can Context Influence a Smart City Project? Case Study of two projects in Brazil	B. Andrade de Figueiredo; R. H. Medrano	2020	x			

When Do IT Security Investments Matter? Accounting for the Influence of Institutional Factors in the Context of Healthcare Data Breaches	C. M. Angst; E. S. Block; J. D'Arcy; K. Kelley	2017	x				
Personalization: UI Personalization, Theoretical Grounding in HCI and Design Research	O. Arazy; O. Nov; N. Kumar	2015	x				
Security-Induced Lock-In in the Cloud	D. G. Arce	2022	x				
Surveilling the SnackKids: Street Entrepreneurship Meets a Neighborhood App	A. Avery	2022	x				
Peer Review: Toward a Blockchain-enabled Market-based Ecosystem	M. Avital	2018	x	x			
FROM DATA WAREHOUSES TO ANALYTICAL ATOMS – THE INTERNET OF THINGS AS A CENTRIFUGAL FORCE IN BUSINESS INTELLIGENCE AND ANALYTICS	H. Baars; J. Ereth	2016	x				
Transformational Issues of Big Data and Analytics in Networked Business	B. Baesens; R. Bapna; J. R. Marsden; J. Vanthienen	2016	x				
Applying a Layered Framework to Disaster Recovery	C. Baham; A. Calderon; R. Hirschheim	2017	x				
The Internet of Things (IoT): A Research Agenda for Information Systems	A. Baiyere; H. Topi; V. Venkatesh; J. Wyatt; B. Donnellan	2020	x				
Corporate Sponsorship of Academic Research: The Trend, Its Drivers, and Its Implications	J. Baker; J. C. Wetherbe	2013	x				
Connecting Systems, Data, and People: A Multidisciplinary Research Roadmap for Chronic Disease Management	I. Bardhan; H. Chen; E. Karahanna	2020	x				
Service Innovation in the Digital Age: Key Contribution and Future Directions	M. Barrett; E. Davidson; J. Prabhu; S. L. Vargo	2015	x				
Concepts for Modeling Smart Cities	V. Bastidas; I. Reychar; A. Ofir; M. Bezbradica; M. Helfert	2022	x				
A Conceptual Framework for Understanding Crowdfunding	T. Beaulieu; S. Sarker; S. Sarker	2015	x				
Governance in the Blockchain Economy: A Framework and Research Agenda	R. Beck; C. Müller-Bloch; J. L. King	2018	x				
Smart Participation Design: Prescriptive Knowledge for Bottom-Up Participation	F. Becker; D. Siemon; S. Robra-Bissantz	2022	x				
Achieving Citizen Centric Smart Cities through Linked Open Data Platforms based on Human-Data Interaction Concepts	M. G. Belizario; R. C. G. Berardi	2021	x				
Complexity and Information Systems Research in the Emerging Digital World	H. Benbya; N. Nan; H. Tanriverdi; Y. Yoo	2020	x				
The Impact of Integration on Application Success and Customer Satisfaction in Mobile Device Platforms	B. Bender	2020	x				
Visions and Voices on Emerging Challenges in Digital Business Strategy	A. Bharadwaj; O. A. El Sawy; P. A. Pavlou; N. Venkatraman	2013	x				
Digital Business Strategy: Toward a Next Generation of Insights	A. Bharadwaj; O. A. El Sawy; P. A. Pavlou; N. Venkatraman	2013	x				
Exploring the Synergy Between Financial Technologies and Financial Inclusion: What We Know and Where We Should Be Heading?	V. Bharathi. S; A. Perdana; M. S. Kulkarni	2023	x				
A Value Co-creation Perspective on Information Systems Analysis and Design	M. Blaschke; K. Haki; S. Aier; R. Winter	2019	x				
Low-Code Platform	A. Bock; U. Frank	2021	x				
Partial Least Squares Structural Equation Modeling Approach for Analyzing a Model with a Binary Indicator as an Endogenous Variable	D. Bodoff; S. Y. Ho	2016	x				
An Open Platform for Modeling Method Conceptualization: The OMILAB Digital Ecosystem	D. Bork; R. A. Buchmann; D. Karagiannis; M. Lee; E.-T. Miron	2019	x	x			
Coordinating Interdependencies in an Open Source Software Project: A Replication of Lindberg, et al	R. Bradley; A. Mockus; Y. Ma; R. Zaretzki; B. Bichescu	2020	x				

User, Use & Utility Research - The Digital User as New Design Perspective in Business and Information Systems Engineering	W. Brenner; H. Österle; C. Petrie; F. Uebernickel; R. Winter; D. Karagiannis; L. Kolbe; J. Krüger; H. Plattner; L. Leifer; H.-J. Lamberti; J. M. Leimeister; G. Schwabe; R. Zarnekow	2014	x	x		
Cascading Feedback: A Longitudinal Study of a Feedback Ecosystem for Telemonitoring Patients with Chronic Disease	K. Brohman; S. Addas; J. Dixon; A. Pinsonneault	2020	x			
Special Issue Editorial	M. K.-. Brown; J. Quesenberry; F. Niederman; T. Weitzel	2018	x			
IT as Curse and Blessing	H. U. Buhl	2013	x			
When Spheres Collide: A Refocused Research Framework for Personal Use of Technology at Work	J. Burleson; B. E. Greenbaum	2019	x			
Next-Generation Information Systems Theorizing: A Call to Action	A. Burton-Jones; B. S. Butler; S. Scott; S. X. Xu	2021	x			
Digital Business Strategies for Incumbent Firms. How a Scandinavian hotel chain competes with the internet giants	B. Bygstad; J. Iden; A. Ulfsten	2020	x			
Trends in FinTech Research and Practice: Examining the Intersection with the Information Systems Field	C. Cai; M. Marrone; M. Linnenluecke	2022	x			
What the History of Linux Says About the Future of Cryptocurrencies	A. Carvalho; C. Sambhara; P. Young	2020	x			
Advancing Qualitative IS Research Methodologies: Expanding Horizons and Seeking New Paths	D. Cecez-Kecmanovic; R. M. Davison; W. Fernandez; P. Finnegan; S. L. Pan; S. Sarker	2020	x			
Assessing Incumbents' Risk of Digital Platform Disruption	C. Cennamo; L. Diaferia; A. Gaur; G. Salviotti	2022	x			
Blockchain for the IoT: Privacy-Preserving Protection of Sensor Data	M. Chanson; A. Bogner; D. Bilgeri; E. Fleisch; F. Wortmann	2019	x	x		
How SME Watkins Steel Transformed from Traditional Steel Fabrication to Digital Service Provision	F. Chasin; M. Kowalkiewicz; T. Gollhardt	2022	x			
Business on Chain: A Comparative Case Study of Five Blockchain-Inspired Business Models	A. Y. L. Chong; E. T. K. Lim; X. Hua; S. Zheng; C.-W. Tan	2019	x			
Privacy Concerns and Data Sharing in the Internet of Things: Mixed Methods Evidence from Connected Cars	P. Cichy; T.-O. Salge; R. Kohli	2021	x	x		
Digital Innovation	R. F. Ciriello; A. Richter; G. Schwabe	2018	x			
Research Perspectives: Through Whose Eyes? The Critical Concept of Researcher Perspective	R. Clarke; R. M. Davison	2020	x			
Introduction to the Special Issue on HCI in a Sharing Society	T. Clemmensen; D. Rajanen; M. Rajanen; J. Abdelnour-Nocera	2019	x			
Integrating Across Sustainability, Political, and Administrative Spheres: A Longitudinal Study of Actors' Engagement in Open Data Ecosystems in Three Canadian Cities	J. Corbett; M. Templier; H. Townsend; H. Takeda	2020	x	x		
A Theoretical Framework to Develop a Research Agenda for Information Systems Innovation	G. J. Costello; B. Donnellan; M. Curley	2013	x			
A Value-sensitive Design Perspective of Cryptocurrencies: A Research Agenda	K. Cousins; H. Subramanian; P. Esmailzadeh	2019	x			
The Regulatory Issues Affecting Mobile Financial Systems: Promises, Challenges, and a Research Agenda	K. D. Cousins; U. Varshney	2014	x			
Evaluation of Action Design Research	S. Cronholm; H. Göbel	2019	x			
Robbing Peter to Pay Paul: Surrendering Privacy for Security's Sake in an Identity Ecosystem	R. Crossler; C. Posey	2017	x			
Identifying Digital Transformation Paradoxes	L. Danneels; S. Viaene	2022	x			
Trust in "Trust-free" Digital Networks: How Inter-firm Algorithmic Relationships Embed the Cardinal Principles of Value Co-creation	A. Das	2020	x			

Online Child Sexual Exploitation: A New MIS Challenge	D. Demetis; J. Kietzmann	2021	x			
Centralization and Decentralization Decisions: Multiple Contingencies for IT Governance in the Public Sector	J. S. Denford; G. Dawson; K. C. Desouza	2020	x	x		
At the Crossroads between Digital Innovation and Digital Transformation	K. Drechsler; R. Gregory; H.-T. Wagner; S. Tumbas	2020	x			
How AUDI AG Established Big Data Analytics in Its Digital Transformation	C. Dremel; M. Herterich; J. Wulf; J.-C. Waizmann; W. Brenner	2017	x			
Modular Architecture of Value-Added Applications for German Healthcare Telematics	S. Dünnebeil; H. Krömer; A. Sunyeav; J. M. Leimeister	2013	x			
Distributed Tuning of Boundary Resources: The Case of Apple's iOS Service System	B. Eaton; S. Elaluf-Calderwood; C. Sorensen; Y. Yoo	2015	x			
Achieving Payoffs from an Industry Cloud Ecosystem at BankID	B. Eaton; H. K. Hallingby; P.-J. Nesse; O. Hanseth	2014	x			
Towards a Process Theory of IS Business Value Co-creation. Insights from enterprise systems adoption in an SME cluster	T. R. Eikebrokk; D. H. Olsen	2020	x			
How LEGO Built the Foundations and Enterprise Capabilities for Digital Leadership	O. El Sawy; H. Amsinck; P. Kraemmergaard; A. Lerbech Vinther	2016	x			
Decision-making Processes in Community-based Free/Libre Open Source Software-development Teams with Internal Governance: An Extension to Decision-making Theory	U. Y. Eseryel; K. Wei; K. Crowston	2020	x			
Digital Innovation as a Fundamental and Powerful Concept in the Information Systems Curriculum	R. G. Fichman; B. L. Dos Santos; Z. Zheng	2014	x			
Becoming a Most Digitalized Country: A History of Digital Organizational Resilience in Denmark	B. Fleron; J. Pries-Heje; R. Baskerville	2022	x			
Examining Ecosystems and Infrastructure Perspectives of Platforms: The Case of Small Tourism Service Providers in Indonesia and Rwanda	C. Foster; C. Bentley	2022	x			
"Impact Engineering" or Social Responsibility? - A Debate on the Responsibilities of Academics and Ways to Combine Scientific with Practice Impact	U. Frank; R. Winter	2015	x			
Three Differentiation Strategies for Competing in the Sharing Economy	A. Frey; M. Trenz; D. Veit	2019	x			
The Routinization of Open Source Project Engagement: The Case of Open Source Risk Management Routines	M. Germonprez; R. A. Gandhi; G. Link	2021	x			
Tapestries of Innovation: Structures of Contemporary Open Source Project Engagements	M. Germonprez; M. Levy; J. E. Kendall; K. E. Kendall	2020	x			
Implementing Disruptive Technologies: What Have We Learned?	A. S. Ghawe; Y. Chan	2022	x			
Information Systems Solutions for Environmental Sustainability: How Can We Do More?	R. Gholami; R. T. Watson; H. Hasan; A. Molla; N. Bjorn-Andersen	2016	x			
Empowering Patients Using Smart Mobile Health Platforms: Evidence of a Randomized Field Experiment	A. Ghose; X. Guo; B. Li; Y. Dang	2022	x			
Design for Empowerment: Empowering Sri Lankan Farmers through Mobile-based Information System	T. Ginige; D. Richards; A. Ginige; M. Hitchens	2020	x			
Poised Between 'a Wild West of Predictive Analytics' and 'an Analytics of Things Westworld Frontier'	M. Goul	2018	x			
Responding to Cybersecurity Challenges: Securing Vulnerable U.S. Emergency Alert Systems	A. W. Green; A. B. Woszczyński; K. Dodson; P. Easton	2020	x			
Laboring Online: Are There "New" Labor Processes In Virtual Game Worlds?	A. Greenhill; G. Fletcher	2013	x			

Special Issue Editorial: Strategies for Surviving and Thriving Within and Between Digital Platforms	V. Grover; K. Lyytinen	2021	x	x		
Special Issue Editorial: Platform Competition in the Digital Era - Overview and Research Directions	V. Grover; K. Lyytinen	2022	x			
Emotions in the Twitterverse and Implications for User Interface Design	A. Gruzd	2013	x			
How Trust Leads to Commitment on Microsourcing Platforms: Unraveling the Effects of Governance and Third-Party Mechanisms on Triadic Microsourcing Relationships	W. Guo; D. W. Straub; P. Zhang; Z. Cai	2021	x	x		
A Field-Based View on Gender in the Information Systems Discipline: Preliminary Evidence and an Agenda for Change	B. Gupta; E. T. Loiacono; I. Dutchak; J. B. Thatcher	2019	x			
How a Traditional Company Seeded New Analytics Capabilities	G. Gust; C. M. Flath; T. Brandt; P. Strohle; D. Neumann	2017	x			
Blueprinting Crowdfunding - Designing a Crowdfunding Service Configuration Framework	P. Haas; I. Blohm	2017	x			
How OpusCapita Used Internal RPA Capabilities to Offer Services to Clients	P. Hallikainen; R. Bekkhus; S. Pan	2018	x			
The Dynamics of Architecture-Governance Configurations: An Assemblage Theory Approach	O. Hanseth; J. R. Modol	2021	x			
The Generative Mechanisms of Digital Infrastructure Evolution	O. Henfridsson; B. Bygstad	2013	x			
How IT Governance can assist IoT project implementation	D. Henriques; R. Pereira; I. S. Bianchi; R. Almeida; M. Mira da Silva	2020	x			
Stepwise Evolution of Capabilities for Harnessing Digital Data Streams in Data-Driven Industrial Services	M. M. Herterich; F. Uebernickel; W. Brenner	2016	x			
Ethics in Information Systems and Design Science Research: Five Perspectives	A. Herwix; A. Haj-Bolouri; M. Rossi; M. C. Tremblay; S. Puro; S. Gregor	2022	x			
In Stars We Trust – A Note on Reputation Portability Between Digital Platforms	M. Hesse; T. Teubner; M. T. P. Adam	2022	x			
Understanding the Behavior of Workshop Facilitators in Systems Analysis and Design Projects: Developing Theory from Process Modeling Projects	A. Hjalmarsson; J. Recker; M. Rosemann; M. Lind	2015	x			
Key Lessons from Bosch for Incumbent Firms Entering the Platform Economy	D. Hodapp; F. Hawlitschek; F. Wortmann; M. Lang; O. Gassmann	2022	x			
Orchestrating Digital Innovation: The Case of the Swedish Center for Digital Innovation	J. Holmström; J. Magnusson; M. Mähring	2021	x			
The Impact of Blockchain Technology on Business Models in the Payments Industry	F. Holotiuk; F. Pisani; J. Moormann	2017	x			
Hourly Wages in Crowdfunding: A Meta-Analysis	L. Hornuf; D. Vrankar	2022	x			
Growing on Steroids: Rapidly Scaling the User Base of Digital Ventures Through Digital Innovation	J. Huang; O. Henfridsson; M. Liu; S. Newell	2017	x			
Platform Sponsor Investments and User Contributions in Knowledge Communities: The Role of Knowledge Seeding	P. Huang; A. Tafti; S. Mithas	2018	x			
A Blockchain-Based Approach Towards Overcoming Financial Fraud in Public Sector Services	H. Hyvärinen; M. Risius; G. Friis	2017	x			
Stability of Transaction Fees in Bitcoin: A Supply and Demand Perspective	N. Ilk; G. Shang; S. Fan; J. L. Zhao	2021	x			
Sustainable Healthcare Systems. A holistic perspective on the use and impact of medication management robots in home healthcare	S. Iqbal; P. Jokela; T. Hammar; A.-L. Nilsson	2021	x			
The role of the agency to overcome alienation in social networking platforms: A self-reported phenomenological investigation	P. J. Jácome de Moura	2022	x			

An Information Systems Design Theory for Service Network Effects	C. Janiesch; C. Rosenkranz; U. Scholten	2020	x				
How to Create Personas: Three Persona Creation Methodologies with Implications for Practical Employment	B. J. Jansen; S.-G. Jung; L. Nielsen; K. W. Guan; J. Salminen	2022	x				
Digital Probes as Opening Possibilities of Generativity	S. Jarvenpaa; W. Standaert	2018	x				
How TradeLens Delivers Business Value With Blockchain Technology	T. Jensen; J. Hedman; S. Henningsson	2019	x				
Conflict and Development: A Headquarter Intervention View of IT Subsidiary Evolution	A. K. Jha; I. Bose	2019	x				
Knowledge Profile in PAJAIS: A Review of Literature and Future Research Directions	J. J. Jiang; T.-P. Liang; J. C.-A. Tsai	2019	x				
Universal Wallets	K. P. Jorgensen; R. Beck	2022	x				
Guidelines for Education in Business and Information Systems Engineering at Tertiary Institutions	R. Jung; C. Lehrer	2017	x				
Supporting the digital transformation of SMEs – trained digital evangelists facilitating the positioning phase	J. Kääriäinen; L. Saari; M. Tihinen; S. Perätalo	2023	x				
The Ambivalent Ontology of Digital Artifacts ¹	J. Kallinikos; A. Aaltonen; A. Marton	2013	x				
Avoiding an Oppressive Future of Machine Learning: A Design Theory for Emancipatory Assistants	K. C. Kane; A. Young; A. Majchrzak; S. B. Ransbotham	2021	x				
Comparing Potential and Actual Innovators: An Empirical Study of Mobile Data Services Innovation	A. Kankanhalli; J. Hua Ye; H. H. Teo	2015	x				
Information Technology Investment and Commercialized Innovation Performance: Dynamic Adjustment Costs and Curvilinear Impacts	P. Karhade; J. Q. Dong	2021	x				
Missing Impact of Ratings on Platform Participation in India: A Call for Research in GREAT Domains	P. Karhade; A. Kathuria	2020	x				
Four Tactics for Implementing a Balanced Digital Platform Strategy	K. Karhu; R. Gustafsson; B. Eaton; O. Henfridsson; C. Sørensen	2020	x	x	x	x	x
A Case for a New IT Ecosystem: On-The-Fly Computing	H. Karl; D. Kundisch; F. Meyer auf der Heide; H. Wehrheim	2020	x				
Information Systems Development as Value Co-Creation	K. Kautz; G. Bjerknes	2020	x				
Pathways to Developing Digital Capabilities within Entrepreneurial Initiatives in Pre-Digital Organizations	R. Keller; P. Ollig; P. Rövekamp	2022	x				
Special Issue Editorial: Addressing Societal Challenges through Analytics: An ESG ICE Framework and Research Agenda	W. Ketter; B. Padmanabhan; G. Pant; T. S. Raghu	2020	x				
Competitive Benchmarking: An IS Research Approach to Address Wicked Problems with Big Data and Analytics	W. Ketter; M. Peters; J. Collins; A. Gupta	2016	x				
Towards a framework for developing visual analytics in supply chain environments	A. Khakpour; R. Colomo-Palacios; A. Martini	2023	x				
Value Co-Creation in Smart Services: A Functional Affordances Perspective on Smart Personal Assistants	R. Knote; A. Janson; M. Söllner; J. M. Leimeister	2021	x				
Exploring Incentives and Challenges for Cybersecurity Intelligence Sharing (CIS) across Organizations: A Systematic Review	F. Kolini; L. J. Janczewski	2022	x				
Extending Digital Infrastructures: A Typology of Growth Tactics	D. Koutsikouri; R. Lindgren; O. Henfridsson; D. Rudmark	2018	x				
Blockchain Token Sale, Economic and Technological Foundations	L. M. Kranz; E. Nagel; Y. Yoo	2019	x				

A Glimpse into Policy Informatics: The Case of Participatory Platforms that Generate Synthetic Empathy	R. Krishnamurthy; A. Bhagwatwar; E. W. Johnston; K. C. Desouza	2013	x				
From Physical Food Security to Digital Food Security. Delivering value through blockchain	A. Kumar; M. Paul; P. Upadhyay	2021	x				
Governance of Social Media Platforms: A Literature Review	A. M. Kumar; S. Gupta	2023	x	x			
Special Issue Editorial: Delivering Business Value through Enterprise Blockchain Applications	M. Lacity; R. Sabherwal; C. Sørensen	2019	x				
Becoming Strategic with Intelligent Automation	M. Lacity; L. Willcocks	2021	x				
How Walmart Canada Used Blockchain Technology to Reimagine Freight Invoice Processing	M. C. Lacity; R. Van Hoek	2021	x				
An Explorative Dive into Decision Rights and Governance of Blockchain: A Literature Review and Empirical Study	S. Lewis; K. Smit; J. van Meerten	2021	x	x			
How to Harness Open Technologies for Digital Platform Advantage	H. Legenvre; E. Autio; A.-P. Hameri	2022	x	x			
Digitalization: Opportunity and Challenge for the Business and Information Systems Engineering Community	C. Legner; T. Eymann; T. Hess; C. Matt; T. Boehmann; P. Drews; A. Maedche; N. Urbach; F. Ahlemann	2017	x				
Offerings That are “Ever-in-the-Making”	J. Lehmann; J. Recker	2022	x				
Designing Digital Market Offerings: How Digital Ventures Navigate the Tension Between Generative Digital Technology and the Current Environment	J. Lehmann; J. Recker; Y. Yoo; C. Rosenkranz	2022	x				
Platform Leadership: Managing Boundaries for the Network Growth of Digital Platforms	C. Leong; S. L. Pan; D. E. Leidner; J.-S. Huang	2019	x				
Robust Action Strategies in a Connected but Unequal World: Revisiting American Pragmatism for Social Justice focused Research in Information Systems	M. Levy; A. Lanamäki; R. Hirschheim	2020	x				
The Building Blocks of Software Platforms: Understanding the Past to Forge the Future	H. Li; W. J. Kettinger	2021	x				
How Instacart Leveraged Digital Resources for Strategic Advantage	T. Li; Y. E. Chan; N. Levallet	2022	x				
“My Precious!”: A Values-Affordances Perspective on the Adoption of Bitcoin	C. Lichti; A. Tumasjan	2023	x				
An Exploration of Capitation Payments for Healthcare Services Using a Multi-agent Simulation	F.-r. Lin; R.-W. Po; M.-C. Lin; W.-Y. Lin; B. Chang	2014	x				
The Dialectics of Technology Standardization	R. Lindgren; L. Mathiassen; U. Schultze	2021	x				
Artificial Intelligence as a Service	S. Lins; K. Pandl; H. Teigeler; S. Thiebes; C. Bayer; A. Sunyaev	2021	x				
Designing Monitoring Systems for Continuous Certification of Cloud Services: Deriving Meta-requirements and Design Guidelines	S. Lins; S. Schneider; J. Szefer; S. Ibraheem; A. Sunyaev	2019	x				
Designing for Crowdfunding Co-creation – How to Leverage the Potential of Backers for Product Development	N. Lipusch; D. Dellermann; U. Bretschneider; P. Ebel; J. M. Leimeister	2020	x				
Building the Next Information Superhighway: A Critical Analysis of China's Recent National Broadband Plan	C. Liu	2016	x				
Investigating the Relationship between the Effectiveness of App Evolution and App Continuance Intention: An Empirical Study of the U.S. App Market	Y. Liu; J. Song; X. Wang; X. Tang	2021	x				
Digital Transformation: Environmental Friend or Foe? Panel Discussion at the Australasian Conference on Information Systems 2019	S. Lokuge; D. Sedera; V. Cooper; F. Burstein	2021	x				

Impactful Research on Transformational Information Technology: An Opportunity to Inform New Audiences	H. C. J. Lucas; R. Agarwal; O. A. El Sawy; B. Weber	2013	x			
Service Innovation: A Service-Dominant Logic Perspective	R. F. Lusch; S. Nambisan	2015	x			
MaCuDE IS Task Force Phase II Report: Views of Industry Leaders on Big Data Analytics and AI	K. Lyytinen; H. Topi; J. Tang	2023	x			
Designing for Digital Transformation: Lessons for Information Systems Research from the Study of ICT and Societal Challenges	A. Majchrzak; M. L. Markus; J. D. Wareham	2016	x			
Taming the Complexity in Search Matching: Two-Sided Recommender Systems on Digital Platforms	O. Malgonde; H. Zhang; B. Padmanabhan; M. Limayem	2020	x			
Resilience in the Open Source Software Community: How Pandemic and Unemployment Shocks Influence Contributions to Others' and One's Own Projects	O. S. Malgonde; T. J. V. Saldanha; S. Mithas	2023	x			
How an Incumbent Telecoms Operator Became an IoT Ecosystem Orchestrator	C. Marheine; C. Engel; A. Back	2021	x			
Guest Editorial: Theories of Digital Transformation: A Progress Report	M. L. Markus; F. Rowe	2021	x			
Internet of Things Adoption for Saudi Healthcare Services	F. Masmali; S. J. Miah	2021	x			
How an Enterprise Blockchain Application in the U.S. Pharmaceuticals Supply Chain is Saving Lives	J. Mattke; C. Maier; A. Hund; T. Weitzel	2019	x			
Digital Industrial Platforms	M. Matzner; T. Pauli; E. Fieft	2021	x	x		
Is Information Systems a Science? Rejoinder to Five Commentaries	N. McBride	2018	x			
Loot Boxes, Problem Gambling, and Problem Gaming: A Critical Review of the Emerging Literature	M. McCaffrey	2023	x			
Improving Collaboration in Ambiguous Settings by Recognizing the Importance of a Common Domain Vocabulary: An Active Learning Exercise	E. H. McKinney; S. G. Green; K. A. Heppard; J. Wu	2023	x			
Models for API Value Generation	N. P. Melville; R. Kohli	2021	x			
New Guidelines for Null Hypothesis Significance Testing in Hypothetico-Deductive IS Research	W. Mertens; J. Recker	2020	x			
Corporate Digital Responsibility	C. Mihale-Wilson; O. Hinz; W. M. P. van der Aalst; C. Weinhardt	2022	x			
The Sustainability of Polycentric Information Commons	V. Mindel; L. Mathiassen; A. Rai	2018	x			
Discursive Fields and the Diversity-Coherence Paradox: An Ecological Perspective on the Blockchain Community Discourse	S. M. Miranda; D. Wang; C. Tian	2022	x			
When and Who Do Platform Companies Acquire? Understanding the Role of Acquisitions in the Growth of Platform Companies	M. Miric; M. Pagani; O. A. El Sawy	2021	x			
Competing Tasks and Task Quality: An Empirical Study of Crowdsourcing Contests	J. Mo; S. Sarkar; S. Menon	2021	x			
Characteristics of Blockchain and Smart Services, for Smart Governments: A systematic review of the literature	A. A. D. Mohamed; Y. M. Alkhateeb; P. Agarwal; A. R. Abdelwahab	2022	x			
Establishing a Data Science for Good Ecosystem: The Case of ATLytiCS	T. Monroe-White; B. Wright; W. Hulsey; E. Kushins; A. Hord	2023	x			
Addressing the Role of Digital Technology in the COVID-19 Recovery	R. Montealegre	2020	x			
A Coevolutionary View of Information Services Development: Lessons from the U.S. National Oceanic and Atmospheric Administration	R. Montealegre; D. Hovorka; M. Germonprez	2014	x			
Episodic Peripheral Contributors and Technical Dependencies in Open Source Software (OSS) Ecosystems	E. Moon	2021	x			

KYC Optimization Using Distributed Ledger Technology	J. P. Moyano; O. Ross	2017	x			
Corporate Digital Responsibility	B. Müller	2022	x			
Managing Ambidexterity in Startups Pursuing Digital Innovation	S. D. Müller; N. Páske; L. Rodil	2019	x			
Tutorial: A Descriptive Introduction to the Blockchain	M. Murray	2019	x			
Information Technology and Product/Service Innovation: A Brief Assessment and Some Suggestions for Future Research	S. Nambisan	2013	x			
Digital Innovation Management: Reinventing Innovation Management Research in a Digital World	S. Nambisan; K. Lyytinen; A. Majchrzak; M. Song	2017	x			
Collective Information System Use: A Typological Theory	B. Negoita; L. Lapointe; S. Rivard	2018	x			
How Schlumberger Achieved Networked Information Leadership by Transitioning to a Product-Platform Software Architecture	J. J. Nehme; S. C. Srivastava; H. Bouzas; L. Carcasset	2015	x			
Assessing IS Research Impact	F. Niederman; K. Crowston; H. Koch; H. Krcmar; P. Powell; E. B. Swanson	2015	x			
A Consumer Perspective on Mobile Service Platforms: A Conjoint Analysis Approach	S. Nikou; H. Bouwman; M. de Reuver	2014	x			
Methodology for the Implementation of Knowledge Management Systems 2.0 - A Case Study in an Oil and Gas Company	S. Orenge-Roglá; R. Chalmeta	2019	x			
The Algorithm and the Crowd: Considering the Materiality of Service Innovation	W. Orlikowski; S. V. Scott	2015	x			
Digital Business Strategy and Value Creation: Framing the Dynamic Cycle of Control Points	M. Pagani	2013	x			
Exploring Factors Influencing Organizational Adoption of Artificial Intelligence (AI) in Corporate Social Responsibility (CSR) Initiatives	V. Pai; S. Chandra	2022	x			
Board Independence and Firm Performance in the IT Industry: The Moderating Role of New Entry Threats	Y. Pan; P. Huang; A. Gopal	2018	x			
The Value of Centralized IT in Building Resilience During Crises: Evidence from U.S. Higher Education's Transition to Emergency Remote Teaching	J. Park; Y. Son; C. M. Angst	2023	x			
The Role of Business Intelligence and Communication Technologies in Organizational Agility: A Configurational Approach	Y. Park; O. A. El Sawy; P. Fiss	2017	x			
Organized Complexity of Digital Business Strategy: A Configurational Perspective	Y. Park; S. Mithas	2020	x			
Platform Ecosystems: How Developers Invert the Firm	G. Parker; M. V. Alstyne; X. Jiang	2017	x			
The Nested Materiality of Environmental Monitoring	E. Parmiggiani; E. Monteiro	2015	x			
In the Backrooms of Data Science	E. Parmiggiani; T. Østerlie; P. G. Almklov	2022	x			
Should Gamification be Personalized? A Self-deterministic Approach	M. Passalacqua; P. D. Senecal, Sylvain; M. Fredette; L. E. Nacke; R. Pellerin; P.-M. Leger	2021	x			
A Ten-Step Decision Path to Determine When to Use Blockchain Technologies	A. B. Pedersen; M. Risius; R. Beck	2019	x			
Connectivity and Continuity: New Fronts in the Platform War	J. Pérez; M. Murray; J. Fluker; D. Fluker; Z. Bailes	2017	x			
Emerging Digital Frontiers for Service Innovation	C. Peters; P. Maglio; R. Badinelli; R. R. Harmon; R. Maull; J. C. Spohrer; T. Tuunanen; S. L. Vargo; J. J. Welser; H. Demirkan; T. L. Griffith; Y. Moghaddam	2016	x			

Digital Strategic Initiatives and Digital Resources: Construct Definition and Future Research Directions	G. Piccoli; J. Rodriguez; V. Grover	2022	x			
Affective Politics and Technology Buy-In: A Framework of Social, Political, and Fantasmatic Logics	E. Pignot; D. Nicolini; M. Thompson	2020	x			
Managing the Scientific Literature Overload: Digital Object Identifiers and Scopus	D. Power; R. Hadidi	2018	x			
Stray Off Topic to Stay On Topic: Preserving Interaction and Team Morale in a Highly Collaborative Course while at a Distance	L. Przybilla; K. Klinker; M. Kauschinger; H. Krcmar	2021	x			
Sharing Economy	T. Puschmann; R. Alt	2016	x			
Celebrating 40 Years of MIS Quarterly: MISQ's History and Future Through the Lenses of its Editors-in-Chief	A. Rai	2016	x			
Editor's Comments: Diversity of Design Science Research	A. Rai	2017	x			
Editor's Comments: Next-Generation Digital Platforms: Toward Human-AI Hybrids	A. Rai; P. Constantinides; S. Sarker	2019	x			
Author Responsibilities in Improving the Quality of Peer Reviews: A Rejoinder to Iivari (2016)	J. Recker	2016	x			
Using Information Systems in Innovation Networks: Uncovering Network Resources	S.-V. Rehm; L. Goel; I. Junglas	2017	x			
Mixed-Methods in Information Systems Research: Status Quo, Core Concepts, and Future Research Implications	L. Reis; C. Maier; T. Weitzel	2022	x			
Determinants of Continuance Intention to Use Open Data Website: An Insight from Indonesia	W. Resti Fitriani; A. N. Hidayanto; P. I. Sandhyaduhita; B. Purwandari; M. Kosandi	2019	x			
Building a Blockchain Application that Complies with the EU General Data Protection Regulation	A. Rieger; F. Guggenmos; J. Lockl; G. Fridgen; N. Urbach	2019	x			
A Blockchain Research Framework - What We (don't) Know, Where We Go from Here, and How We Will Get There	M. Risius; K. Spohrer	2017	x			
P for Platform. Architectures of large-scale participatory design	L. K. Roland; T. A. Sanner; J. I. Sæbø; E. Monteiro	2017	x			
Blockchain Research in Information Systems: Current Trends and an Inclusive Future Research Agenda	M. Rossi; C. Mueller-Bloch; J. B. Thatcher; R. Beck	2019	x			
Platform Provider by Accident - A Case Study of Digital Platform Coring	T. Saarikko	2016	x	x		
Fujisawa Sustainable Smart Town: Panasonic's Challenge in Building a Sustainable Society	M. Sakurai; J. Kokuryo	2018	x			
Digital Agility: Conceptualizing Agility for the Digital Era	H. Salmela; A. Baiyere; T. Tapanainen; R. D. Galliers	2022	x			
High Reliability in Digital Organizing: Mindlessness, the Frame Problem, and Digital Operations	A. Salovaara; K. Lyytinen; E. Penttinen	2019	x			
Digitization and Phase Transitions in Platform Organizing Logics: Evidence from the Process Automation Industry	J. Sandberg; J. Holmström; K. Lyytinen	2020	x			
From Expert Discipline to Common Practice: A Vision and Research Agenda for Extending the Reach of Enterprise Modeling	K. Sandkuhl; H.-G. Fill; S. Hoppenbrouwers; J. Krogstie; F. Matthes; A. Opdahl; G. Schwabe; Ö. Uludag; R. Winter	2018	x			
Grafting: Balancing Control and Cultivation in Information Infrastructure Innovation	T. A. Sanner; T. D. Manda; P. Nielsen	2014	x			
Editorial: A Brief Retrospective (2013-2019)	S. Sarker; P. Palvia; A. Oguz; C. Wu; X. Xiao	2020	x			
Threat Intelligence Sharing Platforms: An Exploratory Study of Software Vendors and Research Perspectives	C. Sauerwein; C. Sillaber; A. Mussmann; R. Breu	2017	x			

Capturing the Complexity of Malleable IT Use: Adaptive Structuration Theory for Individuals	K. Schmitz; J. T. C. Teng; K. J. Webb	2016	x				
Overcoming Challenges to Enable the Potential of Metaverse Platforms: A Qualitative Approach to Understand Value Creation	S. M. Schöbel; F. Tingelhoff	2023	x				
Mini-App Ecosystems	M. Schreieck; A. Ou; H. Krcmar	2023	x				
From Product Platform Ecosystem to Innovation Platform Ecosystem: An Institutional Perspective on the Governance of Ecosystem Transformations	M. Schreieck; M. Wiesche; H. Krcmar	2022	x	x	x		
Literature Reviews in IS Research: What Can Be Learnt from the Past and Other Fields?	G. Schryen; A. Benlian; F. Rowe; S. Gregor; K. Larsen; S. Petter; G. Paré; G. Wagner; S. Haag; E. Yasasin	2017	x				
The Future of Outsourcing in the Asia-Pacific Region: Implications for Research and Practice— Panel Report from PACIS 2014	D. Sedera; S. Lokuge; H. Krcmar; S. C. Srivastava; R. M. N.	2014	x				
The Energy Consumption of Blockchain Technology: Beyond Myth	J. Sedlmeir; H. U. Buhl; G. Fridgen; R. Keller	2020	x				
The Sustainability Imperative in Information Systems Research	S. Seidel; P. Bharati; G. Fridgen; R. T. Watson; A. Albizri; M.-C. Boudreau; T. Butler; L. C. Kruse; I. Guzman; H. Karsten; H. Lee; N. Melville; D. Rush; J. Toland; S. Watts	2017	x				
How to Empower Users for Co-Creation - Conceptualizing an Engagement Platform for Benefits Realization	M. Semmann; C. Grotherr	2017	x				
A Market-Based Approach to Facilitate the Organizational Adoption of Software Component Reuse Strategies	D. Shang; K. Lang; R. Vragov	2022	x				
Leveraging Smart Technology for User Experience Personalization – A Comparative Case Study of Innovative Payment Systems	S. S. C. Shang; L. S. L. Chiu	2022	x				
The Transformative Effect of the Internet of Things on Business and Society	J. P. Shim; M. Avital; A. R. Dennis; M. Rossi; C. Sørensen; A. French	2019	x				
The Internet of Things: Multi-faceted Research Perspectives	J. P. Shim; R. Sharda; A. M. French; R. A. Syler; K. P. Patten	2020	x				
The Transformative Effect of 5G on Business and Society in the Age of the Fourth Industrial Revolution	J. P. Shim; R. van den Dam; S. Aiello; J. Penttinen; R. Sharda; A. French	2022	x				
Information Systems Analysis and Design: Past Revolutions, Present Challenges, and Future Research Directions	K. Siau; C. Woo; V. C. Storey; R. H. L. Chiang; C. E. H. Chua; J. W. Beard	2022	x				
Hearing the Voice of Citizens in Smart City Design: The CitiVoiceFramework	A. Simonofski; E. a. S. Asensio; M. Snoeck; J. De Smedt	2019	x				
Physical Learning Environment Challenges in the Digital Divide: How to Design Effective Instruction during COVID-19?	J. B. Singh; S. K. Sharma; P. Gupta	2021	x				
How Spotify Balanced Trade-Offs in Pursuing Digital Platform Growth	D. A. Skog; J. Sandberg; H. Wimelius	2021	x				
Digital Disruption	D. A. Skog; H. Wimelius; J. Sandberg	2018	x				
The Ecosystem of Software Platform: A Study of Asymmetric Cross-Side Network Effects	P. Song; L. Xue; A. Rai; Z. Cheng	2018	x				
The Curse of the Smart Machine? Digitalisation and the children of the mainframe	C. Sørensen	2016	x				
Platform openness: A systematic literature review and avenues for future research	D. Soto Setzke; M. Böhm; H. Krcmar	2019	x	x	x	x	
How Orange Achieved Long-Term Adaptability and Value Creation by Proactively Transforming its Business	S. C. Srivastava; J. J. Nehme; A. Luherne; B. Kone	2021	x				
Social Participation among the Elderly: Moderated Mediation Model of Information and Communication Technology (ICT)	S. K. Srivastava; P. K. Panigrahi	2019	x				

Rejoined and Regenerated: Response to Responses to “Reviews, Reviewers, and Reviewing: The ‘Tragedy of the Commons’ in the Scientific Publication Process”	T. Stafford	2018	x				
Governance Mechanisms in Digital Platform Ecosystems: Addressing the Generativity-Control Tension	N. Staub; K. Haki; S. Aier; R. Winter	2022	x	x	x	x	
Acquisition of Complementors as a Strategy for Evolving Digital Platform Ecosystems	N. Staub; K. Haki; S. Aier; R. Winter; A. Magan	2021	x				
How Digital Platforms Compete Against Diverse Rivals	K. Staykova; J. Damsgaard	2021	x	x			
Explaining Persistent Ineffectiveness in Professional Online Communities: Multilevel Tensions and Misguided Coping Strategies	M.-K. Stein; E. T. Lim; C.-W. Tan	2022	x				
Coping with Information Technology: Mixed Emotions, Vacillation, and Nonconforming Use Patterns	M.-K. Stein; S. Newell; E. L. Wagner; R. D. Galliers	2015	x				
Digital Entrepreneurship: What is New if Anything?	D. M. Steininger; K. Brohman; J. H. Block	2022	x				
Quantified Vehicles - Novel Services for Vehicle Lifecycle Data	A. Stocker; C. Kaiser; M. Fellmann	2017	x				
Five Inspiring Course (Re-)Designs - Examples of Innovations in Teaching and Learning BI&E	S. Strecker	2019	x				
Artificial Intelligence for Sign Language Translation – A Design Science Research Study	G. Strobel; T. Schoormann; L. Banh; F. Möller	2023	x				
Blockchain Regulations and Decentralized Applications: Panel Report from AMCIS 2018	H. Subramanian; K. Cousins; L. B. Bouyad; A. Sheth; D. Conway	2020	x				
Internet of Things and Competitive Advantage: A Dynamic Capabilities Perspective	Y. Sullivan; S. Fosso Wamba; M. Dunaway	2023	x				
Mapping the Landscape of Health Information Exchange (HIE) Networks in the United States	Z. Sun; D. Compeau; M. Carter	2021	x				
The Making of a Good Impression: Information Hiding in Ad Exchanges	Z. Sun; M. Dawande; G. Janakiraman; V. M. Mookerjee	2016	x				
Token Economy	A. Sunyaev; N. Kannengießer; R. Beck; H. Treiblmaier; M. Lacity; J. Kranz; G. Fridgen; U. Spankowski; A. Luckow	2021	x				
Embracing Digital Innovation in Incumbent Firms: How Volvo Cars Managed Competing Concerns	F. Svahn; L. Mathiassen; R. Lindgren	2017	x				
Role of Social Media on Information Accessibility	F. P. Tajudeen; N. I. Jaafar; A. Sulaiman	2016	x				
The Role of IS Capabilities in the Development of Multi-Sided Platforms: The Digital Ecosystem Strategy of Alibaba.com	B. Tan; S. L. Pan; X. Lu; L. Huang	2015	x				
Delivering Disruption in an Emergent Access Economy: A Case Study of an E-hailing Platform	F. T. C. Tan; B. Tan; A. Lu; L. Land	2017	x				
Developing a Leading Digital Multi-sided Platform: Examining IT Affordances and Competitive Actions in Alibaba.com	F. T. C. Tan; B. Tan; S. L. Pan	2016	x	x			
Empowerment of Grassroots Consumers: A Revelatory Case of a Chinese Fintech Innovation	T. Tan; Y. Zhang; C. S. Heng; C. Ge	2021	x				
Hospital Leadership in Support of Digital Transformation	M. Tanniru; J. Khuntia; J. Weiner	2018	x				
How Should Mydbots Manage Innovations in Consumer Robotics?	P. Tarafdar; I. Bose	2021	x				
Business Model Management: Current Practices, Required Activities and IT Support	N. Terrenghi; J. Schwarz; C. Legner; U. Eisert	2017	x				
The Economics of Multi-Hop Ride Sharing - Creating New Mobility Networks Through IS	T. Teubner; C. M. Flath	2015	x				
Reputation Transfer	T. Teubner; F. Hawlitschek; M. T. P. Adam	2019	x				
Analyzing Crisis Response through Actor-network Theory: The Case of Kathmandu Living Labs	D. Thapa; N. R. Budhathoki; B. E. Munkvold	2017	x				

Global Crises and the Role of BISE	O. Thomas; S. Hagen; U. Frank; J. Recker; L. Wessel; F. Kammler; N. Zarvic; I. Timm	2021	x				
The Contingent Role of Centrality in IT Network Governance: An Empirical Examination	S. Trang; S. Zander; L. M. Kolbe	2015	x				
Data is the New Protein: How the Commonwealth of Virginia Built Digital Resilience Muscle and Rebounded from Opioid and COVID Shocks	M. C. Tremblay; R. Kohli; C. Rivero	2023	x				
A Framework for Intelligent Collaborative Enterprise Systems. Concepts, opportunities and challenges	B. Unhelkar; A. A. Arntzen	2020	x				
Towards the Composition of Services by End-Users	P. Valderas; V. Torres; V. Pelechano	2021	x				
Understanding the Impact of Enterprise Architecture Driven Dynamic Capabilities on Agility: A Variance and fsQCA Study	R. van de Wetering	2021	x				
Improvisational and Dynamic Capabilities as Drivers of Business Model Innovation: An Enterprise Architecture Perspective	R. van de Wetering; B. Roelens; F. de Langen	2023	x				
Big Digital Platforms	W. M. P. van der Aalst; O. Hinz; C. Weinhardt	2019	x	x			
Managing Organizational Explosions During Digital Business Transformations	N. van der Meulen; P. Weill; S. L. Woerner	2020	x				
Business Models - An Information Systems Research Agenda	D. Veit; E. Clemons; A. Benlian; P. Buxmann; T. Hess; M. Spann; D. Kundisch; J. M. Leimeister; P. Loos	2014	x				
Design and Governance of mHealth Data Sharing	A. Vesselkov; H. Hämmäinen; J. Töyli	2019	x	x	x	x	
A Five-Level Framework for Research on Process Mining	J. vom Brocke; M. Jans; J. Mendling; H. A. Reijers	2021	x				
Connecting the Parts with the Whole: Toward an Information Ecology Theory of Digital Innovation Ecosystems	P. Wang	2021	x				
How the AIS can Improve its Contributions to the UN's Sustainability Development Goals: Towards A Framework for Scaling Collaborations and Evaluating Impact	R. T. Watson; S. Elliot; J. Corbett; D. Farkas; A. Feizabadi; A. Gupta; L. Iyer; S. Sen; R. Sharda; N. Shin; D. Thapa; J. Webster	2021	x				
AI Startup Business Models	M. Weber; M. Beutter; J. Weking; M. Boehm; H. Krcmar	2022	x				
Combining Crowd and Machine Intelligence to Detect False News on Social Media	X. Wei; Z. Zhang; M. Zhang; W. Chen; D. D. Zeng	2022	x				
Citizen Science in Information Systems Research	C. Weinhardt; S. Kloker; O. Hinz; W. M. P. van der Aalst	2021	x				
Esport	K. Werder	2022	x				
Data Portability on the Internet	M. Wohlfarth	2019	x				
Design Capital and Design Moves: The Logic of Digital Business Strategy	C. J. Woodard; N. Ramasubbu; F. T. Tschang; V. Sambamurthy	2013	x				
The Scholarly Divide: Insights from the AIS Well-being Project	R. T. Wright; C. Saunders; S. Sarker; A. Kankanhalli; E. A. Whitley; V. K. Tuunainen; H. Z. Henriksen	2023	x				
Using a Digital Services Capability Model to Assess Readiness for the Digital Consumer	J. Wulf; T. Mettler; W. Brenner	2017	x				
"Computing" Requirements for Open Source Software: A Distributed Cognitive Approach	X. Xiao; A. Lindberg; S. Hansen; K. Lyytinen	2018	x				
Role of Social Media in Technology Adoption for Sustainable Agriculture Practices: Evidence from Twitter Analytics	J. Yadav; A. Yadav; M. Misra; N. P. Rana; J. Zhou	2023	x				
Special Issue Editorial: Information Systems Research in the Age of Smart Services	Y.-C. Yang; H. Ying; Y. Jin; H. K. Cheng; T.-P. Liang	2021	x				
Problems, Solutions, and Success Factors in the openMDM User-Led Open Source Consortium	E. Y. Yavuz; A. Barcomb; D. Riehle	2022	x				

Management, Governance, and Value Creation in a Blockchain Consortium	L. Zavolokina; R. Ziolkowski; I. Bauer; G. Schwabe	2020	x				
Understanding Characteristics of Popular Streamers on Live Streaming Platforms: Evidence from Twitch.tv	K. Zhao; Y. Hu; Y. Hong; C. Westland	2021	x				
Intelligence Augmentation: Towards Building Human-Machine Symbiotic Relationship	L. Zhou; S. Paul; H. Demirkan; L. Yuan; J. Spohrer; M. Zhou; J. Basu	2021	x				
Responsible Artificial Intelligence Systems Critical considerations for business model design	M. P. Zimmer; M. Minkkinen; M. Mäntymäki	2022	x				
When IT Evolves Beyond Community Needs: Coevolution of Bottom-Up IT Innovation and Communities	A. Zorina; S. Karanasios	2021	x				
Internet with Transient Destination-Controlled Addressing	Z. Al-Qudah; E. Johnson; M. Rabinovich; O. Spatscheck	2016	x				
Pull Request Governance in Open Source Communities	A. Alami; R. Pardo; M. L. Cohn; A. Wasowski	2022	x				
Adaptive Context Dissemination in Heterogeneous Environments	D. Balakrishnan; A. Nayak	2014	x				
The Effectiveness of Reputation-Based Voting for Collusion Tolerance in Large-Scale Grids	A. Bendahmane; M. Essaaidi; A. E. Moussaoui; A. Younes	2015	x				
A Review on Explainable Artificial Intelligence for Healthcare: Why, How, and When?	S. Bharati; M. R. H. Mondal; P. Podder	5555	x				
Neuroticism, Extraversion, Conscientiousness and Stress: Physiological Correlates	A.-M. Brouwer; M. G. v. Schaik; J. E. Korteling; J. B. F. v. Erp; A. Toet	2015	x				
VRIA: A Web-Based Framework for Creating Immersive Analytics Experiences	P. W. S. Butcher; N. W. John; P. D. Ritsos	2021	x				
MagSign: Harnessing Dynamic Magnetism for User Authentication on IoT Devices	H. Cao; D. Liu; H. Jiang; J. Luo	5555	x				
ParticipAct: A Large-Scale Crowdsensing Platform	G. Cardone; A. Corradi; L. Foschini; R. Ianniello	2016	x				
Automatic Prediction of Impressions in Time and across Varying Context: Personality, Attractiveness and Likeability	O. Celiktutan; H. Gunes	2017	x				
CrowdChart: Crowdsourced Data Extraction From Visualization Charts	C. Chai; G. Li; J. Fan; Y. Luo	2021	x				
Human Effects of Enhanced Privacy Management Models	G. P. Cheek; M. Shehab	2014	x				
BOSSA: A Decentralized System for Proofs of Data Retrievability and Replication	D. Chen; H. Yuan; S. Hu; Q. Wang; C. Wang	2021	x				
Goal-Driven Service Composition in Mobile and Pervasive Computing	N. Chen; N. Cardozo; S. Clarke	2018	x				
A Lightweight Virtualization Solution for Android Devices	W. Chen; L. Xu; G. Li; Y. Xiang	2015	x				
Detection and Localization of Multiple Spoofing Attackers in Wireless Networks	Y. Chen; J. Cheng; W. Trappe; J. Yang	2013	x				
Sentiment Analysis: From Opinion Mining to Human-Agent Interaction	C. Clavel; Z. Callejas	2016	x				
Personality- and Value-Aware Scheduling of User Requests in Cloud for Profit Maximization	P. Cong; G. Xu; J. Zhou; M. Chen; T. Wei; M. Qiu	2022	x				
Personality-Guided Cloud Pricing via Reinforcement Learning	P. Cong; J. Zhou; M. Chen; T. Wei	2022	x				
Mobile Cloud Support for Semantic-Enriched Speech Recognition in Social Care	A. Corradi; M. Destro; L. Foschini; S. Kotoulas; V. Lopez; R. Montanari	2019	x				
Touch and Beyond: Comparing Physical and Virtual Reality Visualizations	K. Danyluk; T. Ulusoy; W. Wei; W. Willett	2022	x				
Enabling Secure and Efficient Decentralized Storage Auditing With Blockchain	Y. Du; H. Duan; A. Zhou; C. Wang; M. H. Au; Q. Wang	2022	x				
Towards Practical Auditing of Dynamic Data in Decentralized Storage	H. Duan; Y. Du; L. Zheng; C. Wang; M. H. Au; Q. Wang	2023	x				
Secure Automated Forensic Investigation for Sustainable Critical Infrastructures Compliant with Green Computing Requirements	M. Elhoseny; H. Abbas; A. E. Hassanien; K. Muhammad; A. K. Sangaiah	2020	x				

Modeling, Recognizing, and Explaining Apparent Personality From Videos	H. J. Escalante; H. Kaya; A. A. Salah; S. Escalera; Y. Gucluturk; U. Guclu; X. Baro; I. Guyon; J. C. S. J. Junior; M. Madadi; S. Ayache; E. Viegas; F. Gurpnar; A. S. Wicaksana; C. C. S. Liem; M. A. J. v. Gerven; R. v. Lier	2022	x				
A Model for an Affective Non-Expensive Utility-Based Decision Agent	P. G. Esteban; D. R. Insua	2019	x				
On Cyber Risk Management of Blockchain Networks: A Game Theoretic Approach	S. Feng; W. Wang; Z. Xiong; D. Niyato; P. Wang; S. S. Wang	2021	x				
IC9600: A Benchmark Dataset for Automatic Image Complexity Assessment	T. Feng; Y. Zhai; J. Yang; J. Liang; D.-P. Fan; J. Zhang; L. Shao; D. Tao	5555	x				
Behavior-based Ethical Understanding in Chinese Social News	X. Feng; T. Gu; X. Bao; L. Li	5555	x				
Regulatable and Hardware-Based Proof of Stake to Approach Nothing At Stake and Long Range Attacks	X. Feng; J. Ma; Y. Miao; X. Liu; K.-K. R. Choo	5555	x				
Variability in Software Systems—A Systematic Literature Review	M. Galster; D. Weyns; D. Tofan; B. Michalik; P. Avgeriou	2014	x				
TrustWorker: A Trustworthy and Privacy-Preserving Worker Selection Scheme for Blockchain-Based Crowdsensing	S. Gao; X. Chen; J. Zhu; X. Dong; J. Ma	2022	x				
Quality Inference Based Task Assignment in Mobile Crowdsensing	X. Gao; H. Huang; C. Liu; F. Wu; G. Chen	2021	x				
Modeling Service Level Agreements with Linked USDL Agreement	J. M. Garcia; P. Fernandez; C. Pedrinaci; M. Resinas; J. Cardoso; A. Ruiz-Cortes	2017	x				
Effective Risk Communication for Android Apps	C. S. Gates; J. Chen; N. Li; R. W. Proctor	2014	x				
Leveraging Android Automated Testing to Assist Crowdsourced Testing	X. Ge; S. Yu; C. Fang; Q. Zhu; Z. Zhao	2023	x				
Improving Humanness of Virtual Agents and Users' Cooperation through Emotions	M. Ghafurian; N. Budnarain; J. Hoey	5555	x				
Behavior Trees and State Machines in Robotics Applications	R. Ghzouli; T. Berger; E. B. Johnsen; A. Wasowski; S. Dragule	5555	x				
Decoding the Perception of Sincerity in Dialogues	C. Girlea; R. Girju	2021	x				
Creative User-Centered Visualization Design for Energy Analysts and Modelers	S. Goodwin; J. Dykes; S. Jones; I. Dillingham; G. Dove; A. Duffy; A. Kachkaev; A. Slingsby; J. Wood	2013	x				
Real-time Multi-map Saliency-driven Gaze Behavior for Non-conversational Characters	I. Goude; A. Bruckert; A.-H. Olivier; J. Pettre; R. Cozot; K. Bouatouch; M. Christie; L. Hoyet	5555	x				
BlockMaze: An Efficient Privacy-Preserving Account-Model Blockchain Based on zk-SNARKs	Z. Guan; Z. Wan; Y. Yang; Y. Zhou; B. Huang	2022	x				
Multimodal First Impression Analysis with Deep Residual Networks	Y. Gucluturk; U. Guclu; X. Baro; H. J. Escalante; I. Guyon; S. Escalera; M. A. J. v. Gerven; R. v. Lier	2018	x				
FlierMeet: A Mobile Crowdsensing System for Cross-Space Public Information Reposting, Tagging, and Sharing	B. Guo; H. Chen; Z. Yu; X. Xie; S. Huangfu; D. Zhang	2015	x				
Identifying Incentives for the Enforcement of Artificial Intelligence Soft Law Programs	C. I. Gutierrez	5555	x				
Design by Immersion: A Transdisciplinary Approach to Problem-Driven Visualizations	K. W. Hall; A. J. Bradley; U. Hinrichs; S. Huron; J. Wood; C. Collins; S. Carpendale	2020	x				
HisVA: A Visual Analytics System for Studying History	D. Han; G. Parsad; H. Kim; J. Shim; O.-S. Kwon; K. A. Son; J. Lee; I. Cho; S. Ko	2022	x				
Collaborative Policy Administration	W. Han; Z. Fang; L. T. Yang; G. Pan; Z. Wu	2014	x				
Cost-Effective Authentic and Anonymous Data Sharing with Forward Security	X. Huang; J. K. Liu; S. Tang; Y. Xiang; K. Liang; L. Xu; J. Zhou	2015	x				
Open Source Intelligence for Malicious Behavior Discovery and Interpretation	Y.-T. Huang; C. Y. Lin; Y.-R. Guo; K.-C. Lo; Y. S. Sun; M. C. Chen	2022	x				

Secure Seamless Bluetooth Low Energy Connection Migration for Unmodified IoT Devices	S. R. Hussain; S. Mehnaz; S. Nirjon; E. Bertino	2018	x			
Human Values in Software Engineering: Contrasting Case Studies of Practice	W. Hussain; H. Perera; J. Whittle; A. Nurwidyantoro; R. Hoda; R. A. Shams; G. Oliver	2022	x			
Batch Allocation for Tasks with Overlapping Skill Requirements in Crowdsourcing	J. Jiang; B. An; Y. Jiang; P. Shi; Z. Bu; J. Cao	2019	x			
A Survey of Task Allocation and Load Balancing in Distributed Systems	Y. Jiang	2016	x			
First Impressions: A Survey on Vision-Based Apparent Personality Trait Analysis	J. C. S. J. Junior; Y. Gucluturk; M. Perez; U. Guclu; C. Andujar; X. Baro; H. J. Escalante; I. Guyon; M. A. J. v. Gerven; R. v. Lier; S. Escalera	2022	x			
An Expressive Virtual Audience with Flexible Behavioral Styles	N. Kang; W.-P. Brinkman; M. B. v. Riemsdijk; M. A. Neerincx	2013	x			
Adaptive Social Learning Based on Crowdsourcing	E. Karataev; V. Zadorozhny	2017	x			
Rapid Development of a Data Visualization Service in an Emergency Response	S. Khan; P. H. Nguyen; A. Abdul-Rahman; E. Freeman; C. Turkay; M. Chen	2022	x			
Drumming in Immersive Virtual Reality: The Body Shapes the Way We Play	K. Kilteni; I. Bergstrom; M. Slater	2013	x			
Affective Body Expression Perception and Recognition: A Survey	A. Kleinsmith; N. Bianchi-Berthouze	2013	x			
Making Digital Artifacts on the Web Verifiable and Reliable	T. Kuhn; M. Dumontier	2015	x			
Programmable and Customized Intelligence for Traffic Steering in 5G Networks Using Open RAN Architectures	A. Lacava; M. Polese; R. Sivaraj; R. Soundrarajan; B. S. Bhati; T. Singh; T. Zugno; F. Cuomo; T. Melodia	5555	x			
Blockchain Security: A Survey of Techniques and Research Directions	J. Leng; M. Zhou; J. L. Zhao; Y. Huang; Y. Bian	2022	x			
ACSEE: Antagonistic Crowd Simulation Model With Emotional Contagion and Evolutionary Game Theory	C. Li; P. Lv; D. Manocha; H. Wang; Y. Li; B. Zhou; M. Xu	2022	x			
Blockchain-Based Secure Key Management for Mobile Edge Computing	J. Li; J. Wu; L. Chen; J. Li; S.-K. Lam	2023	x			
A Misreport- and Collusion-Proof Crowdsourcing Mechanism Without Quality Verification	K. Li; S. Wang; X. Cheng; Q. Hu	2022	x			
Quantitative Personality Predictions From a Brief EEG Recording	W. Li; C. Wu; X. Hu; J. Chen; S. Fu; F. Wang; D. Zhang	2022	x			
Efficient and Secure Deep Learning Inference in Trusted Processor Enabled Edge Clouds	Y. Li; D. Zeng; L. Gu; Q. Chen; S. Guo; A. Zomaya; M. Guo	2022	x			
Disguised as Privacy: Data Poisoning Attacks against Differentially Private Crowdsensing Systems	Z. Li; Z. Zheng; S. Guo; B. Guo; F. Xiao; K. Ren	5555	x			
LiveRender: A cloud gaming system based on compressed graphics streaming	X. Liao; L. Lin; G. Tan; H. Jin; X. Yang; W. Zhang; B. Li	2016	x			
Privacy Leakage in Wireless Charging	J. Liu; X. Zou; L. Zhao; Y. Tao; S. Hu; J. Han; K. Ren	5555	x			
Privacy-Preserving Reputation Management for Edge Computing Enhanced Mobile Crowdsensing	L. Ma; X. Liu; Q. Pei; Y. Xiang	2019	x			
The public option: a nonregulatory alternative to network neutrality	R. T. B. Ma; V. Misra	2013	x			
On the Interplay Between Individuals' Evolving Interaction Patterns and Traits in Dynamic Multiplex Social Networks	L. Meng; Y. Hulovatyy; A. Striegel; T. Milenkovic	2016	x			
Exploring the Use of Chatrooms by Developers: An Empirical Study on Slack and Gitter	M. E. Mezouar; D. A. d. Costa; D. M. German; Y. Zou	2022	x			
Towards Query Pricing on Incomplete Data	X. Miao; Y. Gao; L. Chen; H. Peng; J. Yin; Q. Li	2022	x			
Sticky Policies: A Survey	D. Miorandi; A. Rizzardi; S. Sicari; A. Coen-Porisini	2020	x			

A Multi-Componential Approach to Emotion Recognition and the Effect of Personality	G. Mohammadi; P. Vuilleumier	2022	x			
A Generic Query Model for the Unified Discovery of Heterogeneous Services	M. Pantazoglou; A. Tsalgatidou	2013	x			
The e-LocGov Model for Introducing e-Governance into Local Governments: An Estonian Case Study	I. Pappel; V. Tsap; D. Draheim	2021	x			
Choosing Component Origins for Software Intensive Systems: In-House, COTS, OSS or Outsourcing?—A Case Survey	K. Petersen; D. Badampudi; S. M. A. Shah; K. Wnuk; T. Gorschek; E. Papatheocharous; J. Axelsson; S. Sentilles; I. Crnkovic; A. Cicchetti	2018	x			
Understanding Social OER Environments—A Quantitative Study on Factors Influencing the Motivation to Share and Collaborate	H. Pirkkalainen; J. P. P. Jokinen; J. M. Pawlowski	2014	x			
Blockchain Mutability: Challenges and Proposed Solutions	E. Politou; F. Casino; E. Alepis; C. Patsakis	2021	x			
Framework for Prioritization of Open Data Publication: An Application to Smart Cities	A. E. PRIETO; J.-N. Mazon; A. Lozano-Tello	2021	x	x		
CloudCFI: Context-Sensitive and Incremental CFI in the Cloud Environment	W. Qiang; Y. Huang; H. Jin; L. T. Yang; D. Zou	2021	x			
Proof of Federated Learning: A Novel Energy-Recycling Consensus Algorithm	X. Qu; S. Wang; Q. Hu; X. Cheng	2021	x			
Human-Centric Computing	J. M. Rabaey	2020	x			
Detecting Malicious Facebook Applications	S. Rahman; T.-K. Huang; H. V. Madhyastha; M. Faloutsos	2016	x			
Inner Source in Platform-Based Product Engineering	D. Riehle; M. Capraro; D. Kips; L. Horn	2016	x			
Fake Profile Detection on Social Networking Websites: A Comprehensive Review	P. K. Roy; S. Chahar	2020	x			
MADAM: Effective and Efficient Behavior-based Android Malware Detection and Prevention	A. Saracino; D. Sgandurra; G. Dini; F. Martinelli	2018	x			
The Pictures We Like Are Our Image: Continuous Mapping of Favorite Pictures into Self-Assessed and Attributed Personality Traits	C. Segalin; A. Perina; M. Cristani; A. Vinciarelli	2017	x			
Fee-Free Pooled Mining for Countering Pool-Hopping Attack in Blockchain	H. Shi; S. Wang; Q. Hu; X. Cheng; J. Zhang; J. Yu	2021	x			
VM Scaling and Load Balancing via Cost Optimal MDP Solution	M. Shifrin; R. Mitrany; E. Biton; O. Gurewitz	2022	x			
Affect and Social Processes in Online Communication—Experiments with an Affective Dialog System	M. Skowron; M. Theunis; S. Rank; A. Kappas	2013	x			
SDCon: Integrated Control Platform for Software-Defined Clouds	J. Son; R. Buyya	2019	x			
Learning Person-specific Cognition from Facial Reactions for Automatic Personality Recognition	S. Song; Z. Shao; S. Jaiswal; L. Shen; M. Valstar; H. Gunes	5555	x			
Psychophysiological Reactions to Persuasive Messages Deploying Persuasion Principles	H. A. A. Spelt; J. H. D. M. Westerink; J. Ham; W. A. IJsselsteijn	2022	x			
MLaaS: A Cloud-Based System for Delivering Adaptive Micro Learning in Mobile MOOC Learning	G. Sun; T. Cui; J. Yong; J. Shen; S. Chen	2018	x			
An Efficient and Secure Trading Framework for Shared Charging Service based on Multiple Consortium Blockchains	Z. Sun; P. Zhao; C. Wang; X. Zhang; H. Cheng	5555	x			
Flexible and Efficient Authenticated Key Agreement Scheme for BANs Based on Physiological Features	W. Tang; K. Zhang; J. Ren; Y. Zhang; X. Shen	2019	x			
Understanding How and Why Developers Seek and Analyze API-Related Opinions	G. Uddin; O. Baysal; L. Guerrouj; F. Khomh	2021	x			
A Psychologically Inspired Fuzzy Cognitive Deep Learning Framework to Predict Crowd Behavior	E. B. Varghese; S. M. Thampi; S. Berretti	2022	x			
A Survey of Personality Computing	A. Vinciarelli; G. Mohammadi	2014	x			

Blockchain-as-a-Service for Business Process Management: Survey and Challenges	W. Viriyasitavat; L. D. Xu; G. Dhiman; Z. Bi	5555	x			
Service Workflow: State-of-the-Art and Future Trends	W. Viriyasitavat; L. D. Xu; G. Dhiman; A. Sapsomboon; V. Pungpapong; Z. Bi	2023	x			
A Survey on Energy Internet Communications for Sustainability	K. Wang; X. Hu; H. Li; P. Li; D. Zeng; S. Guo	2017	x			
Systematic Prevention of On-Core Timing Channels by Full Temporal Partitioning	N. Wistoff; M. Schneider; F. K. Gurkaynak; G. Heiser; L. Benini	2023	x			
High-Performance Capabilities for 1-Hop Containment of Network Attacks	T. Wolf; S. Natarajan; K. T. Vasudevan	2013	x			
Current Directions in Personality Science and the Potential for Advances through Computing	A. G. C. Wright	2014	x			
Enabling Data Trustworthiness and User Privacy in Mobile Crowdsensing	H. Wu; L. Wang; G. Xue; J. Tang; D. Yang	2019	x	x		
FCSS: Fog-Computing-based Content-Aware Filtering for Security Services in Information-Centric Social Networks	J. Wu; M. Dong; K. Ota; J. Li; Z. Guan	2019	x			
Design and Performance Evaluation of Overhearing-Aided Data Caching in Wireless Ad Hoc Networks	W. Wu; J. Cao; X. Fan	2013	x			
Eliciting Joint Truthful Answers and Profiles from Strategic Workers in Mobile Crowdsourcing Systems	M. Xiao; W. Jin; C. Li; M. Li	5555	x			
QLDS: A Novel Design Scheme for Trajectory Privacy Protection with Utility Guarantee in Participatory Sensing	Z. Xiao; J.-J. Yang; M. Huang; L. Ponnambalam; X. Fu; R. S. M. Goh	2018	x			
My Privacy My Decision: Control of Photo Sharing on Online Social Networks	K. Xu; Y. Guo; L. Guo; Y. Fang; X. Li	2017	x	x		
Blockchain-Based Transparency Framework for Privacy Preserving Third-Party Services	R. Xu; C. Li; J. Joshi	2023	x	x		
Incentive Mechanism for Spatial Crowdsourcing with Unknown Social-Aware Workers: A Three-Stage Stackelberg Game Approach	Y. Xu; M. Xiao; J. Wu; S. Zhang; G. Gao	5555	x			
MU-TEIR: Traceable Encrypted Image Retrieval in the Multi-User Setting	T. Yang; J. Ma; Y. Miao; Y. Wang; X. Liu; K.-K. R. Choo; B. Xiao	2023	x			
A Reputation-Based Mechanism for Transaction Processing in Blockchain Systems	J. Zhang; Y. Cheng; X. Deng; B. Wang; J. Xie; Y. Yang; M. Zhang	2022	x	x		
Design and Implementation of Efficient Integrity Protection for Open Mobile Platforms	X. Zhang; J.-P. Seifert; O. Acicmez	2014	x			
Finding Critical Scenarios for Automated Driving Systems: A Systematic Mapping Study	X. Zhang; J. Tao; K. Tan; M. Torngren; J. M. G. Sanchez; M. R. Ramli; X. Tao; M. Gyllenhammar; F. Wotawa; N. Mohan; M. Nica; H. Felbinger	2023	x			
Pull Request Decisions Explained: An Empirical Overview	X. Zhang; Y. Yu; G. Gousios; A. Rastogi	2023	x			
Holistic Affect Recognition Using PaNDA: Paralinguistic Non-Metric Dimensional Analysis	Y. Zhang; F. Weninger; B. Schuller; R. W. Picard	2022	x			
MOSES: Supporting and Enforcing Security Profiles on Smartphones	Y. Zhauniarovich; G. Russello; M. Conti; B. Crispo; E. Fernandes	2014	x			
From RBAC to ABAC: Constructing Flexible Data Access Control for Cloud Storage Services	Y. Zhu; D. Huang; C.-J. Hu; X. Wang	2015	x			
A Proof-of-Trust Consensus Protocol for Enhancing Accountability in Crowdsourcing Services	J. Zou; B. Ye; L. Qu; Y. Wang; M. A. Orgun; L. Li	2019	x			
Citizen-centric data services for smarter cities	U. Aguilera; O. Peña; O. Belmonte; D. López-de-Ipiña	2017	x			
Thermodynamically-consistent derivation and computation of twinning and fracture in brittle materials by means of phase-field approaches in the finite element method	B. Amirian; H. Jafarzadeh; B. E. Abali; A. Reali; J. D. Hogan	2022	x			
A data model for collaborative manufacturing environments	B. Andres; R. Poler; R. Sanchis	2021	x			

Discrete Global Grid Systems with quadrangular cells as reference frameworks for the current generation of Earth observation data cubes	R. Béjar; J. Lacasta; F. J. Lopez-Pellicer; J. Noguera-Iso	2023	x			
Model-driven development for the sel4 microkernel using the HAMR framework	J. Belt; J. Hatcliff; Robby; J. Shackleton; J. Carciofini; T. Carpenter; E. Mercer; I. Amundson; J. Babar; D. Cofer; D. Hardin; K. Hoech; K. Slind; I. Kuz; K. McLeod	2023	x			
Framework for collaborative intelligence in forecasting day-ahead electricity price	S. Beltrán; A. Castro; I. Irizar; G. Naveran; I. Yeregui	2022	x			
Cloud-WBAN: An experimental framework for Cloud-enabled Wireless Body Area Network with efficient virtual resource utilization	T. Bhardwaj; S. C. Sharma	2018	x			
Untangling blockchain technology: A survey on state of the art, security threats, privacy services, applications and future research directions	B. Bhushan; P. Sinha; K. M. Sagayam; A. J	2021	x			
Improving Mandatory Access Control for HPC clusters	M. Blanc; J. F. Lalande	2013	x			
Cultivating open government data platform ecosystems through governance: Lessons from Buenos Aires, Mexico City and Montevideo	C. Bonina; B. Eaton	2020	x	x	x	x
Digital platform openness: Drivers, dimensions and outcomes	T. L. J. Broekhuizen; O. Emrich; M. J. Gijzenberg; M. Broekhuis; B. Donkers; L. M. Sloot	2021	x	x	x	x
Coherence or flexibility? The paradox of change for developers' digital innovation trajectory on open platforms	S. Brunswicker; A. Schecter	2019	x			
Electrosense+: Crowdsourcing radio spectrum decoding using IoT receivers	R. Calvo-Palomino; H. Cordobés; M. Engel; M. Fuchs; P. Jain; M. Liechti; S. Rajendran; M. Schäfer; B. V. d. Bergh; S. Pollin; D. Giustiniano; V. Lenders	2020	x			
WinGPYRO: A software platform for kinetic study of forest fuels	D. Cancellieri; E. Innocenti; V. Leroy-Cancellieri	2013	x			
A gray-box performance model for Apache Spark	Z. Chao; S. Shi; H. Gao; J. Luo; H. Wang	2018	x			
A new two-server authentication and key agreement protocol for accessing secure cloud services	D. Chattaraj; M. Sarma; A. K. Das	2018	x			
Securing Spatial Data Infrastructures for Distributed Smart City applications and services	K. Chaturvedi; A. Matheus; S. H. Nguyen; T. H. Kolbe	2019	x			
A novel robust prediction algorithm based on REMD-MWNN for AIOps	L. Chen; W. Wang; Y. Yang; Y. Xu	2021	x			
Modeling methods of self-centering energy dissipation braces using OpenSees	P. Chen; L. Xu; Z. Li	2022	x			
A three-dimensional DNA walker and silver nanoparticles promoting lattice-strain-driven photo-induced electron transfer for high-performance semi-homogeneous biosensing	Y. Chen; L. Xu; D. Qian; G. Sun; Y. Cao; Y. Zhang; Q. Xu; J. Li; H. Li	2023	x			
The impacts of technology platform openness on application developers' intention to continuously use a platform: From an ecosystem perspective	G. Choi; C. Nam; S. Kim	2019	x	x		
A multiple-band perfect absorber for SEIRA applications	H. Durmaz; Y. Li; A. E. Cetin	2018	x			
Social capital and the digital crowd: Involving backers to promote new product innovativeness	N. Eiteneyer; D. Bendig; M. Brettel	2019	x			
Architecture for embedded open software ecosystems	U. Eklund; J. Bosch	2014	x			
Research on multi-objective decision-making under cloud platform based on quality function deployment and uncertain linguistic variables	J. Fan; S. Yu; J. Chu; D. Chen; M. Yu; T. Wu; J. Chen; F. Cheng; C. Zhao	2019	x			
Social interaction in MOOCs: The mediating effects of immersive experience and psychological needs satisfaction	J. Fang; L. Tang; J. Yang; M. Peng	2019	x			

Hierarchical dynamic containers for fusion data	D. Fridrich; J. Urban	2018	x			
Automated identification of circular value chains and synergies	L. Gentilini; C. Polidori; M. Fervorari; M. Colledani	2021	x			
NiftyNet: a deep-learning platform for medical imaging	E. Gibson; W. Li; C. Sudre; L. Fidon; D. I. Shikir; G. Wang; Z. Eaton-Rosen; R. Gray; T. Doel; Y. Hu; T. Whyntie; P. Nachev; M. Modat; D. C. Barratt; S. Ourselin; M. J. Cardoso; T. Vercauteren	2018	x			
Securing IIoT communications using OPC UA PubSub and Trusted Platform Modules	O. Gilles; D. Gracia Pérez; P. A. Brameret; V. Lacroix	2023	x			
Cloud vs Fog: assessment of alternative deployments for a latency-sensitive IoT application	M. Gomes; M. L. Pardal	2018	x			
A hybrid CPU/GPU approach for optimizing sorting throughput	M. Gowanlock; B. Karsin	2019	x			
Investigation on the flow around a submarine under the rudder deflection condition by using URANS and DDES methods	H. Guo; G. Li; L. Du	2023	x			
Micro-Analyzer: Automatic preprocessing of Affymetrix microarray data	P. H. Guzzi; M. Cannataro	2013	x			
Data fusion strategies for energy efficiency in buildings: Overview, challenges and novel orientations	Y. Himeur; A. Alsalemi; A. Al-Kababji; F. Bensaali; A. Amira	2020	x			
A model predictive control for attitude stabilization and spin control of a spacecraft with a flexible rotating payload	P. Iannelli; F. Angeletti; P. Gasbarri	2022	x			
A secure blockchain-oriented data delivery and collection scheme for 5G-enabled IoD environment	A. Irshad; S. A. Chaudhry; A. Ghani; M. Bilal	2021	x			
Improving blockchain performance in clinical trials using intelligent optimal transaction traffic control mechanism in smart healthcare applications	F. Jamil; S. Ahmad; T. K. Whangbo; A. Muthanna; D.-H. Kim	2022	x			
TZMon: Improving mobile game security with ARM trustzone	S. Jeon; H. K. Kim	2021	x			
Managing complexity across multiple dimensions of liquid open data: The case of the Danish Basic Data Program	T. Jetzek	2016	x			
A promising phenomenon of open data: A case study of the Chicago open data project	M. Kassen	2013	x	x		
Data sharing for business model innovation in platform ecosystems: From private data to public good	N. Kazantsev; N. Islam; J. Zwiendelaar; A. Brown; R. Maull	2023	x	x		
A lightweight cryptography (LWC) framework to secure memory heap in Internet of Things	M. Khalifa; F. Algarni; M. Ayoub Khan; A. Ullah; K. Aloufi	2021	x			
Incentive-based resource assignment and regulation for collaborative cloud services in community networks	A. M. Khan; Ü. C. Büyüksahin; F. Freitag	2015	x			
An Overview of Massive Open Online Course Platforms: Personalization and Semantic Web Technologies and Standards	B. Kiselev; V. Yakutenko	2020	x			
Development of an open-source software package for watershed modeling with the Hydrological Simulation Program in Fortran	D. J. Lampert; M. Wu	2015	x			
Self-presentation on Instagram and friendship development among young adults: A moderated mediation model of media richness, perceived functionality, and openness	D. K. L. Lee; P. Borah	2020	x			
Governance strategies for open collaboration: Focusing on resource allocation in open source software development organizations	S. Lee; H. Baek; J. Jahng	2017	x			

Programmable all-thermal encoding with metamaterials	M. Lei; C. Jiang; F. Yang; J. Wang; J. Huang	2023	x			
When services computing meets blockchain: Challenges and opportunities	X. Li; Z. Zheng; H.-N. Dai	2021	x			
Blockchain-based identity management and access control framework for open banking ecosystem	C.-H. Liao; X.-Q. Guan; J.-H. Cheng; S.-M. Yuan	2022	x			
Characterizing client-side caches of audiovisual content sharing services: Findings and suggestions for forensics	Y. Lim; M. Youn; H. Chung; J. Park; G. Horsman; S. Lee	2022	x			
A survey on blockchain-enabled federated learning and its prospects with digital twin	K. Liu; Z. Yan; X. Liang; R. Kantola; C. Hu	2022	x			
Adaptive multi-channel Bayesian graph attention network for IoT transaction security	Z. Liu; D. Yang; S. Wang; H. Su	2022	x			
Cloud agnostic Big Data platform focusing on scalability and cost-efficiency	R. Lovas; E. Nagy; J. Kovács	2018	x			
A blockchain based solution for the custody of digital files in forensic medicine	M. Lusetti; L. Salsi; A. Dallatana	2020	x			
Optimal crowdsourcing contracting for reconfigurable process planning in open manufacturing: A bilevel coordinated optimization approach	Y. Ma; G. Du; R. J. Jiao	2020	x			
Searching for ideas from creative Crowds: The role of examples in problem statements	A. Malhotra; C. Kubowicz Malhotra	2023	x			
Improving Content Privacy on Social Networks Using Open Digital Rights Management Solutions	J. Marques; C. Serrão	2013	x			
Detection and mitigation of field flooding attacks on oil and gas critical infrastructure communication	A. S. Mohammed; E. Anthi; O. Rana; N. Saxena; P. Burnap	2023	x			
Model-based Evaluation of Scalability and Security Tradeoffs: a Case Study on a Multi-Service Platform	L. Montecchi; N. Nostro; A. Ceccarelli; G. Vella; A. Caruso; A. Bondavalli	2015	x			
Renovating blockchain with distributed databases: An open source system	M. Muzammal; Q. Qu; B. Nasrulin	2019	x			
An efficient privacy-preserving control mechanism based on blockchain for E-health applications	H. Naser Alsuqaih; W. Hamdan; H. Elmessiry; H. Abulkasim	2023	x	x	x	
Development of a Laboratory Framework for Testing Simultaneous Localization and Mapping Approaches	P. Neduchal; M. Flidr	2016	x			
Modelling and Investigation on Bouncing Mechanism of a Sphere Robot	N. M. H. Norsahperi; M. A. Abdullah; S. Ahmad; S. F. Toha; I. A. Mahmood	2015	x			
Computer-aided design-while-engineering technology in top-down modeling of mechanical product	Z. Pan; X. Wang; R. Teng; X. Cao	2016	x			
Moving towards digital platforms revolution? Antecedents, determinants and conceptual framework for offline B2B networks	D. F. Peruchi; D. A. de Jesus Pacheco; B. V. Todeschini; C. S. ten Caten	2022	x			
Towards an Analysis of Data Accountability and Auditing for Secure Cloud Data Storage	J. Prassanna; K. Punitha; V. Neelananarayanan	2015	x			
Privacy-enhancing ETL-processes for biomedical data	F. Prasser; H. Spengler; R. Bild; J. Eicher; K. A. Kuhn	2019	x			
Simulation-based modeling of wild blueberry pollination	H. Qu; F. Drummond	2018	x			
Bump in the wire (BITW) security solution for a marine ROV remote control application	M. Rao; T. Newe; E. Omerdic; A. Kaknjo; W. Elgenaidi; A. Mathur; G. Dooly; E. Lewis; D. Toal	2018	x			
Gas diffusion and evaporation control using EWOD actuation of ionic liquid microdroplets for gas sensing applications	F. Ribet; L. De Pietro; N. Roxhed; G. Stemme	2018	x			
Extending τ -Lop to model concurrent MPI communications in multicore clusters	J.-A. Rico-Gallego; J.-C. Díaz-Martín; A. L. Lastovsky	2016	x			

A data-driven agent-based simulation to predict crime patterns in an urban environment	R. Rosés; C. Kadar; N. Malleon	2021	x				
Concepts and Solutions of the Digital Teams Platform to Support Mobile Work and Virtual Teams	S. Schweitzer; M. Gerbershagen; F. Elberzhager; S. Braun	2020	x				
A secure and quality-aware prototypical architecture for the Internet of Things	S. Sicari; A. Rizzardi; D. Miorandi; C. Cappiello; A. Coen-Porisini	2016	x				
An open-source OpenSim® ankle-foot musculoskeletal model for assessment of strains and forces in dense connective tissues	A. Sikidar; D. Kalyanasundaram	2022	x				
Controlled searching in reversibly de-identified medical imaging archives	J. M. Silva; E. Pinho; E. Monteiro; J. F. Silva; C. Costa	2018	x				
Type synthesis of 2-DoF rotational parallel mechanisms actuating the inter-satellite link antenna	Y. Song; Y. Qi; G. Dong; T. Sun	2016	x				
Task Scheduling in Big Data Platforms: A Systematic Literature Review	M. Soualhia; F. Khomh; S. Tahar	2017	x				
A novel method for actuator fault detection considering complex operating processes	T. Sun; Y. Wang; M. Ding	2022	x				
SymEx-VP: An open source virtual prototype for OS-agnostic concolic testing of IoT firmware	S. Tempel; V. Herdt; R. Drechsler	2022	x				
SIMON: Open-Source Knowledge Discovery Platform	A. Tomic; I. Tomic; L. Waldron; L. Geistlinger; M. Kuhn; R. L. Spreng; L. C. Dahora; K. E. Seaton; G. Tomaras; J. Hill; N. A. Duggal; R. D. Pollock; N. R. Lazarus; S. D. R. Harridge; J. M. Lord; P. Khatri; A. J. Pollard; M. M. Davis	2021	x				
Digital maker-entrepreneurs in open design: What activities make up their business model?	P. Troxler; P. Wolf	2017	x				
A blockchain-based trust system for decentralised applications: When trustless needs trust	N. Truong; G. M. Lee; K. Sun; F. Guitton; Y. Guo	2021	x				
Efficient parameter tuning of neural foundation models for drug perspective prediction from unstructured socio-medical data	R. Unnikrishnan; S. K. S; A. V.S	2023	x				
LUCE: A blockchain-based data sharing platform for monitoring data License accountability and Compliance	V. Urovi; V. Jaiman; A. Angerer; M. Dumontier	2022	x	x	x	x	x
Distributed Trust & Reputation Models using Blockchain Technologies for Tourism Crowdsourcing Platforms	B. Veloso; F. Leal; B. Malheiro; F. Moreira	2019	x	x			
A comparison of balance-correcting responses induced with platform-translation and shoulder-pull perturbation methods	D. Verniba; W. H. Gage	2020	x				
The dawn of an open exploration era: Emergent principles and practices of open science and innovation of university research teams in a digital world	R. Vicente-Saez; R. Gustafsson; L. Van den Brande	2020	x				
Design of peer-to-peer protocol with sensible and secure IoT communication for future internet architecture	V. Vijaya Kumar; M. Devi; P. Vishnu Raja; P. Kanmani; V. Priya; S. Sudhakar; K. Sujatha	2020	x				
Obscuring users' identity in VoIP/IMS environments	N. Vrakas; D. Geneiatakis; C. Lambrinoudakis	2014	x				
SimEdgeIntel: A open-source simulation platform for resource management in edge intelligence	C. Wang; R. Li; W. Li; C. Qiu; X. Wang	2021	x				
A mobile malware detection method using behavior features in network traffic	S. Wang; Z. Chen; Q. Yan; B. Yang; L. Peng; Z. Jia	2019	x				
Harvest shopping advice: Neural Question Generation from multiple information sources in E-commerce	Y. Wang; K. Song; L. Bing; X. Liu	2021	x				
A secure visual framework for multi-index protection evaluation in networks	X. Wu; H. Wang; Y. Zhang; R. Li	2023	x				

Data Replica Placement Mechanism for Open Heterogeneous Storage Systems	X. Xu; C. Yang; J. Shao	2017	x				
Keyhole pores reduction in laser powder bed fusion additive manufacturing of nickel alloy 625	H. Yeung; F. H. Kim; M. A. Donmez; J. Neira	2022	x				
Foldable-circuit-enabled miniaturized multifunctional sensor for smart digital dust	C.-Y. You; B.-F. Hu; B.-R. Xu; Z.-Y. Zhang; B.-M. Wu; G.-S. Huang; E.-M. Song; Y.-F. Mei	2022	x				
Group-based Discretionary Access Control for Epidemiological Resources	J. Zamite; D. Domingos; M. J. Silva; C. Santos	2013	x				
Value co-creation and appropriation of platform-based alliances in cooperative advertising	L. Zhang; F.-W. Chen; S.-M. Xia; D.-M. Cao; Z. Ye; C.-R. Shen; G. Maas; Y.-M. Li	2021	x				
Democratic learning: hardware/software co-design for lightweight blockchain-secured on-device machine learning	R. Zhang; M. Song; T. Li; Z. Yu; Y. Dai; X. Liu; G. Wang	2021	x				
Social media security and trustworthiness: Overview and new direction	Z. Zhang; B. B. Gupta	2018	x				
A velocity decomposition method combining potential and viscous flow for the estimation of submarine's hydrodynamic coefficients	J. Zhao; R. C. Zhu; X. Chen; L. Hong	2021	x				
Research on the evolution of the innovation ecosystem of the Internet of Things: A case study of Xiaomi(China)	W. Zhao; L. Yi	2022	x				
Key factors and generation mechanisms of open government data performance: A mixed methods study in the case of China	Y. Zhao; Y. Liang; C. Yao; X. Han	2022	x				
Open-source tribometer with high repeatability: Development and performance assessment	Y. Zhou; Y. Tian; S. Meng; S. Zhang; X. Xing; Q. Yang; D. Li	2023	x				
Node Embedding and Classification with Adaptive Structural Fingerprint	Y. Zhu; J. Wang; J. Zhang; K. Zhang	2022	x				
Theft detection dataset for benchmarking and machine learning based classification in a smart grid environment	S. Zidi; A. Mihoub; S. Mian Qaisar; M. Krichen; Q. Abu Al-Haija	2023	x				
A topology of groups: What GitHub can tell us about online collaboration	N. Zöller; J. H. Morgan; T. Schröder	2020	x				
The Engagement of Complementors and the Role of Platform Boundary Resources in e-Commerce Platform Ecosystems	Martin Engert Julia Evers Andreas Hein Helmut Krcmar	2022	x	x	x	x	x
Exploiting Repositories in Mobile Software Ecosystems from a Governance Perspective	Awdren de Lima Font vÉoRodrigo Pereira dos Santos Arilo Claudio Dias-Neto	2019	x	x			
The Unintended Consequences of Automated Scripts in Crowdfund Platforms: A Simulation Study in MTurk	Haoyu Xie Alessandro Checco Efraxia D. Zamani	2023	x				
Design of an O2O Citizen Participation Ecosystem for Sustainable Governance	Jingrui Ju Luning Liu Yuqiang Feng	2019	x	x			
The Application of the Principles of Responsible AI on Social Media Marketing for Digital Health	Rui Liu Suraksha Gupta Parth Patel	2021	x				
Information Exchange Architecture for Collaborative Industrial Ecosystem	Petri Kannisto David HV stbackaArto Marttinen	2020	x	x			
How to Achieve Swift Resilience: the Role of Digital Innovation Enabled Mindfulness	Dandan Ye Martin J. Liu Jun Luo Natalia Yannopoulou	2022	x				
Measuring the Big Data Readiness of Developing Countries ,À Index Development and its Application to Africa	Anke Joubert Matthias Murawski Markus Bick	2023	x				
Advances in Social Media Research: Past Present and Future	Kawaljeet Kaur Kapoor Kuttimani Tamilmani Nripendra P. Rana Pushp Patil Yogesh K. Dwivedi Sridhar Nerur	2018	x				
A Confirmation Bias View on Social Media Induced Polarisation During Covid-19	Sachin Modgil Rohit Kumar Singh Shivam Gupta Denis Dennehy	2021	x				
Disentangling Capabilities for Industry 4.0 - an Information Systems Capability Perspective	Rocco Huber Anna Maria Ober V\$nderUlrich Faisst Maximilian RV dglinger	2022	x				

Towards Increased Understanding of Open Data Use for Software Development	Maciej GrzendaJaroslaw Legierski	2021	x	x		
Fostering distributed business logic in Open Collaborative Networks: an integrated approach based on semantic and swarm coordination	Francesco P. AppioMario G. C. A. CiminoAlessandro LazzeriAntonella MartiniGigliola Vaglini	2018	x			
Security Privacy and Risks Within Smart Cities: Literature Review and Development of a Smart City Interaction Framework	Elvira IsmagilovaLaurie HughesNripendra P. RanaYogesh K. Dwivedi	2022	x			
Bridging Digital Divides: a Literature Review and Research Agenda for Information Systems Research	Polyxeni VassilakopoulouEli Hustad	2021	x			
On the Combinatory Nature of Knowledge Transfer Conditions: A Mixed Method Assessment	Emily BaconMichael D. WilliamsGareth H. Davies	2021	x			
Applications of Blockchain in Industry 4.0: a Review	Yong ChenYang LuLarisa BulyshevaMikhail Yu. Kataev	2022	x	x		
Classification of Smart City Research - a Descriptive Literature Review and Future Research Agenda	Parul GuptaSumedha ChauhanM. P. Jaiswal	2019	x			
Disruption and Legitimacy: Big Data in Society	Carlos FerreiraAlessandro MerendinoMaureen Meadows	2021	x			
Online knowledge sharing mechanisms: a systematic review of the state of the art literature and recommendations for future research	Yeganeh CharbandNima Jafari Navimipour	2016	x	x		
Cyber-Physical Systems in the Context of Industry 4.0: A Review Categorization and Outlook	Sascha Julian OksMax JalowskiMichael LechnerStefan MirschbergerMarion MerkleinBirgit Vogel-HeuserKathrin M. Mvölslein	2022	x			
Characterization of Cloud Computing Reversibility as Explored by the DELPHI Method	Wafa Bouaynaya	2020	x			
Participation Patterns and Reliability of Human Sensing in Crowd-Sourced Disaster Management	Robert I. OgieHugh ForeheadRodney J. ClarkePascal Perez	2018	x			
A Twitter-Based Study of the European Internet of Things	Funda Ustek-SpildaDavide VegaMatteo MagnaniLuca Rossilrina ShklovskiSebastian LehuedeAlison Powell	2021	x			
Special issue on smart connected hospitality and tourism	Chulmo KooFrancesco RicciCihan CobanogluFevzi Okumus	2017	x			
Challenges in platforming and digitizing decentralized energy services	Ahmed IdriesJohn KrogstieJayaprakash Rajasekharan	2022	x			
Co-Creating Platform Governance Models Using Boundary Resources: a Case Study from Dementia Care Services	Babak A. FarshchianHanne Ekran Thomassen	2019	x	x	x	x
Organizing for openness: six models for developer involvement in hybrid OSS projects	Hanna MvšenpvšvšSimo MvškinenTerhi KilamoTommi MikkonenTomi MvšnnistvðPaavo Ritala	2018	x			
Supporting governance of mobile application developers from mining and analyzing technical questions in stack overflow	Awdren FontvÉoBruno vÅbialogr WieseBernardo Estv°cioMarcelo QuintaRodrigo Pereira dos SantosArilo Claudio Dias-Neto	2018	x			
Crowd intelligence in AI 2.0 era	Wei LiWen-jun WuHuai-min WangXue-qi ChengHua-jun ChenZhi-hua ZhouRong Ding	2017	x			
Smart cities in the new service economy: building platforms for smart services	Ari-Veikko AnttiroikoPekka ValkamaStephen J. Bailey	2014	x	x		
Retailers Dual Role in Digital Marketplaces	Tobias WulfertReinhard Schv°tte	2022	x	x	x	x
Security issues in cloud environments: a survey	Diogo A. B. FernandesLiliana F. B. SoaresJovÉo V. GomesMv°rio M. FreirePedro R. M. Inv°cio	2014	x			
Openness and privacy in born-digital archives: reflecting the role of AI development	Angeliki Tzouganatou	2022	x			

Security-driven distributed platforms for intellectual property resource provision - a case study of TSITE IP	Yang Wang	2021	x			
What motivates „ðfree,Ä revealing? Measuring outbound non-pecuniary openness innovation types and expectations of future profit growth	Martie-Louise VerreyneRui Torres de OliveiraJohn SteenMarta IndulskajA. Ford	2020	x			
On the sustainability of smart and smarter cities in the era of big data: an interdisciplinary and transdisciplinary literature review	Simon Elias Bibri	2019	x			
Building an Industry-†4.0 Analytics Platform	Christoph Grvðger	2018	x	x		
Evolution of Blockchain and consensus mechanisms & its real-world applications	Amrendra Singh YadavNikita SinghDharmender Singh Kushwaha	2023	x	x		
Leading the Charge on Digital Regulation: The More the Better or Policy Bubble?	Cristiano CodagnoneLinda Weigl	2023	x			
Model-driven development platform selection: four industry case studies	Siamak FarshidiSlinger JansenSven Fortuin	2021	x			
A service requirements engineering method for a digital service ecosystem	Anne ImmonenEila OvaskaJarmo KalaojaDaniel Pakkala	2016	x			
A review of smart contract-based platforms applications and challenges	Pratima SharmaRajni JindalMalaya Dutta Borah	2023	x			
The anatomy of the data-driven smart sustainable city: instrumentation datafication computerization and related applications	Simon Elias Bibri	2019	x			
A network analysis of strategic alliance drivers in ICT open ecosystem: with focus on mobile cloud computing and multimedia	Jeongeun ByunTae-Eung SungHyun-woo Park	2018	x			
Utilising stream reasoning techniques to underpin an autonomous framework for cloud application platforms	Rustem DautovIraklis ParaskakisMike Stannett	2014	x			
The European Project in the Materials Informatics Domain: Ontologies and Virtual Platforms	A. O. ErkimbaevV. Yu. ZitsermanG. A. KobzevA. V. Kosinov	2021	x			
The open agent society: retrospective and prospective views	Jeremy PittAlexander Artikis	2015	x			
EdgeKeeper: a trusted edge computing framework for ubiquitous power Internet of Things	Weiyong Yang Êù®ÁªÏÊ∞TTWei Liu ÀàÈãáXingshen Wei Ê≠èÂÖÏÉÖéZixin Guo ÊÉ≠Á≠èÊöiKangle Yang Êù®ÁjΣ%πêHao Huang ÊªÑÁöiLongyun Qi Á•ÁÊæð%jê	2021	x			
Open innovation in software engineering: a systematic mapping study	Hussan MunirKrzysztof WnukPer Runeson	2016	x			
Cloud computing in e-Science: research challenges and-†opportunities	Xiaoyu YangDavid WallomSimon WaddingtonJianwu WangArif ShaonBrian MatthewsMichael WilsonYike GuoLi GuoJon D. BlowerAthanasios V. VasilakosKecheng LiuPhilip Kershaw	2014	x			
Continuous clarification and emergent requirements flows in open-commercial software ecosystems	Eric KnaussAminah YussufKelly BlincoeDaniela DamianAlessia Knauss	2018	x			
Designed Features for Improving Openness Scalability and Programmability in the Fog Computing-Based IoT Systems	Quang Minh TranPhat Huu NguyenTakeshi TsuchiyaMichel Toulouse	2020	x			
A survey of autonomic computing methods in digital service ecosystems	Dhaminda B. AbeywickramaEila Ovaska	2017	x			
Managing to release early often and on time in the OpenStack software ecosystem	Josv© Apolinv°rio TeixeiraHelena Karsten	2019	x			
Blockchain-enabled digital rights management for multimedia resources of online education	Junqi GuoChuyang LiGuangzhi ZhangYunchuan SunRongfang Bie	2020	x	x	x	
Environmentally data-driven smart sustainable cities: applied innovative solutions for energy	Simon Elias BibriJohn Krogstie	2020	x			

efficiency pollution reduction and urban metabolism						
A systematic review of the purposes of Blockchain and fog computing integration: classification and open issues	Yehia Ibrahim AlzoubiAsif GillAlok Mishra	2022	x	x		
A systematic review of artificial intelligence impact assessments	Bernd Carsten StahlJosephina AntoniouNitika BhallaLaurence BrooksPhilip JansenBlerta LindqvistAlexey KirichenkoSamuel MarchalRowena RodriguesNicole SantiagoZuzanna WarsoDavid Wright	2023	x			
A fine-grained task scheduling mechanism for digital economy services based on intelligent edge and cloud computing	Xiaoming Zhang	2023	x			
Open innovation using open source tools: a case study at Sony Mobile	Hussan MunirJohan Linv•kerKrzysztof WnukPer RunesonBjVðrn Regnell	2018	x			
Towards trustworthy blockchain systems in the era of ‚ÄInternet of value,Ä: development challenges and future trends	Hai JinJiang Xiao	2021	x			
Investigations into data published and consumed on the Web: a systematic mapping study	Helton Douglas A. dos SantosMarcelo lury S. OliveiraGlVzria de Fv°tima A. B. LimaKarina Moura da SilvaRayelle I. Vera Cruz S. MunizBernadette Farias LVzscio	2018	x			
Experiences in building a mOSAIC of clouds	Dana PetcuBeniamino Di MartinoSalvatore VenticinqueMassimiliano RakTamv°s Mv°hrGorka Esnal LopezFabrice BritoRoberto CossuMiha StoparSvatopluk ≈+perkaVlado Stankovski	2013	x			
Abstracts from the 8th DACH+ Conference on Energy Informatics		2019	x			
A Responsible Internet to Increase Trust in the Digital World	Cristian HesselmanPaola GrossoRalph HolzFernando KuipersJanet Hui XueMattijs JonkerJoeri de RuitterAnna SperottoRoland van Rijswijk-DeijGiovane C. M. MouraAiko PrasCees de Laat	2020	x			
Cyber governance studies in ensuring cybersecurity: an overview of cybersecurity governance	Serkan Sava≈üSV°leyman Karata≈ü	2022	x			
A survey of blockchain applications in sustainable and smart cities	Shanukha MakaniRachitha PittalaEitaa AlsayedMoayad AloqailyYaser Jararweh	2022	x			
Modelling in low-code development: a multi-vocal systematic review	Alessio BucaioniAntonio CicchettiFederico Ciccozzi	2022	x			
Political machines: a framework for studying politics in social machines	Orestis Papakyriakopoulos	2022	x			
The Library in the Information Ecosystem of Open Science	N. S. Redkina	2021	x			
Cloud-based vs. blockchain-based IoT: a comparative survey and way forward	Raheel Ahmed MemonJian Ping LiJunaid AhmedMuhammad Irshad NazeerMuhammad IsmailKhursheed Ali	2020	x			
Internet of Things and Big Data: the disruption of the value chain and the rise of new software ecosystems	Norbert Jesse	2018	x	x		
Deep learning application in smart cities: recent development taxonomy challenges and research prospects	Amina N. MuhammadAli M. AseereHaruna ChiromaHabib ShahAbdulsalam Y. Gitallbrahim Abaker Targio Hashem	2021	x			
Saudi cloud infrastructure: a security analysis	Wahid RajehHai JinDeqing Zou	2017	x			

Overview of 5G security technology	Xinsheng JiKaizhi HuangLiang JinHongbo TangCaixia LiuZhou ZhongWei YouXiaoming XuHua ZhaoJiangxing WuMing Yi	2018	x			
Open Access Driven Transformation of the Scientific Communication System: Current Status Prerequisites for Change Effects and Prospects	N. D. Trishchenko	2019	x			
Democratizing cognitive technology: a proactive approach	Marcello Ienca	2019	x			
A Web-based distributed architecture for multi-device adaptation in media applications	Mikel ZorrillaNjv•l BorchFranvBois DaoustAlexander ErkJuliv°n Flv≥rezAlberto Lafuente	2015	x			
Consuming Resources and Services from Multiple Clouds	Dana Petcu	2014	x			
A survey on Blockchain mechanisms (BCM) based on internet of things (IoT) applications	C. SathishC. Yesubai Rubavathi	2022	x			
Blockchain-based trust management in cloud computing systems: a taxonomy review and future directions	Wenjuan LiJiyi Wujian CaoNan ChenQifei ZhangRajkumar Buyya	2021	x			
Resilient sensor authentication in SCADA by integrating physical unclonable function and blockchain	Abel O. Gomez RiveraDeepak K. ToshUttam Ghosh	2022	x			
Cyber-physical systems challenges: a needs analysis for collaborating embedded software systems	Pieter J. MostermanJustyna Zander	2016	x			
Ethics-based auditing of automated decision-making systems: intervention points and policy implications	Jakob MvðkanderMaria Axente	2023	x			
What,Ãs news? Encounters with news in everyday life: a study of behaviours and attitudes	Sally Jo CunninghamDavid M. NicholsAnnika HinzeJudy Bowen	2016	x			
Towards evolvable Internet architecture-design constraints and models analysis	Ke XuMin ZhuGuangWu HuLiang ZhuYiFeng ZhongYing LiuJianPing WuNing Wang	2014	x			
Dynamic adaptation of service-based applications: a design for adaptation approach	Martina De SanctisAntonio BucchiaroneAnnapaola Marconi	2020	x			
Extending organizational capabilities with Open Data to support sustainable and dynamic business ecosystems	JfÅnis KamparsJelena ZdravkovicJanis StirnaJfÅnis Grabis	2020	x	x		
Pedagogies for the open knowledge society	Paola Ricaurte	2016	x			
Towards a standard modeling of social health care practice	Mouhamed Gaith AyadiRiadh BouslimiJalel Akaichi	2022	x			
Online teaching research in universities based on blockchain	Li MinGe Bin	2022	x			
,Ã0What are these researchers doing in my Wikipedia?,Ã0: ethical premises and practical judgment in internet-based ethnography	Christian Pentzold	2017	x			
Robust Cross-Platform Workflows: How Technical and Scientific Communities Collaborate to Develop Test and Share Best Practices for Data Analysis	Steffen MvðllerStuart W. PrescottLars WirzeniusPetter ReinholdtsenBrad ChapmanPjotr PrinsStian Soiland- ReyesFabian KlvðtzlAndrea BagnacaniMatVf≈° Kala≈°Andreas TilleMichael R. Crusoe	2017	x			
Pull-the-strings	Luis Leite	2022	x			
The contribution of organizational culture structure and leadership factors in the digital transformation of SMEs: a mixed-methods approach	Bernardo Henrique LesoMarcelo Nogueira CortimigliaAntonio Ghezzi	2023	x			
Big and open linked data analytics: a study on changing roles and skills in the higher educational process	Martin LnenickaHana KopackovaRenata MachovaJitka Komarkova	2020	x			
Toward Self-monitoring Smart Cities: the OpenSense2 Approach	Jean-Paul CalbimonteJulien EberleKarl Aberer	2017	x	x		

Fedidchain: An Innovative Blockchain-Enabled Framework for Cross-Border Interoperability and Trust Management in Identity Federation Systems	Samia El HaddoutiAbdellah OuaguidMohamed Dafir Ech-Cherif El Kettani	2023	x			
An ethico-legal framework for social data science	Nikolaus ForgvStefanie HVšnoldJeroen van-t den HovenTina KrV°gellryna LishchukRenv© MahieuAnna MonrealeDino PedreschiFrancesca PratesiDavid van Putten	2021	x			
Information Resilience: the nexus of responsible and agile approaches to information use	Shazia SadiqAmir AryaniGianluca DemartiniWen HuaMarta IndulskaAndrew Burton-JonesHassan KhosraviDiana Benavides-PradoTimos SellisIlda SomehRhema VaithianathanSen WangXiaofang Zhou	2022	x			
The Evolution of Research on Digital Education	Pierre Dillenbourg	2016	x			
A Survey on the Use of Lightweight Virtualization in I4.0 Manufacturing Environments	Giuseppe Di ModicaLuca Foschini	2023	x			
State of the science on the Cloud accessibility and the future	Amrish ChourasiaDan NordstromGregg Vanderheiden	2014	x			
Towards data-driven tele-medicine intelligence: community-based mental healthcare paradigm shift for smart aging amid COVID-19 pandemic	Lan ChengWK ChanYi PengHarry Qin	2023	x			
Dynamic resource provisioning for cyber-physical systems in cloud-fog-edge computing	Zhanyang XuYanqi ZhangHaoyuan LiWei Jing YangQuan Qi	2020	x			
Blockchain technology in IoT systems: current trends methodology problems applications and future directions	Abraham Ayegba AlfaJohn Kolo AlhassanOlayemi Mikail OlaniyiMorufu Olalere	2021	x			
RETRACTED ARTICLE: Current trends in the digital transformation of higher education institutions in Russia	Alexey MikheevYana SerkinaAlexander Vasyaev	2021	x			
On the Integration of Blockchain and SDN: Overview Applications and Future Perspectives	Anichur RahmanAntonio MontieriDipanjali KunduMd. Razaul KarimMd. Jahidul IslamSara UmmeAlfredo NascitaAntonio PescapV©	2022	x			
Internet of Low-Altitude UAVs (IoLoUA): a methodical modeling on integration of Internet of ,ÁúThings,Àù with ,ÁúUAV,Àù possibilities and tests	Ashish SrivastavaJay Prakash	2023	x			
Studying donations and their expenses in open source projects: a case study of GitHub projects collecting donations through open collectives	Jiayuan ZhouShaowei WangYasutaka KameiAhmed E. HassanNaoyasu Ubayashi	2021	x			
EOS.IO blockchain data analysis	Wanshui SongWenyin ZhangLinbo ZhaiLuanqi LiuJiuru WangShanyun HuangBei Li	2022	x			
Ethical implications of digital infrastructures for pluralistic perspectives	Maria Joseph IsraelAhmed Amer	2021	x			
Blockchain-based secure data transmission for internet of underwater things	Abdul Razzaq	2022	x			
Empowering cultural heritage professionals with tools for authoring and deploying personalised visitor experiences	Elena NotDaniela Petrelli	2019	x			
A systematic review on multi-device inclusive environments	Ig lbert BittencourtMaria Cecv#lia BaranauskasRoberto PereiraDiego DermevalSeiji IsotaniPatrv#cia Jaques	2016	x			
A tailored participatory action research for foss communities	Adam AlamiPeter Axel NielsenAndrzej WaÁłowski	2020	x			
Disclose to Tell: a Data Design Framework for Alternative Narratives	MarV#a de los vÁngeles Briones Rojas	2021	x			
Emerging blockchain-based applications and techniques	Yinsheng Li	2019	x			

The case for the Humanities Citation Index (HuCI): a citation index by the humanities for the humanities	Giovanni ColavizzaSilvio PeroniMatteo Romano	2022	x			
Research on performance optimization of crowd innovation space from the perspective of participation motivation	Xin JinMin ZhangXiao Hou	2021	x			
A human-centered decentralized architecture and recommendation engine in SIoT	Daniel DefiebreDimitris SacharidisPanagiotis Germanakos	2022	x			
B-space: dynamic management and assurance of open systems of systems	Daniel SchneiderMario Trapp	2018	x			
Between Personal and Common: the Design of Hybrid Information Spaces	Polyxeni VassilakopoulouMiria GrisotMargunn Aanestad	2019	x	x		
On the privacy of mental health apps	Leonardo Horn IwayaM. Ali BabarAwais RashidChamila Wijayarathna	2022	x			
The power of big data mining to improve the health care system in the United Arab Emirates	Khawla Eissa AlhajalImmanuel Azaad Moonesar	2023	x			
An in-depth and systematic literature review on the blockchain-based approaches for cloud computing	Jianhu GongNima Jafari Navimipour	2022	x			
Responsible innovation in synthetic biology in response to COVID-19: the role of data positionality	Koen Bruynseels	2021	x			
Developer Role Evolution in Open Source Software Ecosystem: An Explanatory Study on GNOME	Can ChengBing LiZeng-Yang LiYu-Qi ZhaoFeng-Ling Liao	2017	x			
Intertwining globality and locality: bibliometric analysis based on the top geography annual conferences in America and China	Liang ZhuangChao YeScott N. Lieske	2020	x			
Mathematical modeling and performance evaluation of BeRAN for 6G wireless networks	Vuppula RoopaHimansu Shekhar Pradhan	2023	x			
Open innovation from the perspective of network embedding: knowledge evolution and development trend	Ting LiuLiu Tang	2020	x			
Inter-team communication in large-scale co-located software engineering: a case study	Elizabeth BjarnasonBaldvin Gislason BernLinda Svedberg	2022	x			
Are teachers techno-optimists or techno-pessimists? A pilot comparative among teachers in Bolivia Brazil the Dominican Republic Ecuador Finland Poland Turkey and Uruguay	≈Åukasz TomczykVladimir Costas Jv°ureguiCibelle Albuquerque de La Higuera AmatoDarwin Muv±ozMagali ArteagaSolomon Sunday OyelerevñzgV°r Ya≈üar AkyarMariana Porta	2021	x			
The AI gambit: leveraging artificial intelligence to combat climate change,Äioopportunities challenges and recommendations	Josh CowsAndreas TsamadosMariosaria TaddeoLuciano Floridi	2023	x			
Psychological predictors of consumer-level virtual reality technology adoption and usage	James J. CummingsTiernan J. CahillErin WertzQiankun Zhong	2022	x			
An evaluation framework for software crowdsourcing	Wenjun WuWei-Tek TsaiWei Li	2013	x			
Emerging research areas in SIP-based converged services for extended Web clients	Michael AdeyeyePaolo Bellavista	2014	x			
Docker Cluster Management for the Cloud - Survey Results and Own Solution	Renv© PeiniFlorian HolzschuherFlorian Pfitzer	2016	x			
Ethical artificial intelligence framework for a good AI society: principles opportunities and perils	Pradeep ParamanSanmugam Anamalah	2023	x			
Legal dilemmas of Estonian artificial intelligence strategy: in between of e-society and global race	Tanel KerikmVšeEvelin PVšrn-Lee	2021	x			
Ethical assurance: a practical approach to the responsible design development and deployment of data-driven technologies	Christopher BurrDavid Leslie	2023	x			
OpenIaC: open infrastructure as code - the network is my computer	Chunming RongJiahui GengThomas J. HackerHaakon BryhniMartin G. Jaatun	2022	x			

What Empirically Based Research Tells Us About Game Development	Björn Berg MarklundHenrik EngströmMarcus HellkvistPer Backlund	2019	x			
Internet of Things is a revolutionary approach for future technology enhancement: a review	Sachin KumarPrayag TiwariMikhail Zymbler	2019	x			
The computation of boundary spanning for the IT-enabled commercial ecosystem	He LiTaoHua Ouyang	2017	x			
How do people judge the credibility of algorithmic sources?	Donghee Shin	2022	x			
A conceptual-driven survey on future internet requirements technologies and challenges	Antonio Marcos Alberti	2013	x			
What to share when and where: balancing the objectives and complexities of open source software contributions	Johan LinViktor Björn Regnell	2020	x			
Embedding artificial intelligence in society: looking beyond the EU AI master plan using the culture cycle	Simone BorsciVille V. LehtolaFrancesco NexMichael Ying YangEllen-Wien AugustijnLeila BagheriyeChristoph BruneOurania KounadiJamy LiJoao MoreiraJoanne Van Der NagelBernard VeldkampDuc V. LeMingshu WangFons WijnhovenJelmer M. Wolterink	2022	x			
Hydrological stream data pipeline framework based on IoTDB	YuanSheng LouYu QinFeng YePeng ZhangYong Chen	2019	x			
Investigation of RISC-V	V. A. FrolovV. A. GalaktionovV. V. Sanzharov	2021	x			
Comparative legal study on privacy and personal data protection for robots equipped with artificial intelligence: looking at functional and technological aspects	Kaori Ishii	2019	x			
Echo chamber detection and analysis	Giacomo VillaGabriella PasiMarco Viviani	2021	x			
U-cities reshaping our future: reflections on ubiquitous infrastructure as an enabler of smart urban development	Ari-Veikko Anttiroiko	2013	x			
Cognitive architectures for artificial intelligence ethics	Steve J. BickleyBenno Torgler	2023	x			
Discovering community patterns in open-source: a systematic approach and its evaluation	Damian A. TamburriFabio PalombaAlexander SerebrenikAndy Zaidman	2019	x			
A Survey on Software Defined Networking: Architecture for Next Generation Network	Sanjeev SinghRakesh Kumar Jha	2017	x			
Empirical analysis of security vulnerabilities in Python packages	Mahmoud AlfadelDiego Elias CostaEmad Shihab	2023	x			
From Artefacts to Infrastructures	Eric MonteiroNeil PollockOle HansethRobin Williams	2013	x			
A review on architecture and models for autonomic software systems	Pooja DehrajArun Sharma	2021	x			
LegoDroid: flexible Android app decomposition and instant installation	Yi LiuYun MaXusheng XiaoTao XieXuanzhe Liu	2023	x			
From Publics to Communities: Researching the Path of Shared Issues Through ICT	Thomas LudwigChristian ReuterVolkmar Pipek	2016	x			
Web Intelligence meets Brain Informatics: Towards the future of artificial intelligence in the connected world	Hongzhi KuaiXiaohui TaoNing Zhong	2022	x			
Community clouds within M-commerce: a privacy by design perspective	Farid ShiraziAmna Iqbal	2017	x			
Blockchain-assisted secured data management framework for health information analysis based on Internet of Medical Things	Asad AbbasRoobaea AlroobaeaMoez KrichenSaeed RubaieeS. VimalFahad M. Almansour	2021	x	x		
Improving the accuracy of Business-to-Business (B2B) reputation systems through rater expertise prediction	Heidi DikowOmar HasanHarald KoschLionel BrunieRenaud Sornin	2015	x			

Role of hub and spoke model for ICTs in agriculture	Nancy J. AnabelPriyanka MohanR. Rajkumar	2018	x			
Development and implementation processes of digitalization in engineer-to-order manufacturing: enablers and barriers	Sylvi ThunOttar Bakv•sTore Christian BjvTJrsvik Storholmen	2022	x			
Design requirements of a modern business Master,Åds degree course: perspectives of industry practitioners	Shah J Miahlan Solomonides	2021	x			
Tying Knots: Participatory Infrastructuring at Work	Susanne BVJdkerChristian DindlerOle Sejer Iversen	2017	x			
Understanding Nomadic Practices of Social Activist Networks Through the Lens of Infrastructuring: the Case of the European Social Forum	Aparecido Fabiano Pinatti de CarvalhoSaqib SaeedChristian ReuterMarkus RohdeDavid RandallVolkmar PipekVolker Wulf	2022	x			
Reframing data ethics in research methods education: a pathway to critical data literacy	Javiera AtenasLeo HavemannCristian Timmermann	2023	x			
The pursuit of computational justice in open systems	Jeremy PittDV#dac BusquetsRV@gis Riveret	2015	x			
A systematic review of the opportunities and challenges of micro-credentials for multiple stakeholders: learners employers higher education institutions and government	Soovendran VaradarajanJoyce Hwee Ling KohBen Kei Daniel	2023	x			
Societal impact of university research in the written press: media attention in the context of SIUR and the open science agenda among social scientists in Flanders Belgium	Hans JonkerFlorian VanleeWalter Ysebaert	2022	x			
Hybreed: A software framework for developing context-aware hybrid recommender systems	Tim HusseinTimm LinderWerner GaulkeJ#rgen Ziegler	2014	x			
A Code of Digital Ethics: laying the foundation for digital ethics in a science and technology company	Sarah J. BeckerAndrv© T. NematSimon LucasRenv© M. HeinitzManfred KlevesathJean Enno Charton	2022	x			
Thoughts on the development of novel network technology	Jiangxing Wu	2018	x			
Misleading information in Spanish: a survey	Eliana ProvidelMarcelo Mendoza	2021	x			
ICT Services for open and citizen science	Miko=Çaj Morzy	2015	x			
,ÅÚIntelligent Justice,Åù: human-centered considerations in China,Åds legal AI transformation	Nyu WangMichael Yuan Tian	2022	x			
The ethics of AI business practices: a review of 47 AI ethics guidelines	Blair Attard-FrostAndrv©s De los RV#osDeneille R. Walters	2022	x			
A formally verified blockchain-based decentralised authentication scheme for the internet of things	Khizar HameedSaurabh GargMuhammad Bilal AminByeong Kang	2021	x			
On middleware for emerging health services	Jatinder SinghJean M Bacon	2014	x			
Testing android malware detectors against code obfuscation: a systematization of knowledge and unified methodology	Mila Dalla PredaFederico Maggi	2017	x			
Charting the Open Access scholarly journals landscape in the UAE	Mohamed Boufarss	2020	x			
Deep learning for intelligent traffic sensing and prediction: recent advances and future challenges	Xiaochen FanChaocan XiangLiangyi GongXin HeYuben QuSaeed AmirgholipourYue XiPriyadarsi NandaXiangjian He	2020	x			
Towards AI ethics,Åd institutionalization: knowledge bridges from business ethics to advance organizational AI ethics	Mario D. SchultzPeter Seele	2023	x			
Influence of research on open science in the public policy sphere	Daniela De FilippoPablo Sastrv≥n-Toledo	2023	x			
The potential of an artificial intelligence (AI) application for the tax administration system,Åds modernization: the case of Indonesia	Arfah Habib SaragihQaumy ReyhaniMilla Sepliana SetyowatiAdang Hendrawan	2022	x			

Choosing the right collaboration partner for innovation: a framework based on topic analysis and link prediction	Yan QiXin ZhangZhengyin HuBin XiangRan ZhangShu Fang	2022	x			
Principle-based recommendations for big data and machine learning in food safety: the P-SAFETY model	Salvatore SapienzaAnton Vedder	2023	x			
Potential for the use of large unstructured data resources by public innovation support institutions	Wiesław CeteraWłodzimierz GogoÇekAleksander ŁoźniewskiDariusz Jaruga	2022	x			
Applying the blockchain-based deep reinforcement consensus algorithm to the intelligent manufacturing model under internet of things	Tongtong GengYueping Du	2022	x			
Understanding the need for assistance in software modeling: interviews with experts	Maxime Savary-LeblancXavier LeclercqPascal BastienGérard	2023	x			
LimonDroid: a system coupling three signature-based schemes for profiling Android malware	Franklin TchakountzianRoger Corneille Ndjeumou NgassiVivient Corneille KamlaKalum Priyanath Udagepola	2021	x			
Deep learning deep change? Mapping the evolution and geography of a general purpose technology	Joel KlingerJuan Mateos-GarciaKonstantinos Stathoulopoulos	2021	x			
Run-time model based framework for automatic evaluation of multimodal interfaces	Pedro Luis Mateo NavarroStefan HillmannSebastian MüllerDiego Sevilla RuizGregorio Martínez Pérez	2014	x			
System innovations in open WDM DCI networks	Loukas ParaschisHarald BockAbhinava Shivakumar SadasivaraoSharfuddin SyedBernd Sommerkorn-KrombholzJeff RahnBiao LuJoao PedroPaul DoolanParthiban Kandappan	2020	x			
A distributed architecture for efficient Web service discovery	Luciano BaresiMatteo MirazPierluigi Plebani	2016	x			
Cybernetics and systems art in Latin America: the art and communication center (CAyC) and its pioneering art and technology network	José Carlos Marivetequi	2022	x			
RETRACTED ARTICLE: Community guidance model based on interactive multimedia system	Xian Ji	2019	x			
Dark Web Traffic Analysis of Cybersecurity Threats Through South African Internet Protocol Address Space	C. GokhaleO. O. Olugbara	2020	x			
Open source and accessibility: advantages and limitations	Michael HeronVicki L HansonIan Ricketts	2013	x	x		
RepoLike: a multi-feature-based personalized recommendation approach for open-source repositories	Cheng YangQiang FanTao WangGang YinXun-hui ZhangYue YuHua-min Wang	2019	x			
A review of morphogenetic engineering	René DoursatHiroki SayamaOlivier Michel	2013	x			
A digital twin framework for improving energy efficiency and occupant comfort in public and commercial buildings	Anders ClausenKrzysztof ArendtAslak JohansenFisayo Caleb SangogboyeMikkel Baun KjellergaardChristian T. VejeBo Nørregaard Jørgensen	2021	x			
Towards NFC payments using a lightweight architecture for the Web of Things	Tor-Morten GrønliPardis PourghomiGheorghita Ghinea	2015	x			
Information security implications of using NLP in IT outsourcing: a Diffusion of Innovation theory perspective	Baber Majid BhattiSameera MubarakSev Nagalingam	2021	x			
Rule-based tools for the configuration of ambient intelligence systems: a comparative user study	Federico CabitzaDaniela FogliRosa LanzilottiAntonio Piccinno	2017	x			
Software-defined optical networks (SDONs): a survey	Partha BhaumikShuqiang ZhangPulak ChowdhurySang-Soo LeeJong Hyun LeeBiswanath Mukherjee	2014	x			

Multi-dimensional proximity and network stability: the moderating role of network cohesion	Min GuoNaiding YangJingbei WangYanlu Zhang	2021	x			
Leveraging crowd knowledge to curate documentation for agile software industry using deep learning and expert ranking	Akshi Kumar	2021	x			
Data science: developing theoretical contributions in information systems via text analytics	Aya RizkAhmed Elragal	2020	x			
Characterizing networks of propaganda on twitter: a case study	Stefano GuarinoNoemi TrinoAlessandro CelestiniAlessandro ChessaGianni Riotta	2020	x			
Zur disruptiven Digitalisierung von Hochschulforschung	Markus von der HeydeAndreas HartmannGunnar AuthChristian Erfurth	2018	x			
Universal and specific features of Ukrainian economic research: publication analysis based on Crossref data	O. MryglodS. NazarovetsS. Kozmenko	2021	x			
Millimeter-wave wireless communications for home network in fiber-to-the-room scenario	Chao He ĘŸJĚđŌZhixiong Ren %∞∞∞Ā∅ÉŌŃXiang Wang ÁéãÁ••Yan Zeng Ę∅æĀŃ±Jian Fang ĘñπÁÆ#Debin Hou %∞æ∅ĀæΣĀΩ`Le Kuai Ęí∅%∞πĕRong Lu ĘđŪĀÆπShilin Yang Ęù*%∞∞∞iĘúóZhe Chen ĘđàĀñŪJixin Chen ĘđàĀ∞ßĘñ∞	2021	x			
Employee co-invention network dynamics and firm exploratory innovation: the moderation of employee co-invention network centralization and knowledge-employee network equilibrium	Guiyang Zhang	2021	x			
Smart Managed Freight Fleet: Ein automatisiertes und vernetztes Flottenmanagement in einem fvðderierten Datenvðkosystem	Christoph HeinbachHenning GvðslingPascal MeierOliver Thomas	2023	x			
Dynamic traffic forecasting and fuzzy-based optimized admission control in federated 5G-open RAN networks	Abida PerveenRaouf AbozaribaMohammad PatwaryAdel Aneiba	2021	x			
Secured two factor authentication graph based replication and encryption strategy in cloud computing	S. LavanyaN. M. Saravanakumar	2023	x			
Secure the ownership of WoT devices using secure ownership transfer framework	Kaptan SinghDeepak Singh Tomar	2023	x			
The physics of software tools: SWOT analysis and vision	Bernhard Steffen	2017	x			
Successful failure: what Foucault can teach us about privacy self-management in a world of Facebook and big data	Gordon Hull	2015	x			
`	Yifei YuanYajun ZhaoBaiqing ZongSergio Parolari	2020	x			
Identifying learners,Ā∞ topical interests from social media content to enrich their course preferences in MOOCs using topic modeling and NLP techniques	Hajar ZankadiAbdellah IdrissiNajima Daoudilmane Hilal	2023	x			
Visual topical analysis of library and information science	Pin LiGuoli YangChuanqi Wang	2019	x			
Is open science a double-edged sword?: data sharing and the changing citation pattern of Chinese economics articles	Liwei ZhangLiang Ma	2023	x			
Innovation systems in Malaysia: a perspective of university,Ā∞industry R&D collaboration	V. G. R. ChandranVeera Pandiyan Kaliani SundramSinnappan Santhidran	2014	x			
Ethical Foresight Analysis: What it is and Why it is Needed?	Luciano FloridiAndrew Strait	2020	x			
Exposed! A case study on the vulnerability-proneness of Google Play Apps	Andrea Di SorboSebastiano Panichella	2021	x			

EnGINE: Flexible Research Infrastructure for Reliable and Scalable Time Sensitive Networks	Filip Rezabek Marcin Bosk Thomas Paul Kilian Holzinger Sebastian Gallenmüller Angela Gonzalez Abdoul Kane Francesc Fons Zhang Haigang Georg Carle Jörg Ott	2022	x			
Using scaffold innovation-thinking frameworks to integrate food science and technology into the transdisciplinary engineering design classroom	Lisa Bosman Sangjun Eom	2019	x			
Publishing scientific data as linked open data	A. O. Erkimbaev V. Yu. Zitserman G. A. Kobzev V. A. Serebrjakov K. B. Teymurazov	2013	x	x		

Appendix 2: Final list of selected papers

Title	Author	Year
A theoretical framework for evaluating government open data platform	P.-Y. Chu; H.-L. Tseng	2016
On the openness of digital platforms/ecosystems	J. Teixeira	2015
Four Tactics for Implementing a Balanced Digital Platform Strategy	K. Karhu; R. Gustafsson; B. Eaton; O. Henfridsson; C. Sørensen	2020
Platform openness: A systematic literature review and avenues for future research	D. Soto Setzke; M. Böhm; H. Krcmar	2019
Governance Mechanisms in Digital Platform Ecosystems: Addressing the Generativity-Control Tension	N. Staub; K. Haki; S. Aier; R. Winter	2022
Design and Governance of mHealth Data Sharing	A. Vesselkov; H. Hämmäinen; J. Töyli	2019
Cultivating open government data platform ecosystems through governance: Lessons from Buenos Aires, Mexico City and Montevideo	C. Bonina; B. Eaton	2020
Digital platform openness: Drivers, dimensions and outcomes	T. L. J. Broekhuizen; O. Emrich; M. J. Gijzenberg; M. Broekhuis; B. Donkers; L. M. Sloot	2021
LUCE: A blockchain-based data sharing platform for monitoring data License accountability and Compliance	V. Urovi; V. Jaiman; A. Angerer; M. Dumontier	2022
The Engagement of Complementors and the Role of Platform Boundary Resources in e-Commerce Platform Ecosystems	Martin Engert Julia Evers Andreas Hein Helmut Krcmar	2022
Co-Creating Platform Governance Models Using Boundary Resources: a Case Study from Dementia Care Services	Babak A. Farshchian Hanne Ekran Thomassen	2019
Retailers Dual Role in Digital Marketplaces	Tobias Wulfert Reinhard Schvötte	2022

Appendix 3: Thematic analysis of Systematic Literature Review

Topic	Article of SLR	Raw data	Initial code	Theme	Sub-theme
Dimensions of Openness	(Soto Setzke et al., 2019)	Furthermore, openness can be categorized as either providing accessibility or transparency [23, 44, 51]. Accessibility focuses on the degree of discrimination against different roles and determines whether providers, third-party developers, or end users are allowed to join and access the platform [23, 51, 52]. Transparency, on the other hand, relates to the “understanding of what is happening and why” and thus determines whether platform-related governance decisions are comprehensible [23, 51]	Openness can be categorized as accessibility or transparency	<ul style="list-style-type: none"> • Accessibility • Transparency 	
Dimensions of Openness	(Soto Setzke et al., 2019)	while transparency refers to how and to what extent the rules for participating are made comprehensible.	Definition of Transparency: Transparency refers to how and to what extent the rules for participating are made comprehensible.	Definition of transparency	
Dimensions of openness	(Soto Setzke et al., 2019)	Other frameworks focus on specific architectural aspects. Schlagwein, Schoder and Fischbach [16] propose a matrix-based framework for measuring the openness of platform resources along the dimensions of access and control. Access to resources can be exclusive, on a group-basis, or open. Control of resources can be exercised by the platform owner, by a group, or by an external actor.	Openness dimensions access and control	<ul style="list-style-type: none"> • Access • Control 	

Dimensions of openness	(Soto Setzke et al., 2019)	Taking the perspective of complementors, Benlian, Hilkert and Hess [13] develop an instrument for measuring complementors' perceived platform openness along the dimensions of technical platform, distribution channel, accessibility, and transparency	Perceived platform openness dimensions technical platform, distribution channel, accesibility and transparency	<ul style="list-style-type: none"> • Technical platform • Distribution channel • Accessibility • Transparency 	
Dimensions of openness	(Broekhuizen et al., 2021)	Previous research (Boudreau, 2010; Eisenmann et al., 2009) defines actor openness according to two dimensions: access (i.e., who is allowed access to the platform?) and authority (i.e., how much is the actor allowed to do on the platform?). ⁵ A platform is "open" to the extent that: (1) no restrictions are placed on participation in its development, commercialization or use (access); and (2) any restrictions (authority) are reasonable and non-discriminatory regarding entry requirements, requirements to conform with technical standards, or payment of licensing fees (Eisenmann et al., 2009).	Actor openness dimensions access and authority	<ul style="list-style-type: none"> • Access • Authority 	

Dimensions of openness	(Broekhuizen et al., 2021)	Hence, platforms need to define their openness for each of the following five openness dimensions: three actor-based (suppliers, customers, complementary services providers), and two nonactor-based (categories and channels) dimensions. For the former, openness is a two-dimensional construct comprising both access and authority, while for the latter it relates to access only (see Table 1).	Platform openness comprise of dimensions access and authority with five openness dimensions supplier, customer, complementor service provider, category and channel.	<ul style="list-style-type: none"> • Access • Authority 	<ul style="list-style-type: none"> • Supplier openness • Customer openness • Complementor service provider openness • Category openness • Channel openness
Governance mechanisms	(Soto Setzke et al., 2019)	For structuring mechanisms for implementing openness, we draw on the notion of horizontal and vertical openness by Eisenmann, Parker and Van Alstyne [11].	Openness mechanisms can be structured as horizontal and vertical openness.	<ul style="list-style-type: none"> • Horizontal openness • Vertical openness 	
Governance mechanisms	(Soto Setzke et al., 2019)	Horizontal openness refers to allowing rival platform's users to interact with the own platform or allowing additional parties to participate in the platform's commercialization or technical development.	Definition of horizontal openness: Allowing users to engage with the platform and enabling external parties to contribute to the platform's commercialization or technical development.	Definition of horizontal openness	
Governance mechanisms	(Soto Setzke et al., 2019)	Vertical openness refers to granting third- party developers access to resources for developing complementary applications.	Definition of vertical openness: granting third- party developers access to resources for developing complementary applications.	Definition of vertical openness	

Governance mechanisms	(Soto Setzke et al., 2019)	Vertical openness is implemented through boundary resources [49, 53], i.e. the “software tools and regulations that serve as the interface for the arm's length relationship between the platform owner and the application developer” [8].	Vertical openness is implemented by boundary resources as tools and regulations	Platform boundary resources	<ul style="list-style-type: none"> • Tools • Regulations
Governance mechanisms	(Soto Setzke et al., 2019)	In practice, this includes technical boundary resources such as Application Programming Interfaces (APIs), SDKs	technical boundary resources include APIs and SDKs	Technical boundary resources	<ul style="list-style-type: none"> • APIs • SDKs
Governance mechanisms	(Soto Setzke et al., 2019)	and non-technical boundary resources such as technical documentation and support or the provided community	non-technical boundary resources include technical documentation, support and a provided community	Non-technical boundary resources	<ul style="list-style-type: none"> • Technical documentation • Support • Provided community
Governance mechanisms	(Soto Setzke et al., 2019)	<i>From a policy perspective, platform owners can restrict access to resources.. by charging usage fees</i>	Platform owners can use pricing policies to restrict access to resources	Platform rules	Pricing
Governance mechanisms	(Soto Setzke et al., 2019)	<i>From a policy perspective, platform owners can restrict access to resources.. by reserving access to selected groups of developers [18, 57].</i>	Platform owners can restrict access to resources for selected groups.	Platform rules	Gate keeping
Governance mechanisms	(Farshchian & Thomassen, 2019)	Platform boundary resources are Bsoftware tools and regulations that serve as the interface for the arm's-length relationship between the platform owner and the application developer^ (Ghazawneh and Henfridsson 2013).	Definition of platform boundary resources: software tools and regulations that serve as the interface for the arm's-length relationship between the platform owner and the application developer	Definition of platform boundary resources	

Governance mechanisms	(Farshchian & Thomassen, 2019)	Examples can be an Application Programming Interface (API), or a mandatory server –e.g. provided by platform owner –to store data (B.A. Farshchian and Vilarinho 2017).	Platform boundary resource examples are APIs or a mandatory server	Platform boundary resources	<ul style="list-style-type: none"> • APIs • Mandatory server
Governance mechanisms	(Wulfert & Schütte, 2022)	Boundary resources represent a dimension of governance, defining the boundaries between the marketplace owner and the community of external participants, thus facilitating the realization of strategically relevant decisions about ownership, entry into new markets, or community building [33, 46, 62]. Dal Bianco et al. [33] differentiate application, development, and social boundary resources	Boundary resources are separable between application boundary resources, development boundary resources and social boundary resources	Platform boundary resources	<ul style="list-style-type: none"> • Application boundary resources • Development boundary resources • Social boundary resources
Governance mechanisms	(Wulfert & Schütte, 2022)	Application boundary resources (APIs, libraries, etc.) are defined as the minimum required for a software ecosystem to be viable.	Application boundary resources are APIs and libraries	Application boundary resources	<ul style="list-style-type: none"> • APIs • Libraries
Governance mechanisms	(Wulfert & Schütte, 2022)	External modules are developed using technical boundary resources provided by the innovation platform in the form of API services	technical boundary resources are APIs	Technical boundary resources	APIs
Governance mechanisms	(Wulfert & Schütte, 2022)	and providing social (e.g., documentation and technical support) boundary resources	Examples of social boundary resources are documentation and technical support	Social boundary resources	<ul style="list-style-type: none"> • Documentation • Technical support
Governance mechanisms	(Wulfert & Schütte, 2022)	Each application boundary resource is augmented by dedicated social boundary resources, such as documentation, training, or guidelines [33].	application boundary resources are augmented by dedicated social boundary resources, such as documentation, training, or guidelines	Social boundary resources	<ul style="list-style-type: none"> • Documentation • Training • Guidelines
Governance mechanisms	(Wulfert & Schütte, 2022)	Social boundary resources are used for knowledge transfer,	Purpose of social boundary resources	Social boundary resources	

Governance mechanisms	(Wulfert & Schütte, 2022)	development boundary resources for supporting application development	purpose of development boundary resources	Development boundary resources	
Governance mechanisms	(Wulfert & Schütte, 2022)	and application boundary resources for enabling interaction with a focal platform.	purpose of application boundary resources	Application boundary resources	
Governance mechanisms	(Wulfert & Schütte, 2022)	Application boundary resources (APIs, libraries, etc.) are defined as the minimum required for a software ecosystem to be viable.	Definition of application boundary resources: the minimum required for a software ecosystem to be viable.	Definition of application boundary resources	
Governance mechanisms	(Bonina & Eaton, 2020)	Boundary resources have their conceptual foundation in a synthesis of boundary object theory (Star & Bowker, 1999) and innovation networks theory (Chesbrough, Vanhaverbeke, & West, 2006). They refer to “software tools and regulations that serve as the interface for the arm's-length relationship between the platform owner and the application developer” (Ghazawneh & Henfridsson, 2013, p. 174).	Definition of platform boundary resources: software tools and regulations that serve as the interface for the arm's-length relationship between the platform owner and the application developer	Definition of boundary resources	
Governance mechanisms	(Bonina & Eaton, 2020)	In this way, they provide components of platform governance, which conceptualize the two major tasks of platform owners. First, platform owners must provide tools to help resource and support third parties develop platform services	Boundary resource consists of tools that help resource and support third parties	Boundary resources	Tools

Governance mechanisms	(Bonina & Eaton, 2020)	Examples of tools are application programming interfaces (APIs), which provide developers access to core platform modularity, and software development kits (SDKs), which facilitate developers to build software services.	Examples of tools are application programming interfaces (APIs) and software development kits (SDKs)	Tools	<ul style="list-style-type: none"> • APIs • SDKs
Governance mechanisms	(Bonina & Eaton, 2020)	which they do by providing rules for controlling the quality of third party apps and services developed for the platform.	Boundary resource consists of rules for controlling the quality of third party apps and services	Boundary resources	Rules
Governance mechanisms	(Bonina & Eaton, 2020)	Examples of rules include those contained within licensing agreement contracts (Boudreau & Hagiu, 2009)	Examples of rules are License agreements	Rules	License agreements
Governance mechanisms	(Bonina & Eaton, 2020)	We adapt the previous definition so that boundary resources now encompass “the rules and tools that serve as the interface to govern the arm's-length relationship between the platform owner and different members of the platform ecosystem”.	Definition of platform boundary resources: the rules and tools that serve as the interface to govern the arm's-length relationship between the platform owner and different members of the platform ecosystem	Definition of boundary resources	
Governance mechanisms	(Bonina & Eaton, 2020)	This new platform provided an API, as a software tool, through which developers could more easily access datasets and integrate them into innovative apps and services.	API as a tool to access datasets	Tools	APIs
Governance mechanisms	(Bonina & Eaton, 2020)	The approach we take is to extend the concept of boundary resources, as components of platform governance, to encompass the supply side of the platform as well as the demand side.	Boundary resources encompasses the supply- and the demand side to govern a platform	Boundary resources	<ul style="list-style-type: none"> • Supply side • Demand side

Governance mechanisms	(Bonina & Eaton, 2020)	The platform owner uses tools and rules to govern contributors on the supply side and developers on the demand side. It designs tools, as governance components, to resource both groups to help them carry out their respective tasks. It designs rules, as governance components, in order to regulate how both sides carry out their tasks and to control the outcome of those tasks	Use of tools and rules to govern contributors on the supply side and developers on the demand side.	<ul style="list-style-type: none"> • Supply side • Demand side 	<ul style="list-style-type: none"> • Supply side • Demand side
Governance mechanisms	(Bonina & Eaton, 2020)	Components of governance provided by the platform owner which resource and enable ecosystem members to carry out their ecosystem role.	Definition of tools: Components of governance provided by the platform owner which resource and enable ecosystem members to carry out their ecosystem role.	Definition of Tools	
Governance mechanisms	(Bonina & Eaton, 2020)	Components of governance provided by the platform owner, which constrain ecosystem members from carrying out actions that may harm a platform and thereby help to secure it.	Definition of rules: Components of governance provided by the platform owner, which constrain ecosystem members from carrying out actions that may harm a platform and thereby help to secure it.	Definition of Rules	
Governance mechanisms	(Bonina & Eaton, 2020)	the Office released and applied a set of contractual rules to control the activities of developers. These took the form of general terms and conditions introduced in 2012 and licensing terms (Creative Commons Argentina 2.51) in 2013.	Rules took the form of general terms and conditions and licensing terms	Rules	<ul style="list-style-type: none"> • Terms and conditions • Licensing terms

Governance mechanisms	(Bonina & Eaton, 2020)	the office developed information tools, which provided procedures and instruments for the construction of data sets. These included artefacts such as dataset templates that facilitated the collection and formatting of data.	Tools on the supply side in the form of dataset templates	Tools	Dataset templates
Governance mechanisms	(Bonina & Eaton, 2020)	These rules covered several areas governing dataset names, formats, sources and who was responsible for subsequent dataset updates. These technical agreements encouraged ministries to convert their data into usable OGD datasets.	Rules in the form of Technical Agreements concerning format of datasets	Rules	Format of datasets
Governance mechanisms	(Bonina & Eaton, 2020)	terms and conditions of use, which place constraints on how developers can use OGD.	Terms and condition how developers can use the platform	Platform rules	Terms and conditions
Dimensions of openness	(Broekhuizen et al., 2021)	Hence, we define platform openness broadly as the platform's openness toward granting access and authority to suppliers, customers, and complementary service providers, and toward the inclusion of categories and channels.	Definition of platform openness: the platform's openness toward granting access and authority to suppliers, customers, and complementary service providers, and toward the inclusion of categories and channels.	Definition of platform openness	

Governance mechanisms	(Staub et al., 2022)	<p>We focus on the governance mechanisms that were identified as most relevant and describe their application in the literature toward generativity and control...we present governance mechanisms in three fundamental pillars of platform governance...Following Ghazawneh and Henfridsson (2013), the first pillar refers to platform boundary resources that are provided by platform owners to complementors to leverage their development of complementary add-ons... The second pillar refers to platform rules that are established by the platform owner to define the scope of activities for each of the platform ecosystem's actors.... The third pillar refers to the shared ecosystem identity among the actors as a sense of belonging to the platform ecosystem.</p>	<p>Governance mechanisms of platforms governance to address generativity and control are concentrated around the three pillars Platform boundary resources, Platform rules and Ecosystem identity</p>	Governance mechanisms	<ul style="list-style-type: none"> • Platform boundary resources • Platform rules • Ecosystem identity
Governance mechanisms	(Staub et al., 2022)	<p>platform boundary resources that are provided by platform owners to complementors to leverage their development of complementary add-ons. The majority of corresponding studies focus on the platform's interfaces and programming resources</p>	<p>Platform boundary resources consist of Interfaces and Programming resources</p>	Platform boundary resources	<ul style="list-style-type: none"> • Interfaces • Programming resources

Governance mechanisms	(Staub et al., 2022)	platform rules that are established by the platform owner to define the scope of activities for each of the platform ecosystem's actors. The mechanisms dominantly discussed in our literature set are gatekeeping (who can have which access), decision rights (who can do which tasks), intellectual property sharing (who owns the generated innovations), pricing (what is the pricing model), and revenue sharing (who can get which share of revenue).	The governance mechanisms comprising platform rules include Gatekeeping, Decision rights, Intellectual property sharing, Pricing, and Revenue sharing.	Platform rules	<ul style="list-style-type: none"> • Gatekeeping • Decision rights • Intellectual property sharing • Pricing • Revenue sharing
Governance mechanisms	(Staub et al., 2022)	Programming resources (e.g., software tools, software development kits/SDKs, libraries, and documentation) are provided by the platform owner to complementors to help them develop add-ons.	Programming resources are software tools, Software development kits, libraries and documentation	Programming resources	<ul style="list-style-type: none"> • Software tools • SDKs • Libraries • Documentation
Governance mechanisms	(Staub et al., 2022)	platform rules that are established by the platform owner to define the scope of activities for each of the platform ecosystem's actors. The mechanisms dominantly discussed in our literature set are gatekeeping (who can have which access), decision rights (who can do which tasks), intellectual property sharing (who owns the generated innovations), pricing (what is the pricing model), and revenue sharing (who can get which share of revenue).	The governance mechanisms comprising platform rules include Gatekeeping, Decision rights, Intellectual property sharing, Pricing, and Revenue sharing.	Platform rules	<ul style="list-style-type: none"> • Gatekeeping • Decision rights • Intellectual property sharing • Pricing • Revenue sharing

Governance mechanisms	(Staub et al., 2022)	Drawn from organizational identity (Corley & Gioia, 2004), this has been frequently discussed in the existing platform research as relational control that aims to facilitate mutually shared expectations and norms among the ecosystem's actors about what constitutes a legitimate behavior in the platform ecosystem (Lindgren et al., 2015).	Ecosystem identity functions as a relational control mechanism, fostering the development of mutually shared expectations and norms among the actors within the ecosystem.	Ecosystem identity	Relational control
Governance mechanisms	(Staub et al., 2022)	Application programming interfaces (APIs) are specifically designed to "accept a broad class of apps in ways that allow app developers to use the platform's capabilities without having to concern themselves with how those capabilities are implemented in the platform" (Tiwana, 2014, p. 289).	Definition of Application programming interface (API): Application programming interfaces (APIs) are specifically designed to "accept a broad class of apps in ways that allow app developers to use the platform's capabilities without having to concern themselves with how those capabilities are implemented in the platform"	Definition of Application programming interface (API)	
Governance mechanisms	(Staub et al., 2022)	As a mechanism for generativity, interfaces grant complementors with a standardized form of platform access to sell their complementary add-ons (Bender & Gronau, 2017; Ghazawneh & Henfridsson, 2013). In this role, APIs constitute an important part of digital platforms that connects the platform core with external complementors. Also, APIs facilitate data exchange by enabling external developers to access and retrieve data and use it to build their applications (Fuerstenau et al., 2019).	Effect of mechanisms on openness: Interfaces grant standardized platform access and connection and facilitate data exchange	Effects on openness: APIs	<ul style="list-style-type: none"> • Standardized access and connection • Facilitate data exchange

Governance mechanisms	(Staub et al., 2022)	Programming resources (e.g., software tools, software development kits/SDKs, libraries, and documentation) are provided by the platform owner to complementors to help them develop add-ons.	Definition of Programming resources: Programming resources are provided by the platform owner to complementors to help them develop add-ons.	Definition of Programming resources	
Governance mechanisms	(Staub et al., 2022)	Furnishing third parties with development knowledge and tools to foster the generation of innovations is a common strategy for platform owners (Foerderer et al., 2019; Parker et al., 2017). Knowledge resources often aim to provide a holistic picture of the offered platform functionalities to support complementors in identifying possible ideas to create their complementary extensions. For example, Hukal et al. (2020) illustrate how platform owners can signal strategic interests to activate complementors to generate new platform content. Further, scholars particularly highlight the relevance of programming tools as an important means to lower the threshold to develop on the platform (Karhu et al., 2018). Various scholars discuss how a platform owner can incentivize complementors with various types of programming resources, such as Karhu et al. (2018) with regard to Google Android. Further, based on archival data from Apple iOS and Google Android, Ye and Kankanhalli (2018) show that SDKs may also be applied to positively influence complementors in innovating services (e.g., generate ideas or create	Effects of mechanisms on openness: Knowledge resources and tools support complementors in identifying possible ideas to create their complementary extensions. Programming tools lower the threshold to develop on the platform. SDKs may also be applied to positively influence complementors in innovating services	Effects on openness: Programming resources	<ul style="list-style-type: none"> • support complementors in identifying possible ideas • lower the threshold to develop on the platform. • influence complementors in innovating services

		add-ons on the platform).			
Governance mechanisms	(Staub et al., 2022)	Gatekeeping (i.e., input control or bouncer rights) refers to “the degree to which the platform owner uses predefined objective acceptance criteria for judging what apps and app developers are allowed into a platform ecosystem” (Tiwana, 2014, p. 123).	Definition of Gatekeeping: Gatekeeping refers to the degree to which the platform owner uses predefined objective acceptance criteria for judging what apps and app developers are allowed into a platform ecosystem	Definition of Gatekeeping	

Governance mechanisms	(Staub et al., 2022)	Regarding generativity, gatekeeping usually means easing access for complementors, which results in a more open platform and potentially increases diversity in offered complements (Goldbach & Benlian, 2015a). For example, Wessel et al. (2017) explore how a reduction in input control (and thus increasing platform openness) affects key performance indicators for different platform participants. In their study, they investigate a policy change on Kickstarter, a leading crowdfunding platform, where certain restrictions were reduced and platform access was facilitated. Their results suggest that increasing platform openness for complementors rapidly increased the number of crowdfunding projects on Kickstarter but project creators needed to cope with higher uncertainties due to increased competition.	Effects of mechanisms on openness: Gatekeeping means easing access for complementors for a more open platform and increasing diversity in offered complements.	Effects on openness: Gatekeeping	<ul style="list-style-type: none"> • Easing access to the platform • Increasing diversity in offered complements
Governance mechanisms	(Staub et al., 2022)	Decision rights can be defined as the division of authority and responsibilities between the platform owner and complementors. In general, it must become clear who decides about the strategic objectives of the platform or individual apps and how these objectives should be implemented (Tiwana, 2014).	Definition of decision rights: Decision rights can be defined as the division of authority and responsibilities between the platform owner and complementors. In general, it must become clear who decides about the strategic objectives of the platform or individual apps and how these objectives should be implemented	Definition of decision rights	

		<p>According to Ye and Kankanhalli (2018), platform owners should leave complementors sufficient decision rights to ensure their own decision-making autonomy, which ultimately influences the resulting innovation output (in terms of quality and quantity). Further, existing research highlights that platform owners should practice rules with both flexibility and benevolence at the same time in order to maximize complementor dedication (Hurni et al., 2021). As the platform evolves, platform owners may continuously reconfigure decision rights to incentivize complementors in different ways (Sandberg et al., 2020).</p>	<p>Effects of mechanisms on openness: Platform owners ensure decision-making autonomy of complementors to increase their innovation output</p>	<p>Effects on openness: Decision rights</p>	<ul style="list-style-type: none"> • Ensure decision-making autonomy of complementors to increase their innovation output
Governance mechanisms	(Staub et al., 2022)	<p>Intellectual property sharing can happen in two different ways (Niculescu et al., 2018): via direct inter-firm agreements (licensing) or by making a technology accessible to the general public (open source project).</p>	<p>Definition of Intellectual property sharing: Sharing of Intellectual property can happen via direct inter-firm agreements (licensing) or by making a technology accessible to the general public (open source project).</p>	<p>Definition of Intellectual property sharing</p>	
Governance mechanisms	(Staub et al., 2022)	<p>As a mechanism for generativity, platform owners can exchange intellectual property to increase complementors' security to be able to materialize and sell their own offerings (Niculescu et al., 2018).</p>	<p>Effects of mechanisms on openness: platform owners can exchange intellectual property to increase complementors' security to be able to materialize and sell their own offerings</p>	<p>Effects on openness: Intellectual property sharing</p>	<ul style="list-style-type: none"> • Attract more complementors by expanding their intellectual property rights

Governance mechanisms	(Staub et al., 2022)	Pricing policies are used by the platform owner to create incentives for complementors to make personal investments to ensure the prosperity of their own offerings and in turn the whole ecosystem. This includes, for instance, the app pricing model, pricing symmetry, and the selection of a subsidy-side (Tiwana, 2014).	Definition of pricing: pricing policies are used by the platform owner to create incentives for complementors to make personal investments to ensure the prosperity of their own offerings and in turn the whole ecosystem. This includes, for instance, the app pricing model, pricing symmetry, and the selection of a subsidy-side	Definition of pricing	
Governance mechanisms	(Staub et al., 2022)	Pricing is primarily discussed from a generativity perspective. For instance, platform owners can choose to subsidize one, highly valued side of platform participants by granting them free or inexpensive platform access over a certain time period (Thies et al., 2018). This is of particular relevance in the case of consumer platforms, as they often need to reach a critical mass of users to remain competitive. ...Further, platform owners of C2C sharing platforms in monopolistic settings should employ accurate pricing strategies that incentivize both of the platform's sides (i.e., demand and supply) to foster interactions on the platform (Zimmermann et al., 2018)	Effects of mechanisms on openness:	Effects on openness: Pricing	<ul style="list-style-type: none"> • Employ subsidies for one group of platform participants to achieve the critical mass of complementors and users • Accelerate the adoption of the platform among complementors and users
Governance mechanisms	(Staub et al., 2022)	Revenue sharing represents the degree to which the platform extracts revenue that is co-created with the complementors (Oh et al., 2015).	Definition of revenue sharing: Revenue sharing represents the degree to which the platform extracts revenue that is co-created with the complementors	Definition of revenue sharing	

Governance mechanisms	(Staub et al., 2022)	<p>Revenue sharing is primarily discussed with a focus on generativity. A platform owner extracts a part of the co-created value, such as a percentage of sales or service use (Karhu et al., 2018). A main research interest lies in the distribution of revenue shares that ensures complementors are encouraged to develop their high-quality complementary products/services on the platform (Oh et al., 2015)....Platform owners may also adjust the revenue shared with complementors during the evolution of the platform ecosystem or depending on the complementor type (e.g., complementors that are only active in the respective platform ecosystem vs. complementors that are active in several competing platform ecosystems).</p>	<p>Effects of mechanisms on openness: By revenue sharing a platform owner extracts a part of the co-created value that ensures complementors are encouraged to develop their high-quality complementary products/services on the platform</p>	<p>Effects on openness: Revenue sharing</p>	<ul style="list-style-type: none"> • Ensure complementors are motivated by revenue sharing.
Governance mechanisms	(Staub et al., 2022)	<p>Relational control, as informal control, refers to “the degree to which the platform owner relies on norms and values that it shares with app developers to influence their behavior” (Tiwana, 2014, p. 125). Relational control can be divided into self-control (e.g., set your own goals, monitor and sanction or reward yourself) and clan control (Ouchi, 1979).</p>	<p>Definition of Relation control: Relational control refers to the degree to which the platform owner relies on norms and values that it shares with app developers to influence their behavior and can be divided into self-control and clan control</p>	<p>Definition of relation control</p>	

Governance mechanisms	(Staub et al., 2022)	Through shared norms and values (formed by knowledge exchange and learning from each other), a strong community among complementors is likely to enhance their performance through fewer mistakes and less rework. A shared vision and similar ambitions, for instance concerning design and security standards, strengthen a common understanding and bring complementors closer together (Goldbach & Benlian, 2015a). This community feeling can in turn attract new complementors.	Effects of mechanisms on openness: Relational control enhances complementor performance through fewer mistakes and less rework. A shared vision and similar ambitions brings complementors closer together and attracts new complementors	Effects on openness: Relational control	<ul style="list-style-type: none"> • Enhances complementor performance through fewer mistakes and less rework • Brings complementors closer together and attracts new complementors
Governance mechanisms	(Vesselkov et al., 2019)	An application programming interface (API) constitutes a means to allow third parties to access an organization's capabilities or data (Spencer, Krohn, Fisher, & Boyd, 2014)	Definition of Application programming interface (API): constitutes a means to allow third parties to access an organization's capabilities or data	Definition of Application programming interface (API)	
Governance mechanisms	(Vesselkov et al., 2019)	platform providers must carefully design the platform API and other boundary resources.	API is a boundary resource	Boundary resources	APIs
Governance mechanisms	(Vesselkov et al., 2019)	Some mHealth APIs adopted a paid model. Thus, Garmin charged a one-time license fee (US\$5,000) for access to its Connect API, although it shared less granular data for free through its Garmin Health API. Similarly, Samsung charged for the access to its REST API, although we could not locate any publicly available pricing details.	Providers use a pricing model for access to the API	Platform rules	Pricing
Governance mechanisms	(Vesselkov et al., 2019)	On the other hand, some data providers may decide to share the revenue that data	Providers can use Revenue sharing that data consumer	Platform rules	Revenue sharing

		consumer stimulate when using the API (Decision 15)...Thus, revenue sharing does not exist in current mHealth data-sharing platforms.	stimulate by using the API.		
Governance mechanisms	(Engert et al., 2022)	PBRs are defined as “the software tools and regulations that serve as the interface for the arm’s-length relationship between the platform owner and the application developer.	Definition of platform boundary resources: software tools and regulations that serve as the interface for the arm’s-length relationship between the platform owner and the application developer	Definition of platform boundary resources	
Governance mechanisms	(Engert et al., 2022)	Prior studies identified different PBRs and distinguished between application boundary resources, development boundary resources, and social boundary resources (Bianco et al., 2014; Petrik & Herzwurm, 2020a).	Platform boundary resources can be distinguished between application boundary resources, development boundary resources, and social boundary resources		
Governance mechanisms	(Engert et al., 2022)	Application boundary resources allow third-party applications to connect with the platform core, including APIs for accessing specific data (Grzenda & Legierski, 2021).	Application boundary resources allow connection to platform by APIs	Application boundary resources	APIs
Governance mechanisms	(Engert et al., 2022)	Development boundary resources provide the means to developers to develop their applications, such as SDKs or debugging tools.	Development boundary resources comprise of SDKs and Debugging tools	Development boundary resources	<ul style="list-style-type: none"> • SDKs • Debugging tools
Governance mechanisms	(Engert et al., 2022)	<i>PBRs to increase the transparency of regulations for complementors... such as terms and conditions agreement,</i>	Terms and conditions are a form of regulations	Platform rules	Decision rights
Governance mechanisms	(Engert et al., 2022)	These are supported by social boundary resources which comprise documentation, support contacts, or developer forums (Bianco et al., 2014).	Social boundary resources comprise of documentation, support contacts and developer forums.	Social boundary resources	<ul style="list-style-type: none"> • Documentation • Support contacts • Developer forums

Governance mechanisms	(Engert et al., 2022)	Boundary resources are the interfaces, tools and rules used to enable, facilitate and control an arm's-length relationship between a platform owner and third-party complementors.	Definition of Boundary resources: Boundary resources are the interfaces, tools and rules used to enable, facilitate and control an arm's-length relationship between a platform owner and third-party complementors.	Definition of boundary resources	
Governance mechanisms	(Urovi et al., 2022)	One smart contract is generated per published dataset. i) The smart contract keeps track of the access rights to the corresponding dataset. No data can be accessed without obtaining permission via the interaction with the smart contract.	Access rights in contract, where without permission no access to data	Platform Rules	Access rights

Appendix 4: Invitation letter expert interviews

Dear

I hope this message finds you well. My name is Floris van der Krieken, and I am currently pursuing my master's degree in Business Process Management & IT at the Open University in the Netherlands. I am reaching out to you due to your recognized expertise in the field of data platforms, as I am conducting research on governance mechanisms that facilitate openness in data platforms.

The primary objective of my research is to gain in-depth insight into the governance mechanisms that enable openness of data platform ecosystems. In addition, I am particularly interested in understanding why these governance mechanisms are effective in achieving openness in data platforms.

I am writing to request your participation in an expert interview, which would greatly contribute to the insights needed for my research. The interview is expected to last approximately one hour and will be conducted online, at a time that is most convenient for you.

I look forward to your response. Thank you in advance for your consideration of my request.

Best regards,

Floris van der Krieken

Appendix 5: Interview protocol

The purpose of this interview is to collect information on governance mechanisms that enable openness in data platforms. The goal of this interview is to get an in-depth understanding of the mechanisms used to enable openness and the reasoning behind it. Moreover, the interview aims to identify data-specific mechanisms used to manage data platform openness and their underlying reasons. The interview starts with some general questions, from where we work towards the core of the interview.

I will treat your answers with confidentiality and process them anonymously in my report. The interview takes approximately one hour. Do you have an objection to recording the interview, in order to increase the reliability of the research? Do you have any questions before we start the interview?
Introduction questions:

1. What is your job function?
2. Can you describe the platform of the organization you are working for?
If not a data platform, describe that the specific context of the research is about data platforms.
3. How many years of experience do you have with data platforms?
4. Can you explain what kind of experience you have with data platforms?
5. What governance mechanisms have you observed being used in data platforms to facilitate openness?
6. How do these mechanisms work, and what specific roles do they play in enabling openness?

Main questions:

Platform boundary resources	<ol style="list-style-type: none"> 7. Can you provide insights into the use of application programming interfaces as a governance mechanism for enabling openness in data platforms? 8. How are technical development boundary resources, such as software development kits or tools, used as a governance mechanism to promote openness in data platforms? 9. How are non-technical social boundary resources, such as documentation, training, or support, used as a governance mechanism to promote openness in data platforms?
Platform rules	<ol style="list-style-type: none"> 10. How do intellectual property rights, such as licensing, play a role in governing openness within data platforms? 11. Could you elaborate on the use of decision rights, such as terms and conditions, as a mechanism to enable openness in data platforms? 12. Can you provide insights of gatekeeping practices, such as access rights, in data platforms and their impact on enabling openness? 13. How do pricing mechanisms contribute to promoting openness in data platforms? 14. Can you explain how revenue-sharing models are implemented in data platforms to enable openness?
Ecosystem identity	<ol style="list-style-type: none"> 15. Can you explain how relational control is utilized to foster openness in data platforms?
Data-specific mechanisms	<ol style="list-style-type: none"> 16. What are data-specific mechanisms used to manage data platform openness? 17. How do data-specific mechanisms contribute to data platform openness?

Closing:

18. Is there any other governance mechanism related to platform openness that we haven't discussed yet?
19. Are there any additional insights or perspectives you would like to share regarding the mechanisms we covered?
20. Are there any missing elements you would like to add before we conclude the interview?
21. Do you have any recommendations or suggestions to improve the openness of data platforms based on your expertise?

Appendix 6: Overview of participants

#	Job function	Platform description	Years of experience with data platforms	Kind of experience
Int_1	Founder and CEO of financial data platform in the US	We sell financial data feeds covering a variety of asset classes ranging from stocks to options to ETFs, mutual funds and a variety of other data like analyst estimates and ESG data, primarily through an API and a web socket	11 years	I have a significant amount of experience with exchange data market data, the flow of real time you know securities information through exchanges and other government regulatory bodies that control that flow of pricing, public market data. I have a lot of experience with data marketplaces.
Int_2	Principal Consultant at a technological consultancy firm	So per European project there are different platforms that are being used and different standards that are that are being used. And that said in the the whole point of the Data spaces is that we become less dependent of those big three and develop our own data platforms and and they are called data spaces and those built on.	10 years	Performing research. I have worked also mainly on the research into architecture but also into data management and data regulation.
Int_3	Former CEO of a Data marketplace in Canada	So the platform of Company_Int_3 is a SaaS solution that provides every tool needed to be able to take content data that you have and use it in one of three use cases. The first one is packaging the data as products and selling it. The second use case is external data sharing. And then by sharing it, there's some additional calculations and derived contents that's created and that feeds back their risk systems for compliance. And then the third use case is for internal data marketplaces.	30+ years	Product management and general management.
Int_4	Assistant Professor at a Dutch university	So I worked with Narrative, which is one of the data marketplaces in the US. What happens on this marketplace is that there are buyers and sellers who come to this marketplace. Sellers, you can think of companies like MasterCard, visa, Uber. They they have a lot of data with them and they want to monetize it. So they come to this platform on the other hand, on the other side, you have buyers who are interested in such data points. What this form does is essentially facilitates the trade between buyers and sellers.	5 years	So mostly what I do is that I talk to these data platforms, I look at their operations that how they are doing their businesses and eventually my goal is to look at academic problems that I can solve. So I sort of motivate my academic research by looking into the operations of these data platforms. So the problem is motivated from these platforms. Then once I am have modeled the problem I go back and I see whether I have I have accurately modelled the problem and once everything is done, I also go back and give them their feedback, give them the feedback that you know what insights I obtained from solving this problem. And so that's sort of an academic industry collaboration that that usually happens, which is there.
Int_5	Chief Science Officer at a Dutch ministry & Professor at the Rotterdam University of Life Sciences	I am not working with data platforms, we are developing and implementing data platforms for our research, so I'm not using a specific data platform, but one of the things that we are doing, we are trying to design a data platforms and not that you can use it directly, but prototypes of these platforms and then we can see where the bottlenecks are, how the research questions are and if you are trying to develop and to design these kind of platforms for example, well, what are the data sharing issues on such a platform? What are the data quality issues if you bring data from different sources together in one?	30+ years	Well, what we are working on and if you have these data platforms and the idea behind this data platforms is that you have data from different organizations, different sources, you bring them together in one platform, so such that people other people can use the data for their purposes. So what we are working on is saying, OK, if we bring this data together, how can we share this data to other parties in a in a privacy proof way if we at least in governmental data, you cannot share with everybody. You can share it, but you have to obey the rules and regulations. So that's the thing that we are looking at. And beside that we are looking if we bring data from different sources together, sometimes they have some overlap.
Int_6	CEO/CTO of multiple tech companies in Belgium	So one of them is (Client), it's called uh, it's in Belgium. They are in Antwerpen and Gent, they are practically doing I think 70% of the municipalities in Belgium. And they have a platform and it's called (Platform name). So they were focusing on creating the creation of building blocks. When you have building blocks, it means that you scope each requirement or set of requirements to a specific supplier and they have a focus and they are good at it. For example, you have a supplier that's dealing with uh, culture, and so the culture houses access for that	14 years	I had to experience in the beginning with ECM's, so content management systems like Documentum and all fair school, those kind of those domains where you have to also uh, create knowledge bases for specific customers and and to create a semantical value on top of the data. I have experience with time series data a lot where you need to measure stuff and also statistically uh provide insights in data. I have experience in benchmarking and ML where you need to for example benchmark machines like high pressure machines or heat pumps and stuff like that. I have data, experience and governmental data, but also in security documents where you need to provide verifiable claims for a

		<p>one. One that deals with asset management, digital asset management, one that can deal with IoT platforms and how to work with electrical charges for example. And because you have very broad domain of requirements, they went and they defined or they would like to have and they have it for now. They would want to have an API ecosystem. OK, so microservice ecosystem and their and their in their space and that's that's that comes along with a lot of data.</p>		<p>specific attribute or some traits that the individual or a company has. Most of the time I always have a technical architectural role as well in those domains, so most of the times when I did something and the the ones I mentioned to you I had also I was also the technical architect of those.</p>
Int_7	CEO of a Dutch data engineering company	<p>For example, for (Client) we build a platform for that. So in this case data providers are a number of parties that supply product data and product data that sounds a bit simple, but it is not. So they are all kinds of codes for products and you would be surprised how many different types (Client) you can have, but every time, imagine that something changes in the bottle or label. So they have a kind of database with all codes of, well, what is connected. Then you also have laboratory data that have data that test, with those products on those product codes and what they contain.</p>	10 years	<p>So yes, on the one hand I did the data architecture. That really involves designing database systems, so source systems, but also designing data warehouses, so star models based on certain source systems. I haven't worked much with standard pieces of software. I have actually worked much more at a low level, so directly in databases.</p>
Int_8	Partner at a European consultancy firm	<p>for a large construction company in the Netherlands, we have created modern data stack, where we collect data from all their source systems, but also data from outside the organisation. Where we collect data from all their source systems, but also data from outside the organisation, so third party. So the data that is actually not created by them and we combine that data within a central data platform and provide reports on top of it with well dashboards created in Power BI. The data platform itself is created with a medallion architecture. So that means that when we connect the data to the source system, we collect all the data in its raw format, store it in the data lake and make sure that there is a sort of a history mechanism on the lake itself.</p>	20+ years	<p>I've done several roles. So first creating actually the data platform, so more on the execution side that was creating the data platform, making sure that data was connected to the data platform then storing the data and then distributing the data, so creating well working actually as an ETL developer or as the data platform engineer and later on I was also more in the design activities. So creating data platforms, designing it, advising the customer what kind of data platform they should use.</p>
Int_9	Data platform lead at a Dutch maritime contracting company	<p>So we built a central data platform from IT. So from a central data analytics department within IT. What we are currently facilitating on the platform is making sure that the BI use cases and the BI needs of the organisation can be fulfilled. So currently there's no capability for centrally ingested. There's no there wasn't a central data warehouse yet for example. So we've built a data warehouse component on our data platform to make sure that, yeah, the various business teams first focusing on sort of the usual suspects in BI like the finance department, the HR departments make sure that they have the right management insights into the data that we already have but on the other hand also creating reports on for example, the performance of powerful or fleet.</p>	4 years	<p>And after that I my experience is mainly in building data platforms and building the pipelines that run on the data platforms. Mainly focusing on the Yeah, the, more, the, the the clouds infrastructure, the clouds engineering part of data platforms and the ingestion part of data platform. So really the things which you can apply engineering to more of the software engineering part of data platforming as opposed to the more data modelling side of of data warehousing. For example, on a data platform.</p>
Int_10	Head of tech data platforms of a Dutch supermarket chain	<p>It is basically where you can propagate data through the different layers, right? So the the, the, the medallion architecture. Let's say that's what we have chosen where you can take it from source to a clean layer or a goal layer where you're preparing data products. So how do you get your data from the different sources and make it available for consumption is the trick of the data platform.</p>	14 years	<p>So I've done everything from building it to migrating it or running a team that's building it. I've been an architect, I've been a solution architect or a product architect as you would call it, because you're building across the thing. I have been an enterprise, integration and data architect, so that was more on the policies and principles around it. Of course I I didn't start there, but I started as an engineer who was actually doing data warehouse migration or let's say SQL Building SQL Up to that level, and also being a what you call a manager or a director or whatever, now who's maintaining or managing a team that is building it.</p>

Appendix 7: Interview transcriptions

Transcript Interview 1

25 October 2023

00:00:05 Interviewer

OK. Do you have any questions before we start?

00:00:09 Interviewee

No.

00:00:11 Interviewer

No? OK, what is your job function?00:00:15 Interviewee

I am the founder and CEO of Company_Int_1, a financial data company.

00:00:20 Interviewer

Mm-hmm. And can you describe the platform in Company_Int_1?

00:00:27 Interviewee

Yes. So we sell financial data feeds covering a variety of asset classes ranging from stocks to options to ETFs, mutual funds and a variety of other data like analyst estimates and ESG data, primarily through an API and a web socket. So we deliver the data.

00:00:47 Interviewee

To innovators and builders who are using the data to display inside.

00:00:52 Interviewee

Kind of investing tools and websites and apps. And so the mission of the company is to use that data and open that data to help power innovation in the financial services space. So our data kind of comes alive at the hands of our developer clients inside of really cool innovations from really big companies or really small companies. The the whole gamut.

00:01:12 Interviewer

Uhm, does the platform provides modules, uh, besides trading the data? For example analytical modules and can you elaborate on that?

00:01:25 Interviewee

Yeah, we, we do have what we call some derived data sets where we merge data together. A very simple example of that would be like a price to earnings ratio. So we mix fundamental financial statement data with market data and we do we do some analytical data feeds, but the product is on purpose exclusively API data feed.

00:01:46 Interviewee

Access to data. A lot of our larger competitors who have been very entrenched in this space for a long time sell both.

00:01:55 Interviewee

API feed access to data and display kind of terminal workstation software which creates a a very large conflict of interest because they're afraid to open up their API's to developers who then might use the data to build a better software or better workstation terminal. We don't ever wanna compete with our own customers and what they're building.

00:02:16 Interviewee

And so we're very much working on just being a provider of the data and empowering our customers to build whatever they want with it. So we don't end up having a conflict of interest because a lot of our customers are doing analytics and building dashboards and.

00:02:29 Interviewee

Things like that.

00:02:30 Interviewer

So then the the conflict of openness and control becomes even bigger, you say?

00:02:35 Interviewee

Ohh yes yeah, yeah.

00:02:36 Interviewer

OK.

00:02:37 Interviewer

Interesting. OK, how many years of experience do you have with data platforms?

00:02:46 Interviewee

What year is it 2023?

11.

00:02:50 Interviewer

OK. And uh, can you explain what kind of experience you have with data platforms?

00:02:56 Interviewee

Yeah. UM, I've got illegal experience with screen scraping and just trying to pull data from literally 10,000 different platforms to try to get all the data I needed when we were first building our company and and then I have a Lot of deep, deep experience with a an open data mechanism language called XBRL and I'm not sure if that's been part of your research. If not, probably should.

00:03:24 Interviewee

And also significant amount of experience with exchange data market data, the flow of real time you know securities information through exchanges and other government regulatory bodies that control that flow of pricing, public market data. I have a lot of experience with data marketplaces.

00:03:44 Interviewee

And and and especially as with respect through through technology providers like databases and ETL marketplaces, how all that connects together as well with the variety of places people can access and buy data places where it's free and places where it isn't free.

00:04:05 Interviewer

Perfect.

00:04:07 Interviewer

What governance mechanisms have you observed being used in data platforms to facilitate openness?

00:04:16 Interviewee

The most important one for us is XBRL. It's have you do you have experienced XBRL or you have you researched that yet? OK, so XBRL is a programming language that's built off of XML, so it looks very similar to XML. It stands for extensible Business reporting language.

00:04:36 Interviewee

Uhm and it is a way to take data that the public needs and is entitled to, but is in a format that historically was filed in kind of PDF form which is not structured, not digital, not easy to access, and so the XBRL mechanism and language.

00:04:56 Interviewee

It's a standard that was built to make data more accessible. Data that that otherwise would be at a PDF and is very are hard and expensive to get access to in a systematic way. So historically a good example of the types of data that's filed this way is just financial statement data, public company financials in the United States. In 2006 the US SEC mandated the use of XBRL publicly or sorry 2009, and now public companies have to file all this information publicly in XBRL. So instead of putting the data on a PDF that nobody can read or access, they have to code their financial statement.

00:05:37 Interviewee

Data and make it make it accessible on this database. That's the first mile from zero to 1 and saying let's take this public data and actually make it structured and digital so it's accessible. The the second part though is that the taxonomy for data tags in XBRL has like 75,000 items, because a financial statement has tons of line items, revenue, operating revenue, expenses and accountants can kind of just make up whatever it's like the, the the word extensible in XBRL which means they can just kind of you can make up your own tags. And so the data is not really it's it's good that it XBRL makes it structured.

00:06:15 Interviewee

And digital, but it's not comparable because companies change their financial statements over time. If companies file their financial statements differently. So the tags, the XP are all tags are all over the place.

00:06:25 Interviewee

That's what we use machine learning to clean, so we wouldn't even be able to do that if XBRL didn't exist in the 1st place, because we'd have to scrape off of a PDF and find a way to categorize and structure the data. So there's a nonprofit organization, XBRL Global. There's chapters in a bunch of different countries. I'm actually not sure about the Netherlands.

00:06:25 Interviewee

That's what we use machine learning to clean, so we wouldn't even be able to do that if XBRL didn't exist in the 1st place, because we'd have to scrape off of a PDF and find a way to categorize and structure the data. So there's a nonprofit organization, XBRL Global. There's chapters in a bunch of different countries. I'm actually not sure about the Netherlands.

00:06:44 Interviewee

I know that the European Union mandated XBRL in 2000?

00:06:52 Interviewee

I have to look it up. I can't remember the year, but the EU mandated the use of XBRL for all the EU filing countries, and so there's going to be a repository where now in the Europe, all this data is gonna be filed in a structured way through XBRL. Asian countries have been doing this for like 20 years already. They were first, then the US and now Europe is coming.

00:07:11 Interviewee

But the idea is otherwise. These large vendors hire teams of 10,000 people to go out and manually scrape off of PDFs, and it's an expensive way to make that data available to end users. And it's really should be a free resource. I mean, it's if you wanna buy stock in Tesla.

00:07:31 Interviewee

You have the right to know what their stock price is, and you have the right to know what their revenue looked like last year. And it's just, I mean, technically you can go look at it on the Internet, but you can't really systematically pull it.

00:07:46 Interviewee

Use it. It's it's not actually usable for for end users and so I was actually part of writing some legislation in the United States.

00:07:56 Interviewee

That worked on expanding the types of data that we file in XBRL. So in the bond market in the municipal bond market in the US, the.

00:08:06 Interviewee

The data is filed on a 300 page report. If you wanted to buy a bond in the United States, it's just a nightmare. And so we wrote legislation just in the state of Florida.

00:08:17 Interviewee

In the United States. So it's a very small start, but California is now also doing this and it's starting to move in that direction. So I actually wrote the legislation that got passed to mandate that in the state of Florida. So the idea is we have to start small, but data on PDFs isn't good enough. It's just like you have to give the public better.

00:08:38 Interviewee

Better information than that. It's and we've done a lot of research around how structuring and and digitizing data.

00:08:46 Interviewee

In that way actually helps save the economy like a lot of money, there's some really interesting research in Australia about how what the impact of implementing this data standard was for them, but it's saved their their I can't remember the number and I wanna give you the wrong number. You can probably find it online but it saved their economy just hundreds of millions of Dollars because you're not using paper, you're not using PDF. People can get access to information more that actually lowers the cost of capital and ends up saving local governments money and reducing taxes and all kinds of things. And so it seems like a small thing to digitally file this public information, but it actually has very large ripple effects. So.

00:09:29 Interviewee

We're very grateful for XBRL. It's the only reason that we're able to take the next mile and making the data more available to Users.

00:09:35 Interviewer

Mm-hmm.

00:09:37 Interviewee

But it's not great yet, right, like. And you know, I'm very US centric. I'm very focused on working with the SEC here in the United States, for example. And I don't know a lot about the, the, the rest of the global community. But in the US, I mean, the SEC doesn't even know what an API is. So it's kind of tough working with them, right.

00:09:55 Interviewee

It's like XBRL was this big, innovative thing they did, and it's like, great. Well, now, you know, it also would be nice if you just had an API like available on the government website and.

00:10:04 Interviewee

It's like it's like a 20 year uphill battle to even try to get them. It's the government here, right? I mean, it's tough. So. So we just that's why we built we we built the API, we did that we were like give us the XBRL data, we'll clean it, we'll use machine learning because we can't we couldn't otherwise. So thanks for taking it that first step.

00:10:25 Interviewee

But we'll build the API on our own and and so we're in, in a weird way, we're filling that role that really should be the government, right? They should be the one making that data more accessible. So.

00:10:35 Interviewee

So that's the the the main data standard, and like XBRL, has coined the term open data like they're they're really like, they're that's kind of how they market themselves. And so they're very much in this space, I think is a good example for you and and then.

00:10:52 Interviewee

The other area that we touch quite often is a little bit more of a governance issue, I would say, which is that in the US. Actually, there's a guy that you really should if if this specific topic is interesting to you, there's other guy you should interview who has done a lot of work with the government on this in the United States. But there was legislation passed called the Dodd Frank Act in the US.

00:11:19 Interviewee

A long time ago and there was a loophole, as there always is in the United States, there was a loophole that made all of the stock exchanges self regulated.

00:11:29 Interviewee

Which is insane. I mean, name a financial institution that's not regulated. The banks are all regulated. It's like these are huge, multibillion dollar exchanges who just get to do whatever they want and charge whatever they want. And so there's this philosophical debate about, you know, moving away from the financial statement, data and XBRL and into market data.

00:11:50 Interviewee

Yeah, there's a philosophical argument, but that's public data. I mean it's, it's the, it's public markets where I'm buying and you're selling at a certain price. And those exchanges charge inordinate amount of money for for access to that data. And so we're a player in that space and we have to partner with those exchanges.

00:12:11 Interviewee

And pull data from them. But we have to pay a ton of money to be a distributor of the data and then our clients have.

00:12:17 Interviewee

To pay us.

00:12:17 Interviewee

And pay them and it's just like if you don't have \$10,000 a month us to pay for data, you're not going to get accurate stock price information.

00:12:28 Interviewee

Even on the Internet, looking at Yahoo Finance, it's not up. It's not really real time. It's not really the actual accurate best bid or offer real time data. UM and so I think there's a lot of problems with.

00:12:43 Interviewee

That legislation, but it's tough because it's, you know, there's the exchanges are.

00:12:49 Interviewee

They have board seats and they know the people at the SEC, and there's a lot of politics involved with it, and they make a lot of money that way. And so you can't just shut off their multi billion dollar revenue stream. I mean it's a tough, it's a tough philosophical battle, but there are huge, huge issues in the US with the affordability and accessibility of data because of the government.

00:13:14 Interviewee

So that those are the two areas I deal with, mostly XBRL financial statement data and how that can expand globally and open data more at the government level and market data regulations.

00:13:27 Interviewer

UmM and you uh touched APIs already. Uhm, how does that play a specific role in enabling openness? So I did preliminary research and noticed that Application programming interfaces are one of the governance mechanisms to enable openness. Can you provide insights into the use of API as a governance mechanism?

00:14:01 Interviewee

Yes, absolutely. You actually just touched on something when I was talking earlier about the conflict of interest with API's and some of the bigger vendors, which I think is important to note and that could be a really interesting part of your your thesis because.

00:14:22 Interviewee

I mean API's are are they're not that new but they for a lot of people they are the SEC still doesn't know what they are. So it depends I guess you talked to but the I think of APIs as the foundation for innovation especially in the financial services space but you know right there's tons of other type of data that can be opened to insurance.

00:14:44 Interviewee

Data environmental data. I mean it just.

00:14:46 Interviewee

Goes on and on. I'm very focused on the financial side of things.

00:14:50 Interviewee

But you can't really build new ecosystems and innovate without access to data. I mean, and it's the foundation of finance. It's the foundation of insurance. It's the foundation of real estate. It's the foundation of the environment. It's the foundation of everything. And the APIs are really the mechanism to get that data flowing the harder it is to access data the less innovation is gonna happen and moving all of these industries forward, and there's new technologies now that are making it easier to access the data. The API is the easiest possible way. I mean when you come to our website, you get an account, you get API keys and within a minute and a half, you can be pulling data down from our from our databases. I mean, it's extremely fast. You don't have to buy a Bloomberg Terminal and wait for it to get installed.

00:15:38 Interviewee

You know and.

00:15:39 Interviewee

There's even faster ways to do it too, with companies like Snowflake now and their and their newer technologies. But there is also a level of fear.

00:15:48 Interviewee

With API's that I think a lot of larger traditional institutions have, in my opinion they're stifling innovation. They're anti open data because they make a lot of money because of their their Moat that they have around it and and that that conflict of interest is with is an interesting point with APIs because.

00:16:09 Interviewee

Bloomberg wants you to buy their Bloomberg Terminal, but there is now this whole new generation of data consumers who don't wanna just look at the data inside of a terminal they want. They want the data they wanna access. They wanna do something with it. They wanna analyze it in their models, run it through AI, and they want to display it on websites. They want to do things and build things with the data.

00:16:29 Interviewee

I always say if you're if you want to look at the data, go to Yahoo finance by a Bloomberg Terminal. Were not for you if.

00:16:35 Interviewee

You want to build something with the.

00:16:38 Interviewee

You want to build your own model, your own analysis. Then you need an API. That type of usage is just exploding right. Everybody's a retail, everybody's a trader and everybody started investing in the stock market. Fin tech is exploding. Everybody's building apps and things like that. That's not gonna stop. But there's really five companies globally that own over 70% of that Financial data market and Bloomberg is one of them. They're the easiest prey to talk about because they're such a well known brand.

00:17:05 Interviewee

But they they do sell Open Access to data through these APIs. Now I say open, but they're still rules, right? Like you can't resell our data. You can pull the data down via API and integrate it and display it and show it and use it, but you can't turn around. Start your own data business and you can't resell. But there are rules. It's open, but it's not that open.

00:17:25 Interviewer

I come to the rules in in in three questions like platform rules is next on the topic.

00:17:32 Interviewee

Yeah, yeah, yeah. So essentially, I mean the point that I already kind of talked about was.

00:17:38 Interviewee

If you open an API to somebody to allow them to have real Open Access to this data, they could build the Bloomberg Terminal with the data they could build their own version of the Bloomberg Terminal with the data. So I do think it's a really in in this segment. The financial segment, it is a huge problem that the companies that own all of the data are afraid to open it up via APIs.

00:17:59 Interviewee

That's that's a a conflict for sure.

00:18:03 Interviewer

OK. And and how are technical development boundary resources such as software development kits or tools used as a governance mechanism to promote openness?

00:18:15 Interviewee

Yeah, I think that that when it comes to those resources, it's all about the speed and and openness of access. So API is the first step. You have to have the delivery mechanism, but the education and the usability around the delivery mechanism, if that's not there, it's not really open. If you can't use it easily. Is it really open? That's where I think the SDK's and software development kits come in. Having documentation, tutorials, guides, resources, things like that. There are a lot of developers who just need the API and they're off to the races. They can go, but that that's not not always that easy. Not every API looks the same way. You have to learn the syntax and and learn everything, so that's a huge part of what we do at Company_Int_1 is make that easier. We think of ourselves as like.

00:19:02 Interviewee

When we first were researching the data space, we got app trial access to Bloomberg's API and it wasn't documented. So like forget SDK is they didn't, they didn't even know what a software development kit was. They didn't have.

00:19:15 Interviewee

API documentation we had the how do we use it?

00:19:19 Interviewee

I mean, it was insane. They there were no resources around it at all.

00:19:22 Interviewee

So we have SDK's in every major language we have GitHub, we have tutorials we have how 2 guides we have videos, we have everything that you would need to get free, fast access to the data. So I think those resources are critical.

00:19:38 Interviewee

Forget about the government providing them, though. That's the thing, right? I mean, I don't. I don't know that at that level it's ever gonna be realistic, but it is. It is all about the usability and the question of.

00:19:51 Interviewee

What? What does open data mean in accessible data mean? Well, it has to be easy to use and fast. And I think that's where the resources are important.

00:20:01 Interviewer

Would you place non-technical boundary resources such as documentation, training, support and technical uh another play under the same number on like do they have the same effects to open up data platforms?

00:20:21 Interviewee

I think it depends on the user Uhm.

00:20:24 Interviewee

It depends on the technicality of the user. We have very technical engineers who use our data, who don't care about the written resources. They just want the code and then we have some other users who are more in the business side of things that need the rest of it. I do kind of bundle it together because you often have a business oriented person and a technical oriented person working on the same project and they wanna know different things, but there's a lot of crossover such that I think I bundle it all together.

00:20:54 Interviewer

OK.

00:20:57 Interviewer

Now to the platform rules. How do intellectual property rights, such as licensing, play a role in governance governing openness within data platforms?

00:21:09 Interviewee

That's a good question too. I think there's a lot of ways to tackle that one, we.

00:21:15 Interviewee

We have a set of terms of service on our website for our API and it mostly is just don't resell our data right? I mean could there be people who are pinging our API and just storing everything in a database and creating their own asset? Yes. But we've chosen to be an API company, you know, and that's just part of the game and.

00:21:34 Interviewee

And I think there's pricing models in our industry that differentiate between historical data and real time updated data. You kind of let go of your historical data when you sell it, you often sell it as a bulk download as CSV, something like that. You charge an upfront fee. It's not a subscription model, right. If you want recurring access to the data, the licensing changes it's you have to keep paying us to keep getting updates on the data.

00:22:00 Interviewee

Some people don't even want to store the historical data, so it's less of a concern they they want the subscription, they wanna ping the historical data whenever they want it, right? In some cases, it makes sense to store it on their side and cache it or whatever, but in other cases it's just I'll pay the subscription out into perpetuity to be able to ping this data whenever I want to. So there's different pricing mechanisms with historical and delayed in terms of the rules and the licensing around it and then.

00:22:26 Interviewee

You know, there's really no way to reverse engineer the data that's coming through the API because, at least in our case, our IP is happening in the machine learning phase when we're cleaning the data in our database before it gets displayed and presented and put through the API. So you can't reverse engineer into that. It's just a trade secret that sits on our platform.

00:22:49 Interviewee

But the the other I guess category in this question in, in our industry, in the finance space is display rights. So that's a, it's a tool that a lot of bigger vendors use to make more money which is are you using the data internally or externally. That's a huge factor in the licensing and the pricing.

00:23:09 Interviewee

They don't want because they do open themselves up to risk. If they're selling you data through an API and then you're integrating it into a website and displaying it like Yahoo Finance does, anyone could scrape it right? So they're they're opening themselves up to a little bit of a liability there, and they've used that as an excuse to charge.

00:23:26 Interviewee

A display fee or an external use fee, and so you have to very closely license yourself with the exchanges or license yourself the data provider to know what your use case looks like, and then that another mechanism that uses pro and non Pro users so you can have like an integration fee, an exchange fee, a display fee and then a per user fee. So every single person who looks at the data. This is mostly true for market data, for stock price data.

00:23:57 Interviewee

Every person who looks at the data you have to pay \$0.50 a month for a dollar a month. I mean it varies on exchange but and then they want you to track who the pro and the non professional user is. If it's a professional trader or just somebody who's investing in a stock, it's insane. And then you put the onus on the innovator to track all of that. I mean, you need software just to be able to do that.

00:24:17 Interviewee

In and of itself, and so there's all types of different ways that these data providers try to make more money off of people by licensing the data in certain ways. So there's a lot of different rules and regulations. It's hard to follow. And I mean.

00:24:30 Interviewee

You can if you if you break the rules and don't pay your fees correctly, you're in big trouble. You can get audited and blacklisted from the industry and it's just a nightmare, so it's complicated to license these data sets.

00:24:45 Interviewer

Uhm.

00:24:46 Interviewer

And could you elaborate on the use of decision rights such as term terms and conditions like is it is it separated from licensing uh or and how how does it work work as a mechanism?

00:25:02 Interviewee

99% of our customers just accept the terms and conditions on our website, which is normal. We work with very, very, very big organizations. They have their own complete separate contract that like huge huge financial institutions and those are few and far between. Most people just accept the terms and services on our website, terms of services and conditions. I guess kind of the same thing.

00:25:27 Interviewee

But they do sign an agreement or a contract also that has more detail on their specific usage of the data depending on what type of data they they buy. So.

00:25:38 Interviewee

Everybody accepts the terms of services and conditions on the website and then also in addition to separate contract for the data.

00:25:44 Interviewer

And and what do you cover by the terms and conditions like to, to related to openness?

00:25:51 Interviewee

Yeah. I mean some of.

00:25:53 Interviewee

Some of it is like we're not a registered investment advisor. None of this data is advice. You know, we we have to be really careful about those rules and regulations. the United States. You can't. You can't resell the data different. Those are kind of the main categories on the terms and conditions.

00:26:11 Interviewee

I don't know the rest. I haven't looked at it in a very long time, honestly, but it's mostly just this is not investment advice and don't resell. Our data is really kind of the core of it.

00:26:24 Interviewee

We have a duty philosophically to be an API company, right? I mean, it's like we and our and our mission and our philosophy is very much like we want you to display the data we that's the point we want you to go build something with it and innovate. So we're very different from other data providers in that sense. It's a little, it's a much our terms are much looser.

00:26:45 Interviewer

Again, can you provide insights of gatekeeping practices such as access rights to the to the data platform and their impact on enabling openness?

00:26:56 Interviewee

What do you mean by exit? Exit right. Ohh access, right? Yeah, that's OK. I'm trying to think.

00:27:01 Interviewer

Excuse me.

00:27:06 Interviewee

UM, can you repeat? Can you repeat that one?

00:27:10 Interviewer

Yeah. So can you provide insights of gatekeeping practices such access rights in data platforms and their impact on enabling openness?

00:27:21 Interviewee

Yeah, so access rights for us is really just as simple as API keys, and so we can we control access through API keys. I will say in our industry, the one thing that.

00:27:35 Interviewee

Is a Gray area is you can share your API keys like I could send you my API key right now and you could go get free data from Company_Int_1 and so we have to implement a lot of mechanisms in the background to track usage, right? If we can see everybody's API key and then a correlation to how much data they're pulling, when, where, what type of data.

00:27:56 Interviewee

So much, and if all of a sudden there's a million API calls being made, we can shut the access off, message them and ask. But we do have to monitor that.

00:28:03 Interviewee

Not only because we don't want our data to be stolen, but also it costs us money. Servers, all of the delivery. I mean, every time you make an API call we we pay money for that. And so we have to protect ourselves as a business, but we rarely see issues with that. I mean, we don't have a lot of nefarious people coming in and trying to steal data from us. We do get a lot of bots though.

00:28:03 Interviewee

Not only because we don't want our data to be stolen, but also it costs us money. Servers, all of the delivery. I mean, every time you make an API call we we pay money for that. And so we have to protect ourselves as a business, but we rarely see issues with that. I mean, we don't have a lot of nefarious people coming in and trying to steal data from us. We do get a lot of bots though.

00:28:24 Interviewee

We'll get like 20 sign UPS, 20 API keys and then see it and and we we can tell because the usage is really systematic looking, right. There'll be a 100 API calls this moment, then a hundred that moment and 100 this and it'll roll through the API keys. So we have to do a lot of monitoring for bots that come in and try to take the data through the API keys.

00:28:41 Interviewee

That that is the access mechanism for us is the API key. It's very simple, it's just you can find it in your account. You can access to the data. I would say that the impact on openness is.

00:28:53 Interviewee

Again, a lot of the in the API industry just comes back to the speed and the ease, ease, speed and usability of access. You don't have to get on the phone with a sales person. You can chat with us on the website, get your API keys, boom, your access is open.

00:29:08 Interviewee

And and we're and we're in an industry where, like historically, if you wanted to buy data, you'd have to go golfing with somebody, a salesperson to to like to do a sale, do business and get the data. And then like, maybe a couple weeks later, it's like the timeline. I mean, when you're working with innovators, they want the data now. So you just the API key is a really simple way to say.

00:29:28 Interviewee

And we can we can enforce trial access to, we can attach whatever rights to the API key that we want to. We can say you can only access these 30 stacks or you only have so many API calls. So that's a control.

00:29:39 Interviewee

Mechanism for us.

00:29:39 Interviewee

To to get the data open and flowing quicker while still.

00:29:43 Interviewee

Protecting the business.

00:29:45 Interviewer

OK.

00:29:47 Interviewer

Uhm and how you explained it already a bit, but how do pricing mechanisms contribute to promoting openness in data platforms?

00:29:57 Interviewee

They stop openness a lot because of the complication of the combination of exchange fees, display fees per user fees, and then the technology delivery fee, which would be our fee. And it's confusing because we'll talk to a customer and we'll say this, if you want to buy stock prices data from us, it's \$500 a month.

00:30:18 Interviewee

But if you're going to display the data, you have to pay the exchange \$250 a month and you have to pay them for every user, and you have to track that. And then you have to report it to them. So it's like it's not just \$500 a month, it's we have all these government bodies and and and or, I mean exchange bodies around us that charge things too. And so.

00:30:37 Interviewee

We we have to do a ton of education. I have a bunch of YouTube videos that are like.

00:30:42 Interviewee

The complete guide to exchange fees right here are all the products you can buy. Here's where all your money is going. Here's what to plan and budget for because nobody knows we have to walk people through it. It's way too complicated so it gets easier when you're buying like historical data. But when you get into the real time like streaming data space, that's when it gets really expensive and really complicated with the pricing and the licensing.

00:31:06 Interviewee

But otherwise, if it's delayed data or historical data or more kind of just static data, it's a lot easier and you just pretty much pay us for whatever our delivery fee is to deliver the data.

00:31:22 Interviewee

There's ways around like we've done really crazy stuff like we've taken small exchanges that don't charge exchange fees and woven all of their data together to make a much higher quality comprehensive product to deliver without any exchange fees. But the fact that we even have to try to do that to make things easier for our customers is crazy.

00:31:44 Interviewee

So yeah, it it depends on the type of data in our space that the licensing I I would say in the financial space the licensing and the pricing totally prohibits openness in data and there aren't a lot of companies that are out there trying to make it easier to get access to it.

00:32:01 Interviewer

Ok, thank you.

00:32:04 Interviewer

And do you work with revenue sharing models as well and how does that enable openness in data platforms?

00:32:13 Interviewee

Yeah, we do. We have like 10 different partners who we deliver their data simply just because it's very difficult to source and not something that our team even wants to do and we can get that data to market faster by working with those. Again, it's about fastness and easiness, easiest of use, a lot of providers that are out there in the market building data sets.

00:32:41 Interviewee

In a very old way, which is why we don't want to do the the the sourcing of the data like a good example is analyst estimates. So what do all the professional analysts that are covering these stocks think the stock price is gonna do well, you gotta go talk to all of them.

00:32:55 Interviewee

That that's like a manual data collection approach. We don't. We're never gonna do that. We're a smaller tech team. You do that. We don't wanna call all the analysts and type in the data into a database. It's like against our entire ethos and mission. So we work with other companies that do.

00:33:12 Interviewee

That and so you're basically putting the player who's willing to do the manual stuff together with the more innovative company that has really good delivery mechanisms and APIs and by working together with the revenue share agreement, you're able to take that data, which is otherwise very difficult to access and terribly supported and with bad technology, and bring it into an ecosystem where it can be delivered more easily. That's why we've established revenue sharing partnerships to unlock more data sets for our customers that we otherwise wouldn't source for them. So it leads to more comprehensive access to the types of data you need.

00:33:50 Interviewer

OK.

00:33:51 Interviewer

And a question about the ecosystem identity. So the identity of the platform, can you explain how relational control is utilized to force the openness in data platforms?

00:34:02 Interviewee

What do you mean by relational control?

00:34:06 Interviewer

So the it refers to the degree to which the platform owner relies on norms and values that is shared with the the users and the developers, and to influence their behavior. And well, that can be divided it into self-control and clan control.

00:34:27 Interviewee

Yeah. So I think if I'm understanding your question correctly, we we very much have a different relationship with our customers and kind of our identity and who we are and how how they think of us and work with us than any other data provider, because our entire mission is to make data more open and more accessible.

00:34:48 Interviewee

And so we have kind of an innovators's personality, I would say, and kind of like the phrase we use is we can't wait to see what you build with our data. Like, no, we're not gonna charge you to display it. Yes, we're gonna make this easier for you. Yeah. You can get access in 30 seconds. Go. Here's your API keys. All your resources, like we're gonna make it easy fast.

00:35:07 Interviewee

Like for you, go innovate like the financial industry needs you like it. It needs to be changed and brought into the future. And so we have kind of a camaraderie attitude with our clients and and I think they appreciate the openness a lot. It's.

00:35:21 Interviewee

It's and I think also helpful that I identify personally as the CEO directly with them because I was an app developer who couldn't afford data and that's why I started a company in the 1st place. So they know the company is being run by people who've had the same problems with open data and we get what they're going through. So we're gonna build the best experience for them and there's a lot of trust in.

00:35:46 Interviewee

If that makes sense.

00:35:48 Interviewer

Yes, It does.

00:35:49 Interviewer

Uhm and what about data specific mechanisms. So you said already XBRL. Are there other more data specific mechanisms you use to enable openness and data platforms.

00:36:05 Interviewee

Yes, XBRL is the main one. That's huge. I think XBRL should be like everything should be filed that way in the financial industry. It's just a standard that should be global. It works. It's proven to work, it's huge.

00:36:18 Interviewee

That's probably the main one. We have a lot of proprietary systems in the back end for data quality checks and analysis and and things like that. And and it's and it's relatively straightforward. I think you'll have an appreciation for this just in the financial industry. It's like is the stock price like we have standard standard deviation away from the norm. If it goes beyond.

00:36:38 Interviewee

Check alert. I mean there's very simple things that we do like that and it gets a lot more advanced, but that's an easy way to think about the the data quality side of things.

00:36:45 Interviewee

We have done a lot of work, I would say I have to give a shout out just to LLNS and natural language processing models that have come out that have been open source because not only do I use ChatGPT just in general for the business, but we've leveraged a lot of open source.

00:37:06 Interviewee

Natural language processing models when it comes to the qualitative side of the data. So at the SEC, there's a lot of quantitative information, apples revenue, Tesla's, you know, expenses and debt ratio and all of that information is all quantitative. But the reports are filled with a lot of qualitative information.

00:37:24 Interviewee

As well as news management discussion and analysis company descriptions, you know different things like that. There's a lot of qualitative information and without open source natural language processing technology, we'd have not even tried to get at some of that data, right. So we're able to.

00:37:44 Interviewee

Build API endpoints that enable you to, you know, keyword search through documents and and ask questions and get answers on the text and things like that. And so we have some solutions in that space that we've started to build. So I would.

00:37:58 Interviewee

A large language model and natural language processing technology has played a huge role in opening a lot of that data.

00:38:10 Interviewee

I think a lot of users still rely on companies like us to use it and deliver it to them. I don't think a lot of.

00:38:17 Interviewee

People are building their own UM.

00:38:19 Interviewee

Their own tools for it. But it's it's, you know, it's the fact that XBRL is an open source language used globally. It's the fact that the natural language processing tools are open source for everybody too, that allow.

00:38:31 Interviewee

Providers like us to take it to the next level so that that would, that's the other one. I would mention that has been important for us.

00:38:38 Interviewer

OK. Is there any other governance mechanism related to platform openness that we haven't discussed yet?

00:38:49 Interviewee

I would note that in again in my space, which is just specifically financial data, there's been a huge trend in the concept of data marketplaces and we used to run a data marketplace that used to be the model that we used for our Platform and I think.

00:39:06 Interviewee

I'm not a huge fan. I think that it's been a big trend and and the the messaging is that let's make a big open data marketplace and it's just not the way that it works in finance. You don't shop for financial data the way you would shop for a pair of socks on Amazon, right? I mean, it's a pretty serious, you might be taking that data and investing it and making decisions on a multi billion dollar portfolio or putting the data inside of an app that 10 million users are looking at every day.

00:39:34 Interviewee

It's it's just, that's not the kind of model that works well. I think that data marketplaces are transformative models for opening data in some industries for some types of data. And and it's almost like in the financial industry, there's like a battle of the marketplace now. Fact Set, which is kind of like a Bloomberg.

00:39:55 Interviewee

As their own open fact set, they called it. They literally they named. They used the word open, right, like open FactSet. Amazon launched a data marketplace. There's 6000 data products on the on the AWS Data exchange.

00:40:11 Interviewee

That's decision paralysis. I mean, the data providers like us are frustrated because, like, no one's gonna find us in that ecosystem, it's too much data. It's too many different types of data. So I don't think anybody has solved like the catalog problem is what I call it is that there's so many different like we don't. And Treo doesn't have all the data you need. We don't sell 100%. You're gonna have to go to multiple places to get what you need, and so there's also an issue and a challenge with open data in the cataloging and like the source where where is it? Where do you have to go to get all of this data? Snowflake has a catalog. AWS has a catalog, Bloomberg has a catalog. We have a kind of a catalog on our website and it's like it's kind of all over the place and it's really tough for people to know where to go to get their data. So I don't know.

00:40:55 Interviewee

That the marketplace is a solution to that, but.

00:40:57 Interviewee

Everybody is trying it right now in the space to try to categorize things better.

00:41:02 Interviewer

That then it comes down to data components, right? And do you use specific tools for that like Informatica or Colibra or others?

00:41:14 Interviewee

I think it kind of depends on what you mean by data governance. We don't use Informatica but what what what part of data governance do you?

00:41:20 Interviewer

Mean. Well, to make it data catalog accessible and to to make data findable. And the the roles and responsibilities over the data to govern that.

00:41:33 Interviewer

And yeah to well to structure your your data catalog so you know where the data is what what the data is about and who's responsible for it.

00:41:45 Interviewee

Yeah, that's something that we're constantly trying to improve on our end. I mean, I think the documentation is a huge key in that which is.

00:41:53 Interviewee

First of all, there needs to be a Security Master that links all of the data together in the background, which is very complicated, right? You need to be able to use the the symbol TSLA and be able to pull from multiple data sets.

00:42:06 Interviewee

Just by using that ticker, and there's a whole string of identifiers, and so Security Master having a good Security Master is critical and that having good documentation linking everything together and then just designing good marketing within the catalog is important. When you when you look at a product, you need to know.

00:42:22 Interviewee

And we have this on our website, this the direct source where it's coming from the time that the the data is updated, the latency, the frequency of the updates, the coverage, the history, like what products it works well with what you should consider buying in addition how to get support, how to get help like it's actually really I think a lot of the governance that matters in this space.

0:42:43 Interviewee

It's marketing. It's just putting all that information in an easy way to access and being really transparent.

00:42:49 Interviewee

Right. Because the transparency is not sending a lot of, that's the, that's another big issue with governance is transparency a lot of the big vendors, they don't, they don't tell you what the price is, they don't give you any detail. They make you get on the phone with them and then they'll try to sell you and and.

00:43:00 Interviewee

Quote you a price? That's not. That's like, here's the data. Here's the volume. Here's the history. Here's the access methods. Here's the price. Here's your API keys. That's what makes things faster and open is putting all that information up front in front of customers. Some marketplaces do a good job of it and some don't. It just depends on the company and how much they're investing in in the marketing.

00:43:18 Interviewer

Are there any additional insights or perspectives you would like to share regarding the mechanisms we covered?

00:43:22 Interviewer

Not really.

00:43:24 Interviewer

OK.

00:43:27 Interviewer

Are there any missing elements you would like to add before we conclude this interview?

00:43:34 Interviewee

I don't think so. I think we covered a lot of the you know the the XBRL and market data stock exchanges are the the key in the financial data space, the key, the key mechanisms for openness and the business model. So I think we covered covered a lot. I hope it was helpful.

00:43:52 Interviewer

Do you have any recommendations or suggestions to improve the openness of data platform based on your expertise?

00:43:59 Interviewee

Ohh boy.

00:44:02 Interviewee

I think XBRL should be a global standard for way more asset classes than just it's applicable to a lot of things other than just financial statements. So I think there's a huge resistance at the. I think there's there's a really important role for the government to play, but expecting the government to understand technology.

00:44:22 Interviewee

It's a fool's game, so there's there's an education aspect at the government level, and I and I don't have the answers to that. If we have the answers to that, I think we, our societies would be running in a very different way. But you know, figuring out the role of the government is important and educating them on technology, making XBRL a global standard, removing the self regulation of the stock exchanges in the United States.

00:44:46 Interviewee

And and I don't have the answer on the the platform centralization marketplace, it's it's a huge problem that the financial industry has been facing for a long time. This is just way too many places to go to get the data. There is no central platform. I mean the SEC as the government does a bad job of it, right because it's the government.

00:45:07 Interviewee

And there's some data you can get there, but you have to go to 12 other places to get it. So I don't have other than removing the self regulation of stock exchanges, making XP or all global standard for a lot of different types of data I think would make a huge impact.

00:45:07 Interviewee

And there's some data you can get there, but you have to go to 12 other places to get it. So I don't have other than removing the self regulation of stock exchanges, making XBRL a global standard for a lot of different types of data I think would make a huge impact.

00:45:21 Interviewee

But I don't know if anybody has cracked the code on the right model. I mean, I talked to, I talked to probably five or six companies a week that are trying to build.

00:45:29 Interviewee

A marketplace. So.

00:45:31 Interviewee

I don't think anyone's figured it out yet.

00:45:34 Interviewee

That's a lot.

00:45:37 Interviewer

OK, this what's my final question, do you do you have any comments or questions related to this interview?

00:45:45 Interviewee

No, I think this is a fantastic topic and I commend you on on tackling it. I'm happy to be a resource if you wanna e-mail me if you have questions pop up, feel free to shoot me an e-mail and I'll do my best to to help. But yeah, I just, I'd love to to read it when you're done.

00:46:02 Interviewee

And good luck. Thank you.

00:46:04 Interviewer

Very much thank you for having.

00:46:05 Interviewer

Uh, giving me some of your time. UM and and have a good day.

00:46:11 Interviewee

You too great to meet you. Good luck. Thanks.

00:46:13 Interviewer

Nice to meet you. Thank you very much.

00:46:16 Interviewer

Bye bye.

Transcript Interview 2

25 October 2023

00:00:09 Interviewer

What is your current job function?

00:00:13 Interviewee

My current function officially is in French chef d project Europea, which means European project officer

00:00:22 Interviewee

And in that role I both develop project proposals and consortia around digitization of industry and smart manufacturing. And I lead a big research project at the moment for 24 partners in Europe on on that same topic.

00:00:43 Interviewer

Can you describe the platform of the organisations where you worked with data platforms?

00:00:54 Interviewee

I can. I can give a I can give try.

00:00:58 Interviewee

Platforms is always a bit tricky because when it's when it's something a platform there, there's many types of platforms, that are being used, in larger companies internally and in Smaller companies, often externally. If we talk about data in practise than large research companies and and also large companies are still very dependent on the large tree. So Microsoft, IBM and and Amazon to do their their data analysis. Most most research platforms are still often Microsoft.

00:01:37 Interviewee

So no, no open source, but Microsoft 365 as a as a as an environment, at least in SharePoint as a as a as a place as a platform. It really depends on what you when when do you define something as a as a data platform and so.

00:01:56 Interviewee

Of course, there's platforms that manage applications like Docker. There's there's GitHub, there's a place for for code. Data repositories are still very difficult, so it's it's. It's hard to say.

00:02:13 Interviewee

Let me put it differently.

00:02:14 Interviewee

So so there's platforms to to, to store and and collect project data. That's often Microsoft 365.

00:02:21 Interviewee

There's there's different platforms to develop and the developer platforms are still either, uh, people develop in AWS or they they develop in local software. But using GitHub or local version thereof. So at TNO there is a local GitHub Repository and library, et cetera at the company that I work currently It is really project dependent, because it's a European project. So per European project there are different platforms that are being used and different standards that are that are being used. And that said in the the whole point of the Data spaces is that we become less dependent of those big three and develop our own data platforms and and they are called data spaces and those built on.

00:03:13 Interviewee

Open standards and are a bit it's not defined yet, but think of IDSA, Gaia X Catena X. Also the fiware platforms, that those are the kind of the things that are that that are being developed in that area.

00:03:30 Interviewer

Mm-hmm. OK. And how many years of experience do you have with data platforms?

00:03:41 Interviewee

That that's tricky. I I would say I have been working in that domain for a very long time, but my experience with it really depends on what you call experience with. So I I work.

00:03:58 Interviewee

Uh on research on those platforms, but I don't necessarily work with those platforms. It depends a bit, uh. So I would say 10 years, maybe 15 depends in how you.

00:04:09 Interviewer

My next question is explain the the experience that you have so mainly with those platforms. So mainly research?

00:04:09 Interviewee

Yes.

00:04:18 Interviewer

OK. And, uh, what kind of research to the mechanisms or to the architecture or?

00:04:25 Interviewee

So I'm I'm not in the production part of it.

00:04:30 Interviewee

I haven't been coding for a long time.

00:04:36 Interviewee

I work mainly I have worked also mainly on the research into architecture but also into data management and data regulation. So I worked a lot on the translation from ethics and regulation to practise, and then you have to think about, for instance.

00:05:00 Interviewee

Privacy preserving technologies, or think about, for instance, fair principles and findable principles in in that I also worked a bit on the open science area in the in the research, and there it was really about.

00:05:21 Interviewee

How do you get from from different types and sources of data? So really ontologies towards standards. So I also worked a bit in the in how do you organise different stakeholders around the the nice activity of standardisation.

00:05:40 Interviewer

OK. What governance mechanisms have you observed being used in data platform platforms to facilitate openness?

00:05:55 Interviewee

That's a very good question.

00:05:59 Interviewee

Not so many, actually.

00:06:06 Interviewee

What? What I have what I've noticed. I mean, in the world of.

00:06:13 Interviewee

It it really depends on where you are so so some larger research institutes have kind of.

00:06:21 Interviewee

A A. A certain kind of principles around that. So one would be open where possible close to where necessary for instance. So that's that's a, that's a politic. That's kind of a a principle that that you want to live by but but to put that in.

00:06:39 Interviewee

Practise. That's quite.

00:06:41 Interviewee

That's quite tricky, so one.

00:06:44 Interviewee

Data governance mechanism is is really the in research projects. It's really one. One part is the the the standard research ethics. So that really comes from the world of how do you treat data when when when you have.

00:07:03 Interviewee

Participants of research, but also also how do we make data sharing agreements between Parties. So that's one side is quite contractual and the other is really data protection impact assessment, which is kind of generic to say what is this, what it is that we do with our data, where we share it. And sometimes this this is really problematic this this is still not very well.

00:07:28 Interviewee

Understood. So.

00:07:30 Interviewee

A lot of research projects that share data.

00:07:35 Interviewee

Quite quickly, think they're not dealing with personal data and therefore think this is not necessary and therefore think construct other kind of mechanisms or or governance principles that are often built around standard or or contractual clause or or these kind of things whereas often they still have to do data protection impact assessments. So you see the three kind of mechanisms are General research ethics. OK, what comes from research ethics and practise? So if I'm a research group and I have some private parties in my research group and we're doing a collective research project where we do a lot of data, how do we actually manage that? And that is often on the legal departments talks to the other legal departments. And then there is a there is a kind Shared agreement, either put by the funder or put by the the biggest institution, or the the one who does the most work, or it's really in the context of European research. This this if you do Horizon Europe projects, there's a really strict administrative process that you need to follow and Often, almost always, it comes down to data protection impact assessment that kind of sets the guides as well for what you do locally within that.

00:08:55 Interviewee

Project. And then there's the company itself that often has.

00:08:59 Interviewee

Their own kind of.

00:09:01 Interviewee

Data management and and strategy. That said, there is of course quite some rules nowadays that have been slowly trickling down and and being implemented also by European research. So for instance, the the European research is is pillar 1-2 and three and the pillar two you have Horizon Europe and you have the.

00:09:21 Interviewee

It's not entirely true. It's not entirely Peter 2, but the Digital Europe programme.

00:09:26 Interviewee

I did some projects under the Digital Europe programme and the Digital Europe programme says if you start collecting or sharing data or doing data analysis, you cannot make use of external platforms such as Microsoft, so the so there is the European Commission that forces which.

00:09:26 Interviewee

I did some projects under the Digital Europe programme and the Digital Europe programme says if you start collecting or sharing data or doing data analysis, you cannot make use of external platforms such as Microsoft, so the so there is the European Commission that forces which.

00:09:41 Interviewee

It's not bad to search for European alternatives and then in collective project you have to have for all for so this conversation we would be having, we should we should not. We could not be having it on teams we could we could not use Microsoft 365 we could not use IBM and we could not use Amazon for any of the development, storage or sharing of data.

00:10:02 Interviewee

So that forces as a data mechanism to find alternative partners. There are not so many. They're two in Switzerland and one I think starting up in in Scandinavia somewhere that are actually offering the cloud services that we got so used to in in all our daily work. So, so these mechanisms start trickling down by not by law, but by by.

00:10:23 Interviewee

Rules of the Thunder, which is almost the same of course, so that I've seen and for the rest of what I have seen is.

00:10:33 Interviewee

On the for the for the for the technical development teams that before we start a project, there's an establishment of a way of working. So then then there is a decision, and often there's also quite increasingly for European companies guidelines or what kind of tools you should use and what kind of tools you should avoid.

00:10:53 Interviewee

So no Palantir, no, no Amazon. If it's possible, no, et cetera, et cetera. So so this is starting to come, but then the the question that you always get back from the development team is yes. But then give me the alternative that is as fast and as simple to use and as easy and etcetera and that is often very difficult so.

00:11:13 Interviewee

We have quite some.

00:11:15 Interviewee

Rules coming in from either law or from funding agencies or from practise, because often development teams also know that they and they want to use more, more open standards and open ways of of data governance. But then the practical problem is always what are? Where are the European alternatives? Where are they?

00:11:35 Interviewee

So at some point you're blocked right? At some point you cannot. If you have big data processes or you have heavy Computing there, there are some alternatives in in in high performance computing as well, but they're not easily found yet.

00:11:50 Interviewee

So uh, it, that's that's really the the, the, the, the the gist is in the practise.

00:11:55 Interviewer

And if you look, for example to the platform of Gaia X.

00:12:08 Interviewer

Where they are the platform owner right? And do you see governance mechanisms that they have put in place for the data providers or data consumers?

00:12:32 Interviewee

It's. Yeah, that's a good question. There's and that's a bit where the the the, the data space comes in. Then we we they just published uh the blueprints of all the building blocks also the governance building blocks and also the there are about I think 11 data spaces that.

00:12:52 Interviewee

That are up and running for for almost a year that are trying to also establish these rules. The problem of Gaia X is that.

00:13:02 Interviewee

Gaia X is nothing, right? It's a it's Gaia X itself does not have data, neither is it a data marketplace. So we have a big problem. Conceptually, we're thinking about data marketplaces because everybody wants to have one, and all they do is put some data on a website and figure that's a data platform. But that's not the data.

00:13:21 Interviewee

Platform that the platform needs a business model needs an active outreach needs a particular need, needs to supply and demand, needs a needs. Also established rules needs to also be established in law etc. It's not so easy.

00:13:34 Interviewee

Gaia X for the moment, for for me, is a field project because first of all, if you look at the type of partners that are in there, they're all the partners that we wanted to avoid. So because they have a European office, there are so-called European, but Huawei is not the European company and neither is IBM to be honest, uh, so. So for me to feel attempt but the the problem is that Gaia X and similar initiatives

00:14:05 Interviewee

they don't know what they want, so they don't know what they are. So on the one hand it's a it's an open standard and that's nice. But on the other hand they an open an open standard in itself, it's a voluntary thing, right. So you need to.

00:14:26 Interviewee

Have a reason why this is being taken up and you need to also offer the infrastructure and the architecture to be able to do this.

00:14:34 Interviewee

And that is heavily underestimated. So they thought that in a couple of years they could do the same. What the American companies cost 20 years. And and I think around a trillion of investment you you can't do that. You need to get into the to the to the not only the hearts and minds but you need to get into the daily operations.

00:14:54 Interviewee

And and and software and platforms of of of what we use as developer tools.

00:15:00 Interviewee

You know and and so that means you need to recreate and operating system. You need to infiltrate and and replace the big data centres. You need to have the hardware. You need to have the the bandwidth you need to have the education, the developers and and already start working in your standard and not in that one of Amazon. And all this was was not thought of so then it it for me it it's it's a nice little hobby project with with a good intention but with a complete naive idea on how this how this would work because.

00:15:35 Interviewee

Already what you said, right? The data consumers and the data providers, this division is not so clear and that there is not that there is a market for data to a certain extent, but often the data, the one who generates data is also the one who consumes data or is there's there's not such a single lineal Market from the data is not a product that you pick up here and you put it in a shelf or somebody buys it, uses it and then disposes of it. It's not a linear thing. So so the, the.

00:16:07 Interviewee

The the issue of projecting the physical work of products and markets to data and then thinking that would would work the same for me is a big mistake because data is a is a non fund. It you can use data forever right? It's not. It's not going to get worse so so and data is valueless.

00:16:28 Interviewee

Unless somebody does something with it together with another data with the software and and and makes value out of it for a particular reason in a particular time. But the data is not, it's not like a beautiful face that I have somewhere and has a particular value that Stays. my data is maybe valuable for one second and after that it's useless. So the whole thinking about Data as a market and and thinking about the value of data in terms of a product is wrong, but thinking of it only as a service is also wrong because it's more than that. So we don't know yet conceptually how to think about data as a as a good or as a surface.

00:17:06 Interviewee

Well, because it's neither.

00:17:09 Interviewee

But or as a fuel because also not a fuel is a bit strange. A fuel you you process and you use and then it turns into something else. It's it's not a fuel for data as oil is stupid, stupid, stupid, ridiculous metaphor which makes no sense. Data is not an oil. So this, this and this all this Problematic. This reflects on what Gaia X tries to do because because there is no, there is no supply demand definition in in, in. In general in data. What you can say is that there's a supply demand definition in Information needs that data might service and especially now with the rise of AI, there's a lot of data that needs to be prepared for AI for particular reasons for particular types of processes in a particular market to answer a particular question, yes. Then then there is a supply and demand, but you need to do this long till to both sides.

00:18:04 Interviewee

And then say, OK, if we want to make this open or if we want to help data suppliers and and or do the matchmaking between supply and demand on data.

00:18:13 Interviewee

A. There's a lot of already lots of startups and companies who do this a lot better than Gaia X does. Also in Europe, and B it can never be generic. You need to be. You need to be an expert in a particular market. You need to be an expert in the the information gap that they have and you need to be an expert in what is the What is the willingness to pay for for certain data over a certain amount of time? And this is this is a puzzle that.

00:18:42 Interviewee

Uh, the the the Americans might have a bit of a game ahead of us and but but we want to do it differently yet in Europe. But that's good. We want to do it open and we don't want to do it. The winner takes all style if we want.

00:18:56 Interviewee

To do.

00:18:56 Interviewee

That, and that's what Francesca Brilla, who already in 2018 told the commission.

00:19:03 Interviewee

It's all nice and all, but then you need to really start to fund research on novel types of business models for data, because the only one we know now is either Chinese, which is it's all government or American, which is all winner takes all and the race to the bottom. We need a third way. And this third way.

00:19:23 Interviewee

It is connected to a capitalist system, is connected to bigger questions than only the question of data. So it's a systemic question that on on what what you how you want to be invent that system.

00:19:36 Interviewee

And that will take, we're not there yet but so. So to come back to Gaia X, yes, they do a bit on governance to help the the the suppliers and they do that mainly by offering standards, right. So we can.

00:19:51 Interviewee

Take your data and and treat it in a way that it that it adheres to a particular standard and would also Gaia and IDSA and fire have been doing is a lot of connectors, right? So they call that the semantic treehouse. That's a project that really tries To make the semantics from one data sets generic and connectable to other data sets and and then the the. So that's one thing, yeah.

00:20:14 Interviewer

In terms in terms of API's or?

00:20:19 Interviewee

API's is is one part of it that there is definitely a level in which you can make these kind of connections. Uh the other is maybe more fundamental and that is Really, we could call that ontology mapping, let's say. So I have a particular terminology and I have a particular logic in my data architecture and the other one has a different kind of ontology and and logic and they they need to be mapped. And that's not so easy because they're often.

00:20:53 Interviewee

They look, they look similar but they are not because they are particular jargon that is connected to A to a field or there's a particular legacy terminology that doesn't fit. So. So this is quite complicated. So that and that is not.

00:21:11 Interviewee

That, that, that precedes a bit the API level. If you can already connect on API level then it's you know then you then you already are quite far, but that's that is the the stuff that Gaia-X tries to do and other initiatives in Europe as well like you say like Fiware before that and now the data spaces that build up on top of all these, it's really to say OK what is the?

00:21:32 Interviewee

Common architecture. What are the common standards, open standards and and what is then the? What are then the building blocks of which level we can connect these data sets?

00:21:41 Interviewer

OK. And and that ontology mapping can be seen as a governance mechanism to enable openness?

00:21:50 Interviewee

I would say yes, some some would say no.

00:21:54 Interviewee

So I would say so. So it the discussion in in data spaces, which is kind of interesting at the moment is that there's a couple of elements, right. So there is a there's a whole set of new data regulation that comes in so that the Digital service act, the, the, the Data Governance Act, The AI act, etcetera and this is arranged that the Interoperability Act is coming. So there's like six or seven or eight data regulation. There's actually 27 data regulation laws that are coming that defined data governance principles and often that consists of a rule book that consists of.

00:22:34 Interviewee

You need to have an oversight body. You need. You need to have XY and Z.

00:22:39 Interviewee

A lot of technical people see this as apart from the technical work and what we try to say in DSSC is that your data governance is also a technical governance question. So they they are built in the way you build your your technology. But there's quite some resistance there because often still in the in the world of technology and developers. developers see governance ss a soft kind of external site. So artist, law

people and the ethics people and the and the and the and the business people and and and and that link is not often made. So if you speak about data governance.

00:23:13 Interviewee

Often if you speak to technology developers or the technical side, they see that as not part of their work or as some, especially the people at TNO and Fraunhofer they're really pushing for the agenda that this is at the core of your technological architecture and design, the governance is where where you start because you.

00:23:32 Interviewee

Also, technically decide how you connect to somebody else. What are the rules, what are? How do you make sticky policies? For instance, how do you use smart contracts to to actually enforce automatically enforce particular rules that you have agreed upon, but also what are the the common technical terms that we agree upon, et cetera. So so this is a bit of a, I would say, yes, some would say no.

00:23:56 Interviewee

But it it's a it's a hot, heated debate at the moment.

00:24:00 Interviewer

Thank you. I found some governance mechanisms to enable openness in digital platforms and I'm interested is if they also apply for data platforms. We discussed it already a bit, but can you provide insights into the use of application programming interfaces As a governance mechanism for enabling openness in data platforms?

00:24:31 Interviewee

That's a good question. I'll make it. I'll try.

00:24:38 Interviewee

Yes, yes or no? Do you have to question a bit what's what's an API? Is it used to be?

00:24:50 Interviewee

From a technical point of view, it used to be a way to say, OK, I'm opening up my kitchen a bit to to allow others to connect to particular parts of my my software or my platform or my data set or whatever I put online, and I built an API around it and it became a bit popular in the Web 2.0.

00:25:10 Interviewee

Where you could say OK and for researchers it would be interesting, but also for companies to connect to certain apps and services via the API.

00:25:18 Interviewee

Uhm, but it was often in the developer side and then in the in the later stage the legal departments found out that this was maybe not so handy, so the API became also a place where where kind of techno-legally some rules were enforced. So so it's basically an interface to say.

00:25:38 Interviewee

How how do we interact with each other and what what do we allow you to do and what do I disallow you to do technically, but sometimes also legally you know so. So they're a bit intertwined and.

00:25:51 Interviewee

So in that sense, it is an interesting space what an API does not do Of course, is touch upon the fundamentals behind a particular surface. So you you get to see a part of A company. You see it part of a service or part of a digital technology that is has certain things have been chosen to to be made open and to connect, but that doesn't always tell you everything about the structure and the logic and the and the the reasoning of an entire organisation or company or or software etcetera.

00:26:25 Interviewee

So yes, it is one interface where can enforce rules and maybe also thereby enforce openness to say, OK, we only work together if you use these in these open standards and everything you do in this API will be published there and there or will be made available.

00:26:44 Interviewee

But so so yes, it is an option Services will work via API so so it's one of the solutions, but it it's definitely not the the panacea for all solutions, and that is also on the other side is also interesting because.

00:27:00 Interviewee

Open. Does it mean free? Yeah. So open is is also a bit of a contested term in software development because open standards doesn't mean it's available or accessible or free for all open. Those can also come with different levels of access to different level of of agreements that you have or different business models that you build on top of that. So.

00:27:22 Interviewee

I think open is is a bit of a sometimes misunderstood word in the in the world of software development. Not to say that it's bad, but but sometimes it needs a bit more nuanced than open doesn't mean It's solved for all.

00:27:36 Interviewee

So, so so APIs open for sure, but this this can still lead to exclusionary business models. It can still lead to.

00:27:46 Interviewee

Lock in models. It can still lead to lack of sovereignty. It's it's not the solution always there.

00:27:54 Interviewer

But I what I've defined by openness in my research is the provision of access to the platform resources by third party entities through open source or specific interfaces.

00:28:00 Interviewee

OK.

00:28:09 Interviewer

How are technical development boundary resources such as software development kits or tools used as a governance mechanism to promote openness and data platforms?

00:28:30 Interviewee

If they are used.

00:28:32 Interviewee

Uh, I mean, it's super interesting because yes, it is. I think it's it is a way to promote openness.

00:28:38 Interviewee

This I think we should promote that a lot more open SDK's and open use of open, open toolkits. The problem is especially now if you look at the development of many AI based Applications. So let's call it advanced analytics, because most of the time it's not AI, it's.

00:28:58 Interviewee

It it's not entirely true, I'm going to say that there's a lack of them. So. So yes, there are. It is sometimes used, especially in research projects, especially also in European research project. There's a lot of attention going to three levels. One is architecture. So how can we?

00:29:14 Interviewee

Actually create software architecture high performance computing.

00:29:18 Interviewee

And also.

00:29:21 Interviewee

Exchange of data and exchange of of AI trained models for particular sciences for us to exchange that and on top of that, how can we offer open source development kits to actually start working with this stuff the the the tragic history of of open source software development is often.

00:29:40 Interviewee

Starts in a in a in a in a university lab with enthusiastic nerds in in a corner, little female or other that that put all that time and money and and not even money that often time weekends because they need this to because they want It to work.

00:29:58 Interviewee

It works to a level that it works if you're an expert, then you're equally nerdy and you, you know install this and and play and play around with it when it becomes interesting, there's either a spin out or or it's being bought or just plainly stolen from by a company.

00:30:18 Interviewee

And then later on it's not usable anymore, and there's particular reasons for that. The reason is that open software is an open source developed software kit.

00:30:29 Interviewee

Are instable are not maintained and are built by hobbyists and they are the best experts in The world.

00:30:37 Interviewee

But to maintain software like that you need professionalism. You need, you need somebody who is paid to maintain it and does not do that on a Sunday evening, right?

00:30:46 Interviewee

It's the famous Wikipedia problem.

00:30:50 Interviewee

And and that's why most of these kids that are successful end up being commercialised because that's the only system we know, because there's no public funding to say, oh, we have open software development kits, let's let's give them 10 million a year to keep it up. No, it needs to now. We now unfortunately live in a world where everything has to be valorized and commercialised. So so there there's a million graveyards, software development kits around their research labs and on the Internet that are beautiful and perfect and well coded and hyper functional for one specific research problem, but are just not scalable because there's no funding for it and there.

00:31:28 Interviewee

So there's there's still the the the floss development at the the free Deep Open Software society and they they are still trying but often it still remains in the margins and they cannot compete with the big the big boys when it comes to really putting mass mass skill Developing kits that everybody will use and that is a big risk. So yes, they should be used and sometimes they are used and they're still being promoted in in the context of research, research, funding and research projects. So I think European large projects or national projects that that really say, OK, we're going to fund this.

00:32:05 Interviewee

That to to really develop these kind of kits and and keep them up, but often they, they, they, they they don't make it.

00:32:17 Interviewee

But the sad the, the sad, the we, we we didn't. We didn't think of a better, better way to deal with that so far.

00:32:25 Interviewer

And how a non-technical social boundary resources such as documentation, training and support used as a governance mechanism to promote openness in data platforms.

00:32:38 Interviewee

Ah, that's a very good question. I like that question. It's.

00:32:44 Interviewee

It is getting better. I think there's there's quite some education also.

00:32:54 Interviewee

And if you ask how is it being used? And then the question is by whom he, so so.

00:33:00 Interviewee

I think it is being used more and more in in university or or applied research education to say, OK. And then there's you, you can you can develop and you don't always need to use the the the the standard platforms. But there's also alternatives. There's there's a lot happening in professional education and awareness, also on a policy level to.

00:33:22 Interviewee

It's not always put in practise, but at least on a on a on national and European level, you see a big awareness and also researches and funding going to educating about alternatives and educating about the, the, the notion and the need for digital sovereignty and for for an alternative plan. There's even a huge plan.

00:33:43 Interviewee

Of the the the, the, the recent European Commissioner von der Leyen who had the whole digital decade, which has a large part earmarked for for digital sovereignty, so there there's a lot of attention on that level

00:33:56 Interviewee

I think what's lacking is really on the basic education, like in high schools or in in grammar schools, where unfortunately we all signed killer contracts with software companies for 10 years down the line and that are forcing particular solutions down our throat.

00:34:18 Interviewee

The Google scandal and education, but also Apple.

00:34:20 Interviewee

Did that back in.

00:34:21 Interviewee

The day where this is going in the other direction, so we make ourselves extremely dependent on such external private companies.

00:34:31 Interviewee

To to develop and and offer all our digital education tools and development and and that is problematic. So, so there we are quite soft. So it's it's an in first pyramid now so that the the the more, more.

00:34:48 Interviewee

Elevated stages and a bit more mature forms like larger European companies also in cybersecurity, but also in governments they really start realising that this needs to change, but in lower education we are we are completely doing the opposite and the way this is done, I don't know that that it's also where do people find information on this that's either.

00:35:10 Interviewee

You have a niche market of people who are interested in in what is offered by associations or by neighbourhood clubs or by online courses or by really activist privacy activism. These kind of things that are really in developing these tools.

00:35:27 Interviewee

And they had an effect over the last 1020 years.

00:35:30 Interviewee

Trickle down a bit to a more generic understanding.

00:35:33 Interviewee

But what we could do definitely better is also use social media platforms to promote and educate such messages a lot stronger. And there's there's, there's very few campaigns, a bit on this information, but but that's a bit far from this topic. There's quite quite little information on.

00:35:53 Interviewee

Our dependency of these kind of platforms and the fact that there is there's an alternative and in terms of development and in, in education.

00:36:02 Interviewee

There's quite something to be done in another. There's a good example of the of the AI parade where there's this bus that's showing all kinds of tools and effects of AI in local libraries per per province. These kind of things. So it's interesting, it's going to have a small effect, but at least it starts there to to get people engaged and in in the discussion.

00:36:24 Interviewee

I think for companies and startups it's very difficult because again, what we and I I did a lot of research on this as well on the Willingness to share data, the willingness to to participate in in alternative platforms, they just don't have the money and the time and they they hire a developer that's completely trained in in the AWS platform, that's what they're going to use.

00:36:47 Interviewee

Right. They're not going to spend, uh, if they spend need to spend a half a year on retraining somebody to to programme in an open source platform, that's a bit shaky for them. It's way too risky. So, so that it needs us a a larger level of investment and and structural reform. So you can't just depend on.

00:37:08 Interviewee

On non-technical education it's it's important, but it's not.

00:37:11 Interviewee

Enough to to really.

00:37:13 Interviewee

Change the course of the ship, let's say.

00:37:16 Interviewer

Right. How do intellectual property rights such as licencing play a role in governing openness within data platforms?

00:37:30 Interviewee

They play a huge role at the moment, actually, because there's the, the, the.

00:37:36 Interviewee

One of the reasons why often companies are Hesitant to to to really engage in open platforms is exactly that so.

00:37:50 Interviewee

There was, of course the famous uh Creative Commons licencing scheme that allowed for different levels of openness there and and and where you could decide what you want to be reduced, how you want people to reuse it. This this in a particular sector, this really worked, but it hasn't. It hasn't really been taken up for the product or the more or technical side. So it really worked for media products, right, so to say.

00:38:19 Interviewee

Open unless there was once an idea to really enforce this on a European level, but this never to to lobbying. This never happened. So so it's it's.

00:38:31 Interviewee

It's still a Big issue because especially now if you look at the open source and open publishing sector.

00:38:43 Interviewee

They're really struggling due to a couple of things, so one is.

00:38:51 Interviewee

Uh, that the big companies who are who are ironically old Dutch like like Elsevier and like Cluer and these kind of they they their profit margins are are higher than ever. It's ridiculous while we still and and they use all that power to buy to buy expensive Lawyers to fight against the open source science platforms like science hub.

00:39:17 Interviewee

Uh so. So I worked as a scientist for a long time and I had to. I had to buy my own articles back once they were published. I wrote the god damn thing. What the fuck is this?

00:39:29 Interviewer

For example.

00:39:36 Interviewee

It in that world. In publishing, for instance, are completely ludicrous. So we do all the work for free. Even the refilling, the makeup of the artists. If you need your beautiful designs, also feel book. It's the research. It's designed, that book. Do all the all the spell checking and all the peers.

00:39:53 Interviewee

Doing and all everything, everything, everything we do that's for free in the weekend, the only thing else fear does is press the button to print and they get all the and they force libraries to have expensive licences to be able to access this book. And that's why you, a student, have to pay 100EUR for a book, but we never see that money, not one euro of it.

00:40:13 Interviewee

This is. Yeah, that's the that's the most. It's the most ridiculous business model around. But the point is a bit that that in in. That's why a lot of scientists developed open source platforms like like the social science hub.

00:40:31 Interviewee

The irony is that again same thing.

00:40:34 Interviewee

It became big and then it was bought by Microsoft and Microsoft then has a deal with Elsevier and we're back to Square 1. So so this is.

00:40:41 Interviewee

A bit this is a bit how how these models develop. And then there was Archie fix that that that as a response for all the hard sciences. That said, we don't want to do this and Europe also now forces in your projects to to to publish.

00:40:56 Interviewee

Open source, so then the publishing house to say, OK, you can publish open source with us, but then you need to pay. So now in your project proposal you need to have a a particular budget to pay Elsevier to publish open source. So they win again. So they keep winning and so so it's and that's only in the world of publishing. So imagine in the world of of online commerce.

00:41:17 Interviewee

Or of social media. There's a lot of ways in which open source platforms and.

00:41:26 Interviewee

Are being misused because somebody else then finds a model around it, so the IP is not on the platform itself, but fire services that they built on top of it that they make the entire network dependent of, and then you're back to square 1 so so that that makes that very difficult to to really Make these kind of worlds the schemes like Creative Commons really work. Creative Commons was was a was a great idea to have, not not black and white, but six levels or 7 levels of types of licencing that this is something that that needs revisiting because.

00:42:06 Interviewee

Well, I could talk for hours about this, but the the issue is a bit that that the the ways in which.

00:42:13 Interviewee

And this is misunderstood and and the way companies also see this as a risk is is still a bit staggering and that that that is due to reasons of global business model. So if there would be a local or more more European or national market, it would work because you you you are not threatened, you're not threatened to.

00:42:34 Interviewee

To get stolen or be bought off directly by a hedge fund and etc. So so so this this is what makes it A bit tricky so.

00:42:45 Interviewee

The issue of licencing is is is is an IP also in the context of data spaces is now big big discussion because there's another thing that plays a part in in this this open platforms, especially around data platforms, is that what happens if?

00:43:05 Interviewee

If somebody builds an interesting added value on top of on top of that, so I'm we're three or thirty with 30 parties sharing data on that platform.

00:43:16 Interviewee

And we have a data spaces agreement that we can all use that data and and but it remains closed in that group. But somebody creates a killer app on top of that and and sees the market Potential how do how then do we? How do we deal with this? This has not been this has not been. This is still, we don't know yet. We don't know yet, there's no model yet.

00:43:37 Interviewee

Either you make a contract for all of them, or you make a framework contract, or you make some kind of percentage margin or you. But the exact what if half of it is public parties and the other are private parties? What if some some of it is, is paid by taxes and and innovation money and the other is funded by private funder? So there's a lot of models there that are still under development to think about, OK, but what happens if if something good comes out of this and we still want it to remain semi open or completely open, but we want to also allow companies to to make a profit out of it because that's the point of it to Also take the added value but not in a way that is based on Palo Alto. Winner takes all model. This is still a big question mark and what people also forget often when it comes to IP is that.

00:44:33 Interviewee

The the most successful companies in in the digital space in that come from the Us half of their employers are lawyers and so so we also forget about how big the legal departments are of these companies that do things, know that they're illegal and and just higher and higher and higher to to either drown the fish or or make it so difficult that you will never pursue an IP. OK, there's there's no there's.

00:44:56 Interviewer

I noticed by finding in the interview participants. Yeah.

00:45:03 Interviewee

I can imagine, yeah, yeah.

00:45:06 Interviewee

So so so.

00:45:08 Interviewee

Long story short, yes, very important, but but not a solved issue yet.

00:45:13 Interviewer

OK, a a few more to go. Could you elaborate on the use of decision rights such as terms and conditions as a mechanism to enable openness in data platforms?

00:45:28 Interviewee

It's a bit linked.

00:45:39 Interviewee

It's it's a, it's a it's uh.

00:45:43 Interviewee

It's a. It's a topic that becomes more interesting the the the bigger the player gets. So in the world that I have been working in, this was never yet an issue because we always worked on on research and development that was pre market or not at the at the tier level that was so high that it was ready to be marketed.

00:46:03 Interviewee

So then it becomes a bit like like the cookie button, right? Yeah. OK, whatever. Whatever. Whatever. Here we go. But. But of course, terms and conditions are one way to to promote this.

00:46:18 Interviewee

The equivalent there of in research projects is is your are your are your agreements.

00:46:23 Interviewee

Your, your, your grant agreements on on which you work, and there you often say we.

00:46:29 Interviewee

We we agreed that on the terms that that our data is put on that platform that the results are made open and that any party can for three years after the project can use these results without any problem and et cetera, et cetera. But these are terms you make in in the proposal agreement.

00:46:47 Interviewee

Openness is definitely their way to promote this because it forces parties to say whatever comes out of our shared activity. It needs to remain open to the parties and sometimes also open to the public, depending on the funder which is a very good idea. I think in in terms of condition, terms and conditions in in platforms this this time becomes a legal contractual battle and if you start a business with another company, right?

00:47:16 Interviewee

So so this is One way of doing it, but that's really B2b, I think B2C.

00:47:20 Interviewee

or B to I don't know somebody who's developing in the in the in the attic just using a platform they they often don't realise or are not are not skilled enough to understand Terms and agreements, because it's often legal, legal, speak right. It becomes very legalistic and if you don't understand that as a developer for instance, you you often don't.

00:47:47 Interviewee

It's in Github, must be fine. I'll start coding. Right. So it's a bit like that. So so there's of course the read me as a place to put some conditions, and often there's also in the world of developer there are some kind of licencing and conditions that they are still used right that are also kind of respected.

00:48:11 Interviewee

Terms and conditions is maybe too legalistic as a term to do it upfront. It could work as a as a contractual place to really force openness. If you are more serious. If you engage in more serious forms of of collaboration.

00:48:28 Interviewer

Can you provide insights of gatekeeping practises such as access rights in data platforms and the impact on enabling openness.

00:48:40 Interviewee

Uhm

00:48:42 Interviewee

I think gatekeeping is is a is a is an interesting topic also in relation to a new type of job that it's or a new type of actor that's called the data intermediary. So there's lots of new companies that that call themselves data intermediaries. And this is exactly what they do. So they take data from one side And they offer it to another side and they are the gatekeepers because they offer a particular service on top, like we're going to change your data and treat it in a way so that it fits the data set of the other or that. So they offer particular service but there, but they're also Gatekeepers and you also see that in the EOSC platform, for instance, the European Open Science cloud.

00:49:23 Interviewee

Which is a huge endeavour and one big work package there is to talk about AAI. So AAI means authentication and authorization infrastructure and that is a that is a a place where this gatekeeping , Offer some some possibilities for multi sided platform so you can say.

00:49:46 Interviewee

You are a data supplier or you you send data to us or you make it accessible only these in these parts or only particular metadata or only for a certain amount of time, or only if you have a certain clearance and it's that that is definitely it has an influence on what we discussed before on on, on the levels of openness. So So what do you mean with open and how do you, uh? Because there's particular reasons. Also a good reason for some data to not be open and so.

00:50:17 Interviewee

Either for competitive commercial reasons or for national interest reasons, or for how you don't want to just open all your health data to to anyone you want to make sure that you have control over that.

00:50:33 Interviewee

And these AAI schemes that are quite technical and quite developed nowadays, and they they they are definitely of a hot topic and and and much in development and in terms of connecting these these levels of openness and types of collaboration to what that means in practise and what that means in practise.

00:50:53 Interviewee

For instance, the CBS has this the Central Bureau of Statistics in the Netherlands. If you want to use their micro data, you need to do a course of two days. You need to get a certificate and you need to adhere to a particular rules. So you need to swear an oath, but also you need to pass a technical test and only then in a secured environment you can use particular type of data.

00:51:13 Interviewee

CBS data is open data is public data, but you can't just use it. So I think that's a very good example on How these are connected so so that also gives a particular responsibility to the party, opens up the data to say we can't just open it up to everyone. We need to certify and test and and and create certain levels of who can do what with what types of data.

00:51:38 Interviewer

Thank you.

00:51:39 Interviewer

Uhm, and how do pricing mechanisms contribute to promoting openness in data platforms?

00:51:51 Interviewee

It's a difficult world of pricing.

00:52:00 Interviewee

I don't know if it helps it it.

00:52:04 Interviewee

There's, there's one. There's I'm going to share.

00:52:08 Interviewee

This is a European startup that went to the US as.

00:52:11 Interviewee

They all do, unfortunately.

00:52:13 Interviewee

And they are. They are one of the first to really opened up a marketplace for data, so their service.

00:52:17 Interviewer

Ah, I asked them. So sorry I asked them.

00:52:19 Spreker

OK.

00:52:20 Interviewer

For the interview, do you know them?

00:52:22 Interviewee

Yeah, I know them. Yeah.

00:52:26 Interviewee

Yeah, I can. I can help you connect to Alex.

00:52:29 Spreker

Thank you very much.

00:52:31 Interviewee

That's I know them quite well.

00:52:37 Interviewee

And and it's interesting because, uh, they said what? You know what I said before I really echoed also what their philosophy is. They say the added value is really in the matchmaking. So everybody has data and nobody has data. Often they don't know that they have data. But what what this pricing, it's a bit double what the pricing has done on one side.

00:52:55 Interviewee

It has made a lot of companies aware that they have the data that actually has a value and that if they open this up or semi open it up or partially open it up or have different levels of of access to it that it's actually becomes an asset. The other side is that for some types of data they they became an asset where it's it, maybe it would have been better if they would have been just open and free for all, so it has a bit of a double sided effect in that sense. One example is again the CBS.

00:53:29 Interviewee

That in 2018 No 17 uh, due to a political force, so that that was the force of politics. That said, you need to make more money on your own countries. So they started to build paid data services on public data.

00:53:45 Interviewee

And then some rightfully so, some advocacy groups said oh, oh, that's not the way this is publicly collected data. You can't just make a business model on top of that and sell it to private. So this programme was shut off again by, by by a court later on because because of this kind of thinking to say, OK, this data as an asset data as a.

00:54:05 Interviewee

Data as a valuable thing, means that we should close it off and build business models on top of it so so that there was one negative effect and one positive effect is is that the the the realisation that that there is pricing and that your data that you didn't even care about for instance in predictive maintenance this was a big things so.

00:54:27 Interviewee

A lot of the factories had machines and they were collecting data on the status of machines and the temperature and energy used, and they thought this was useless. And actually this is now a big Data. So Siemens has built an entire data model business model on predictive maintenance and machines and they didn't know that before, right? So that is the positive effect in a way. If you see that as a positive thing, I think so, but yeah.

00:54:50 Interviewer

And it's about care of the machine and providing service, right?

00:54:53 Interviewee

Yeah. So the services on top, yeah, and yeah, and and Alex was one of the one who who jumped in that gap. And they really do matchmaking so they.

00:55:04 Interviewee

You know, all kinds of sectors.

00:55:06 Interviewer

Just out of curiosity, do you think because you want to treat data as an asset you want to manage data as an asset? Do you think data should be on the balance sheet?

00:55:21 Interviewee

Ah, I haven't. I haven't heard that one before. That's interesting. Maybe it should. Yeah, it should be it. Yeah, maybe.

00:55:32 Interviewee

It should it it and if you say and yeah, it maybe it should, but then it's very difficult to establish the value of that of that.

00:55:40 Interviewer

And that's data monetization, right?

00:55:42 Interviewee

Yeah, yeah, but it should be an asset on the balance sheet.

00:55:49 Interviewee

Yeah for sure.

00:55:48 Interviewer

I find it interesting.

00:55:53 Interviewer

Can you explain about how revenue sharing models are implemented in data platforms to enable openness?

00:56:07 Interviewee

Hmm, there's one interesting example, and maybe you know if it already, smart connected supplier network CSN. It's a Dutch initiative actually, but it's a bit bigger than that.

00:56:16 Interviewee

They started a data space data platform for the smart industry, really the production industry.

00:56:27 Interviewee

So the small industrials and it's working really well and they have developed different revenue streams so that that they different revenue streams.

00:56:38 Interviewee

They have case studies where they where they proven that it works that and and and they're quite different.

00:56:47 Interviewee

Revenue streams can help in these data platforms and and in this openness by understanding that often you cannot do this alone. so you might be a I don't know a shipbuilder and you have big machines. You have a lot of data, but you have no expertise or not, not even knowing what to do with it. And then.

00:57:06 Interviewee

It becomes interesting to to hire a third party or to pay per service or to pay per amount of data to to have a service built on top of your data. So you you give this data for free to a third party and they are they are allowed to build a service on top of it and you get a percentage for instance so it.

00:57:26 Interviewee

It helps in rethinking a bit these these revenue streams, but that there's always a a division between.

00:57:32 Interviewee

Somebody who who generates data so a machine data or exhaust data as we call that. So the data that was a byproduct but can become a product you need somebody who sees the value in that in an ecosystem and a new types of service which is indeed often in terms of can be.

00:57:52 Interviewee

Market scans can be material, can be predictive, maintenance can be, can be. I don't know long term impact assessment can be a legal need that you need to comply to and then there needs to be.

00:58:10 Interviewee

As I said that there's often this new kind of in healthcare, in France, there's these data intermediaries and then there is a user of the data. So in healthcare there's, there's the, the hospitals who have no idea to deal with it themselves, but also not the doctor or the patient. So that there's a multisided platform and there's different actors there. The revenue stream in Healthcare is absolutely not clear.

00:58:30 Interviewee

I can talk for a day about that but but but in other markets it's a bit more clear. So there's a good example of wind energy. For instance, there's a wind turbine world where.

00:58:44 Interviewee

Actually, the European Commission is now developing a a law to force the OEM's, so the so the manufacturer of the different parts to open up their data that their machines produce in order to have services built on top of that for by companies who offer a lot of models on energy use. The revenue stream is there then, that you get either a percentage or you or you buy your access to the data in the licencing mode.

00:59:15 Interviewee

Uh, or you say, uh, it's a risk based thing, so we're not paying anything, but we're not buying anything and just until it starts making money and then we need to in a partnership for instance, uh, this also happens a lot that you you are the one who has the big machine and you have some data and instead of selling it.

00:59:35 Interviewee

You create a partnership with somebody who does something with the data and then you create a spin off or you see a bit what happens. in terms of openness, it really depends on the sector, in the market, on how this affects openness or closeness for now.

00:59:51 Interviewee

There's different examples where it goes either way.

00:59:58 Interviewer

Can you explain how relational control is utilised to foster openness in data platforms, so the in terms of the ecosystem identity?

01:00:16 Interviewee

Now it's very fashionable to say that you're open and that you share your data, so it it definitely helps. The German industry was dealing with this. So the German industry was all in in the, in the in manufacturing and in in machinery. And they realised that they missed the boat a bit on digitization. So now Bosch and Siemens and all of them, they have big programmes on funding startups to play around in their data, to open up their data sets to so to really have this start up kind of accelerating and that they did them really well in terms of branding and tapping into completely different.

01:00:51 Interviewee

Markets and players that they would not have recognised before, however, with the come of AI, they're also closing it down quite quickly because they saw a risk of corporate espionage of risks, of opening the kitchen too much. Cyber security risks, et cetera

01:01:12 Interviewee

So on the one hand, it opened up a network towards different types of actors, that is smaller ICT horizontals and startups. But now you see also a big attention going to infrastructure. So who owns actually all this?

01:01:25 Interviewee

So this is the again IBM, Apple and Amazon. The discussion and Microsoft discussion, but also the cyber security issue. So, who is actually how can we still manage and control what happens to our data when third parties start building on top of this.

01:01:45 Interviewee

And should we not then think more in terms of networks of innovation. So be a part of that. What I said before create spin offs but not completely linked to our company. And if it's successful, we'll buy it back, etcetera, etcetera. So you see that that is a bit of what movement in terms of relation intern you see a lot more attention with big companies, but also governments.

01:02:09 Interviewee

On having departments dealing with this with data management with data innovation, with data, with, with, with opening competitions, networks having different roles. So you see also that internally it it creates different structures and not within an organisation because they realise they they need to do something with this the IT department is no longer the one in the basement with the cables, let's say.

01:02:33 Interviewee

Right.

01:02:33 Interviewer

Coming to the end of the interview, what are data specific mechanisms used to manage data platform openness? So specific to data.

01:02:49 Interviewee

Well, there's the the fair data principles, right. And and they come also with protocols, there's data specific.

01:03:05 Interviewee

It's a bit difficult because standards of course are one way and that is important, so data standards but more than more than standards are a result of an interest, but also of an architecture and the specificity about data and those architectures is that it's, as I said, the data is is endlessly reusable. So it's you need to rethink a bit on what this, what this is and the specifics there are.

01:03:38 Interviewee

And we need to think about a bit about what happened in recent discussions.

01:03:44 Interviewee

It has a bit to do with the types of formats that you use, the metadata and how you describe it. So the openness is really connected to all and these kinds of things. So how do you describe your data. To metadata descriptions that are that are open and readable machine findable.

01:04:08 Interviewee

So that is, I think a very important one.

01:04:10 Interviewee

We lost all the beautiful flash movies because they're not readable anymore. There's these kind of things. So the standards on metadata descriptions on openness and connecting is very important. And another thing that was very popular and then died out, and it's coming back a bit is sticky policies, which means you describe Rules and technical rules that you connect to a data set that automatically allow or disallow particular actions with that data. So that is a thing that is very data specific. That is one way of promoting openness or controlling it at least because you can have trace of what is happened, who can do what with it? There are different names for that now they call it sticky policies.

01:05:06 Interviewee

They used to call US ticket policies, now they call it something else.

01:05:08 Interviewee

Yeah, the smart contract. So for us it's a bit of in the same in the same area, but that's more the API.

01:05:21 Interviewee

Well, I think it's mainly that, but that there's some other thing that, that that moves in the back of my head. But if it comes to me, I'll, I'll let you know.

01:06:04 Interviewer

Are there any other governance mechanisms related to platform openers that we haven't discussed discussed yet?

01:06:17 Interviewee

Probably, there's a lot of attention now on AI governance and AI governance models that will have an impact also on data governance models.

01:06:27 Interviewee

In general, data governance models that I've seen start a bit with ethics. That's that's always a bit vague, but then often chuckled in legal data governance models, technical data governance models and business governance models. So that's a bit the three elements that I always see coming back and I think we discussed all three, what's a bit missing is still the societal governance models.

01:06:55 Interviewee

That has to do a bit with willingness to share with how we think about data with is it a public or private thing? Is it how, how do we think about data as on the balance sheet, should we be paid for that if it's public data?

01:07:16 Interviewee

Should there be a public equivalent data comments these kind of elements that we haven't discussed yet that can be interesting depending on what you're looking into. Also as a source of information, because that's quite a common popular term nowadays data comments.

01:07:35 Interviewer

And how does that apply as a government governance mechanism when you are in a data platform owner for example?

01:07:48 Interviewee

Hmm, if you're data platform owner, you could think of a bit more your wider ecosystem and also another thing that comes in more and more now, is societal impact assessment. So what happens to all this data that you're using that maybe also is owned and comes from somewhere else. And another thing is environmental impact assessment. So how much energy does actually your whole data centre consume? Should you be taxed for that? How do you know how much data you're using, how much processing power, where do you store all this data?

01:08:35 Interviewee

That's kind of the technical physical side of data and data warehousing and this kind of things.

01:08:40 Interviewee

Do I store it on the edge so closer to my place or on the cloud where it's I don't know where. So the whole edge cloud discussion and that's for data centres is is an important discussion that will also come to data platform owners because they will have to show in their annual reports where their data is and how many environmental impact there has been. These things will become mandatory.

01:09:06 Interviewee

So that's one way and the social one is a bit more difficult, but it depends on the the application, but it could be interesting if you're dealing with personal data. Also to think of societal impacts of thinking a bit beyond only privacy compliance but a bit larger on how does this affect people and should I have some kind of responsibility on that?

01:09:30 Interviewer

Yeah. In terms of ethics and et cetera.

01:09:33 Interviewee

Yeah, but also how long do you save data shared data? Do you inform people about it? Do you delete it? Do you inform them about whether it's been shared or not? These kinds of things.

01:09:44 Interviewer

Are there any additional insights or perspectives you like to add to share regarding the mechanisms we covered.

01:09:54 Interviewee

No, I think I could keep going for another day but I keep it at this.

01:09:59 Interviewee

I think for now I'm a bit empty, but maybe if I'll think of stuff I'll send you an e-mail. But for now this is a bit what the brain can do at the end of the day.

01:10:08 Interviewer

Then I don't even dare to ask the next question. I assume there are not missing elements you like to add before we conclude the interview? Do you have any recommendations or suggestions, suggestions to improve the openness of data platforms based on your expertise?

01:10:32 Interviewee

I think what I said, it's very important we need to think of business models that work in in that open because open is is not cannot be for free it it it has cost

01:10:43 Interviewee

And we need to think about politically if that cost is public or private and or both and how that then works.

01:10:51 Interviewee

I think the openness of data in certain elements is really important, but there's different levels of openness and to promote that.

01:11:01 Interviewee

We need stronger political will, but we also need stronger and better examples.

01:11:11 Interviewee

If you call a business model or operational models on how to do this, that's two and the third is we need technical alternatives.

01:11:19 Interviewee

So and that will that will that will take some time so.

01:11:22 Interviewee

So this this openness in software can only work if there's openness in hardware, and this will take a long time, but we need to do that.

01:11:32 Interviewer

Thank you very much.

Transcript Interview 3

27 October 2023

00:00:03 Interviewer

Do you have any questions before we start the interview?

00:00:07 Interviewee

No, I'm I hope I'll be able to to see your work once it's final

00:00:13 Interviewer

Yes, it's openness is the the yeah, the the objective the research. So I want to share my interview with the participants as well.

00:00:26 Interviewee

OK, good

00:00:27 Interviewee

Well, OK

00:00:30 Interviewer

What is your job points you know?

00:00:33 Interviewee

So right now I am I guess in between jobs. My my official function is I have my own little company called Wenzel Wear.

00:00:43 Interviewee

So my official Title is Big Kahuna and Hot Sauce chili chemist.

00:00:50 Interviewer

And what was your former job?

00:00:54 Interviewee

I was the CEO for 12 years of tech Smith slash Company_Int_3, which is one of the leading or leading provider of online data stores and marketplaces.

00:01:18 Interviewer

Can you describe the platform of Company_Int_3?

00:01:22 Interviewee

Yes. So the platform of Company_Int_3 is a SaaS solution that provides every tool needed to be able to take content data that you have and use it in one of three use cases. The first one is actually packaging the data as products and selling it.

00:01:44 Interviewee

A good example of that if you go to the CME Group website and you ask for the type of historical data, then their instance of our platform is called data mine.

00:01:57 Interviewee

And today you can go and get up to 75,000 distinct data products that are available and their users. There's anything from farmers that want to have some ideas of of what's happening with different crops and value, to universities, traders. So data monetization is the first one. The second use case is external data sharing.

00:02:26 Interviewee

So a good example of that would be the Keynesian banks that adopted our platform to share amongst themselves all of the data and information related to OTC data products and OTC training.

00:02:42 Interviewee

And then by sharing it, there's some additional calculations and derived contents that's created and that feeds back their risk systems for compliance.

00:02:53 Interviewee

So they're literally saving billions in capital costs with the solution by sharing the information together. And then the third use case is for internal data marketplaces. So in mid size to large organizations actually exposing and making the data available to your users and employees, even if they're not data scientists or developers.

00:03:23 Interviewee

Which is more and more useful with the, the proliferation of AI tools and machine learning tools often time all You need. For example I'm running a server in my office in on my desk and it's a GPT for all and all you need to do is just literallyn Push data in it to be usable in the models that you're working with.

00:03:50 Interviewer

Other extra modules as analytical modules on the platform core as well.

00:03:55 Interviewee

We used to have them, but because if you have analytical modules you have to develop the expertise. So we had that in finance today. The Relelate platform is used by firms in different verticals. So financial institutions is obvious. We have one customer that has an application that is used to scan baseball cards and trading cards and Pokémon cards and tells you the value etcetera. And they've accumulated all of this data And now they have a data store to sell, we have another customer that's in the sports business and they and fan that basically every information you might have related to fans of sports teams and sport Leagues. and part of their business now is allowing their customers to share that content, primarily with suppliers and with firms that that are looking to get better sponsorship deals.

00:05:07 Interviewee

There's in the in the medical field. There's use cases that there's use cases everywhere. I mean, I think that for me, the philosophy behind this is really, really simple. Today, if I look at at this, this is a pen, this is a product. This is easily identifiable. if I look at a a DVD of a movie. Technically it's a piece of plastic with data, but we identify it as a movie or as a DVD. The difference between that and data is its product types.

00:05:44 Interviewee

We've defined a form in which the data gets delivered And it's accepted. So we don't think of it as an abstract concept like data anymore.

00:05:53 Interviewee

And and for me, every piece of data eventually should be and will be packaged and it will become common to exchange data between individuals, between companies, between companies and individuals.

00:06:14 Interviewer

How many years of experience do you have with data platforms?

00:06:19 Interviewee

Well, if we start with my first one, so it's 38 years.

00:06:24 Interviewer

That's a lot.

00:06:26 Interviewee

Yeah, it is.

00:06:29 Interviewer

You explain to the the what kind of experience you have with them already, so I'll skip that part. What governance mechanisms have you observed being used in data platforms to facilitate openness?

00:06:45 Interviewee

OK, that that's so let's talk about the evolution of data management and organizations and it used to be and and it still is in many, many, many cases where data sits in various specific repositories that are Specific to applications, what we've seen is the cloud came in And had companies and organizations change how they're doing things, but as surprising as it might seem, we're still in the early days of the cloud.

00:07:22 Interviewee

Right.

00:07:24 Interviewee

I was shocked when talking to some of the largest banks in Canada and as recently as two, two years ago, they were Doing nothing on the Cloud, they're all waiting to see what others were doing before jumping in. And so we're still really the cloud adoption is still really, really early.

00:07:44 Interviewee

But what we see the trend is first is they say, OK, well, there's ways to save money. Let's take the data repositories and move them to the cloud. So all you've done is change the architecture, but it's still separate repositories.

00:08:00 Interviewee

The next step is to say great. Let's centralize all of the data to have one common access mechanism, and there it is, then creating the concept of a data lake.

00:08:13 Interviewee

So ok, I'm good. I went from my data and elsewhere data to cloud save some money. Great. But no, no huge advantages. Moving it all into one location with a central access with a data lake and Right, it's now there, but only the developers and the data scientists can now use it. So the next step is, let's put in a data governance layer.

00:08:43 Interviewee

And for example, in our platform, we had built our own data governance layer. More and more what we see is companies that do just that provide them a solution. So here's the data governance layer. Here's where you document all the data, all the fields And you put in the the attributes to say this can go internally , To this group, This can go internally everywhere, this can be shared externally to these categories and for data openness, that's the most important part that you need to have Is to be able to, you need to figure out exactly what data you can make available to.

00:09:23 Interviewee

What has Pi? What doesn't? And that's, that's where a lot of companies right now are. Still, organizations are still, in my opinion, struggling because those concepts are new. So now you set that up and it's great. So now I know exactly who can get access to what, but I haven't solved the problem of Actually getting making the data available to the outside world.

00:09:48 Interviewee

Right. And there's a there's a number of options to do that and Company_Int_3 we had built a solution for that. So then the next layer on top of it is a way to expose to in an easy fashion what the data is. So the the data store data marketplace is a great way to do it.

00:10:11 Interviewee

E-commerce and being able to go to on Amazon and finding products, clicking on it is the way that individuals are used to working with data. If you do it on a programmatic fashion then you use other technologies you put in all the contracts and data contracts and the metadata to be able to make it systematically accessible.

00:10:35 Interviewee

Now, so now you have an individual that can go find Data you need to add in there the ability to ask for the data and have some sort of approval mechanism in there to to validate who has access to what and that's often time a combination of the entitlement system and the data Governance application and then once that is there based on how the data has been packaged or prepared for delivery, then it's actually the fulfillment.

00:11:11 Interviewee

Does my user does he want to get a copy of the data in his own cloud instance? Is he looking for history? Is he looking just for updates? Is he using a snowflake and he wants to access the data but all the processing will actually be done on my side, so it's offering different technical means to connect to the content.

00:11:36 Interviewer

Right.

00:11:41 Interviewer

Can you provide insights in the use of application programming interfaces as a governance mechanism for enabling openness in data platforms?

00:11:52 Interviewee

Well, so when we look at that, so let's let's take a step back and and let's see what are the enemies of openness?

00:12:01 Interviewee

OK. And and I think it's an interesting way to to look at it, because what I've seen is there's actually 2. The first one is you you need to change your culture. You need to understand what needs to happen for data to be made available to others and to be open and.

00:12:28 Interviewee

There's not enough expertise yet that hasn't been done often enough for most organizations to actually have that expertise in house. I mean, you're doing consulting for data, and I'm I'm sure you see that as well. So the first one is actually nurship.

00:12:48 Interviewee

And uh.

00:12:50 Interviewee

Same here, it's it's our data. It's in house. It has its uses and we don't need to do anything. So that's the first one. The second one is actually technology.

00:13:03 Interviewee

There's a number of organizations out there, whether it's Amazon, Snowflake, Google, I mean, all of the players and where your data might actually sit and how you might actually get to access the data with specific technologies.

00:13:21 Interviewee

Whether it's It's redshift or Google Computing or again Snowflake, and they all are trying to block customers into an ecosystem.

00:13:37 Interviewee

And as an organization architects their data platform, they will usually select one Sometimes more, but usually it's they're trying to get into one ecosystem to simplify their development life. So I decide that I want to use Snowflake and Snowflake is the most extreme In trying to not be open. it provides to the business value that it provides is. Here is a very simple mechanism for you to create your data lake and centralized all of the data when you push the data in the technology We will make the data usable.

00:14:24 Interviewee

and here's different ways to be able to work with the data afterwards, but as a company they have for a long, long time had the philosophy that once it's there we don't actually want this to interface with any other technologies.

00:14:42 Interviewee

So if somebody wants to take the data and then make it available onto have a copy onto Amazon S3 or do the.

00:14:51 Interviewee

They literally tried to block you and built in mechanisms in there and same thing with Amazon. They want you to adopt their ecosystem to put the data on S3. They'll offer their marketplace if you want to expose it to the outside World. If you want to do it and expose it internally then you have to build all of the tools yourselves or use a tool like Company_Int_3 or one of the competitors because there's a few others now out there. Same thing with Google. So they're trying to lock you into their technological ecosystem.

00:15:29 Interviewee

And that is probably the biggest blocker to have true openness.

00:15:37 Interviewee

One thing I've discovered I've learned in 38 years career is you cannot assume to know how an end user will actually use data.

00:15:50 Interviewee

And I have often so often been surprised, thinking that, OK, well, they come in, they do this, and this is what they're going to be doing it on their side. And This is why they need to get data this way or that way. And then you propose a mechanism, you propose a data format and then somebody comes back and says, but I need it like this.

00:16:10 Interviewee

And I have two choices I can say here, I will adapt my solution to package the data differently for your use case or you get access to the data as is and then you figure It out on your side.

00:16:25 Interviewee

That there was one metric that I'm sorry, I tend to Raven talk a lot. So do interrupt me.

00:16:32 Interviewer

No problem, no problem. Thank you very much.

00:16:35 Interviewee

There's one metric in the 1990s that is still used today because it's still relevant.

00:16:46 Interviewee

Which is that if you're a data scientist, you're actually spending 80% of your time trying to format and make data usable, and only 20% of your time actually using the data.

00:16:59 Interviewee

And openness means for me that you provide everything so that the 80% gets reduced and you can spend more time in actually using the data. That means the data is documented. Accessing the documentation is easy. There's multiple ways to access the data. That was going to suit the last mile, in how you're going to be taking it in and processing it or using it remotely and then and.

00:17:33 Interviewee

Yeah. And then maybe, one other note is everything related to IP.

00:17:40 Interviewee

Because that's the you're going to see companies that have built complete solutions to be able to control usage of the data, right? There's one firm I help once in a while and they had originally built a solution that allow applications whether it's Excel or others to go and get data, but everything that actually happens with the data is controlled and if you want to cut access to the data and remove the data that the user has, you can actually do that remotely.

00:18:24 Interviewee

And why? Either because the data is super sensitive or because as IP you don't want anybody to take the content and start redistributing it or reselling it at that point it's it's not a. It's like trying to stop a terrorist, I mean. If he's really motivated, he's going to blow himself up. There's not much you can do. I guess that those are more like contractual issues and legal issues.

00:18:54 Interviewer

We come later on that or that will be later part of the interview as well, the access rights etcetera.

00:19:02 Interviewee

So maybe.

00:19:02 Interviewer

May be good to know. I defined openness as the provision of access to platform resources by third party entities through open source or specific interfaces in my research.

00:19:14 Interviewee

OK, OK.

00:19:17 Interviewer

Uhm

00:19:20 Interviewer

How are technical development boundary resources, such as software development kits or tools, used as a governance mechanism to promote openness in data platforms?

00:19:34 Interviewee

I'd say it's it helps intermediaries, but it's not really something that we've seen. It's really more for, for me the openness of the data really goes through having the right governance, knowing exactly where the data can go, and then having a mechanism to be able to find the data by the user, request it and then get some sort of technical access to get APIs.

00:20:14 Interviewer

And how are?

00:20:16 Interviewer

Non-technical social boundary resources such as documentation, training or support used as a governance mechanism to promote openness in data platforms?

00:20:29 Interviewee

It's derivation of governance, because that's where you're exposing what is accessible or not. So it is necessary. It is part of what you need to expose so the user knows what he can get and how he can get it.

00:20:54 Interviewer

How do intellectual property rights, such as licensing, play a role in governing openness within data platforms?

00:21:05 Interviewee

It plays an important role in how you're exposing the data. But in the platform itself, again, it goes back to you have to have a governance layer of some sort that's implemented with control mechanisms to be able to define and review and politics and processes to say.

00:21:28 Interviewee

Here we will need to do a yearly review of our governance is, do we still continue to share these fields or those fields that are those are all procedures and processes?

00:21:46 Interviewer

Can you elaborate on the use of decision rights such as term and conditions as a mechanism to you to enable openness in data platforms platforms.

Yes it.

00:22:01 Interviewee

When you make data available, it's always under certain conditions and those can vary greatly. For example, I can say here you have access to the data and you can use it and then it's it what happens to any type of derived content.

00:22:22 Interviewee

Some firms will say it's internal use only. Here's all the the the conditions for you to to use it, and you don't like that either come back to us and negotiate and put in more money, or if you don't like the conditions then that's it. The conditions can vary greatly. The most extreme that I've seen is is actually one large provider, financial data that claim the rights to anything that was derived using the content that you accessed from them.

00:22:59 Interviewee

So that was specifically was IDC, which was large provider of indices and different types of financial data and they basically said for example if you connect and you get a stock price of yesterday IBM traded At \$100 per share.

00:23:22 Interviewee

And if you were to go and calculate and use this and to calculate any type of the right number that the right number is ours.

00:23:33 Interviewer

That sounds a bit strange.

It sounds a bit strange, but it was also a way to control.

00:23:45 Interviewee

How people use the data?

00:23:49 Interviewee

If your business is strictly providing data, then you want to maximize your return and that becomes then the pricing model. Do I charge you per access? Do I charge you per display? Do I charge you per number of company in my firm that can have access to the data?

00:24:09 Interviewee

So I need to build that in another mechanism, something that I've seen which was like purely contractual, was if you're if you're accessing exchange data. So New York Stock Exchange for example sends out a feed of data and there's multiple intermediaries.

00:24:29 Interviewee

That we'll look at the, we'll have access to the data, so I'm AT&T and the telco company and I had to set up a dedicated link.

00:24:42 Interviewee

I'll go through there, then I might have what's called a ticker plant that will take it, make it usable, etcetera. And what they'll say is well, the company that provides you the ticker plant or that as a service.

00:24:56 Interviewee

Do they actually do any processing on the data? Do they just pass it as is or will they calculate an average? Ohh if you do a calculation then that company also becomes an intermediary that's using the data so they will have to pay us also for the privilege of getting the data available to you.

00:25:21 Interviewee

So contracts and the legalese can introduce some some very oftentimes strange and what might appear illogical business terms. But in the end, it's really for one reason and one reason only.

00:25:40 Interviewee

To try to make more money from the data that's being sent out to the users by bringing as many intermediaries as you can.

00:25:51 Interviewer

Can you provide insights of gatekeeping practices such as access rights in data platforms and the impact on enabling openness?

00:26:02 Interviewee

Yeah, of course. Well, that, that's it's basically identifying through the whole chain.

00:26:09 Interviewee

Who is liable to be perceived as a user versus just a strictly a processor? And in that case, what are all of the commercial terms and variants of the commercial terms for the different types of usage?

00:26:30 Interviewer

And we talked a bit already about pricing mechanisms. Can you explain how revenue sharing models are implemented in data platforms to enable openness?

00:26:46 Interviewee

That's where I'll say, I don't know if it's related to openness. And of course, if there is no royalties or payments or if the less expensive the data is technically the more open it is because there's a wider audience.

00:27:08 Interviewee

That can have access to it if they actually need it.

00:27:13 Interviewee

Again, if I go back to my my financial background and if I look at stock market data, when Stock Exchange goes and says here this stock was traded at this price, they claim ownership of that piece of information.

00:27:32 Interviewee

Even if the original ownership is really the participants that in that specific trade.

00:27:40 Interviewee

But as an exchange, I will claim ownership and I will say here, if you are a retail end user and you're not using this for professional reasons.

00:27:53 Interviewee

And the data has this level of freshness, then it's worth so much if it's less than 20 minutes old then you will pay us one cent per quote that you see with a maximum. Ohh you happen to be a professional trader or you've created a a trading group with your friends. Well sorry it's not 1 cent anymore. Now you have to pay us \$36 per month for the privilege of looking at this data.

00:28:21 Interviewee

But if you get it after 20 minutes, then it's free, but you might be missing some fields which will which you have to pay for or pay the real time fees if you get it at the end of the day and you just get the final trading value then it's a different price.

00:28:42 Interviewee

So there's different ways of doing that and putting different levels ultimately.

00:28:48 Interviewee

It should. It is meant for the owner of the IP to go and again maximize how much money they make and it's it is often to the detriment of the users.

00:29:08 Interviewer

Do you want to elaborate a bit more about the pricing mechanism so you?

00:29:11 Interviewer

You discussed them already briefly.

00:29:17 Interviewee

Yeah, well, the pricing is really based on how as a the owner of the content, I will perceive the value the data that I'm providing, I mean am I providing it first of all because I have to provide it? Am I providing it as a public service? Am I providing it as a support to my operations and am I providing it as a way to literally make money?

00:29:47 Interviewee

Or am I providing it one? One of the things that I think we'll we'll start seeing more and more.

00:29:52 Interviewee

Is how access to data to the data benefits the core use of an organization's main activity. If I go back to the CME Group, so CME Group adopted, we signed with them in 2015.

00:30:12 Interviewee

For our data marketplace platform. It was launched in 2016 and I think two years ago.

00:30:20 Interviewee

So the CME Group came back out and and publicly said here we have realized that 50%, So half of the customers that they have today, started their journey by accessing data on the data platform.

00:30:40 Interviewee

And it's an easy way to get familiar with the activities of that group and then to say great, now I want to do business with them this way and that way. And because I have data, I know how, how I'm going to need to integrate my systems to be able to do that.

00:31:02 Interviewer

Can you explain how relational control so is in terms of ecosystem identity, Is used to foster openness in data platforms?

00:31:15 Interviewee

Think if you put in the word control and openness mutually.

00:31:23 Interviewee

Then instead of control or openness, you actually want to talk about quality.

00:31:31 Interviewee

Right. If data is available, is it actually good data and is it actually usable? And if I go and I'm looking for a specific type of content?

00:31:45 Interviewee

I find a database that's publicly accessible. I go and get that data and I realize I do a lot of work and effort and then realize that, well, you know what, everything is bad because the data itself is unreliable. I think all of the controls and everything that you put before it also helps to legitimize data, and if data is recognized as being usable, then it will be more open because it will be more desirable.

00:32:22 Interviewee

What are data specific mechanisms used to manage data platform openness?

00:32:31 Interviewee

Connectivity the the first one, how is the data delivered?

00:32:35 Interviewee

If I deliver the data using an API then my the credentials, the entitlements, the data contracts. For example, if I if I'm using JSON I mean that technology is what is used to be able to enable the openness and apply the controls and then behind that behind the controls then that is how has to be linked to my governance.

00:33:03 Interviewer

And and how?

00:33:04 Interviewer

Do they do they contribute to data platform openness specifically?

00:33:09 Interviewee

Well they don't actually contribute to openness. There is no standards other than if I'm using JSON and I've got a data contract or on how to go and get to it.

00:33:26 Interviewee

For me, and I'm I'm not a developer, but for me that's more or less the limit of openness between data platforms. I think it's really more of a philosophy, and if I go back to the idea of this is a product.

00:33:46 Interviewee

This this is a product, but it's actually a good example. It's a running shoe. OK, so let's say this is a Nike running shoe which it is not. Then if I'm Nike and I'm selling my shoes.

00:33:58 Interviewee

I will have my own stores, so they will come to the source to go and get the data.

00:34:05 Interviewee

But I will also have the same shoes available in sports retailers at Walmart and and the large retailers as well and they are the openness doesn't go through the technology.

00:34:20 Interviewee

It goes through the distribution and it really goes, which is if I copy that to data then then it's me setting up more distribution mechanisms and setting up more business relationships to enable easier access to my product.

00:34:41 Interviewee

So if I copy that to data so I decide to take I have a data product, I have a a database of I don't know, COVID data. There was so much so many things with COVID, it's important. Well, here I'm an organ a medical organization in the CDC or and.

00:35:10 Interviewee

Here I go to my website. Here's a way to access the data. I will also go on to the Amazon Data exchange and I'll make it available for free and there might cost might be to actually host a copy of the data.

00:35:28 Interviewee

On S3, because that's one of the requirements I want to have it on the snowflake data marketplace. Well then my cost is having a copy of the data in that technology and that's where then we go back to the original argument I was making that all of those technological players are trying to lock people into their environments and to be able to use how they expose the data. I am forced to use some of their underlying technology or services.

00:36:03 Interviewer

And for example.

00:36:04 Interviewee

For CC.

00:36:05 Interviewer

X DBC is in the financial world.

00:36:07 Interviewee

Used to use in the world.

00:36:10 Interviewer

In the US.

00:36:14 Interviewee

What?

00:36:21 Interviewer

A format.

00:36:23 Interviewee

Well, there there's a number of formats that exist, so one of the ways to try to provide openness is through lobby groups and organizations. So in the financial world. So let's let's move sideways in the financial world.

00:36:43 Interviewee

And central banks years and years ago created a common format called SDNX.

00:36:54 Interviewee

Statistical something. Some data markup language and it meant that if I'm a country, I'm a central bank, I have information on any type of economics information. Then I will try to have it in that format, so it can be shared with the other central banks and I want to be able to go and get access to their data.

00:37:21 Interviewee

To be able to factor in what they're doing into my economic policies right now.

00:37:32 Interviewee

If you have a a true standard, it's great, but anytime the data starts becoming complex and then standards usually nobody will, they'll be in one standard. It will be the minimalist approach and everything you use in house will be inspired by the standard.

00:37:54 Interviewee

But there'll be different variants of it. Fix is another example. So fix is a standard as to how to do data transactions to exchange data. If you do any type of OTC or fixed income products or bonds and others.

00:38:14 Interviewee

That's standard way of doing it, and there's open source, fixed engines, etcetera, but once you implement it, you will always have some of your secret sauce that you add to it, so those standards and standards by.

00:38:32 Interviewee

I call them lobby groups, but it's really like I don't know if you're in, you're probably familiar with fisd.

00:38:41 Interviewee

Financial Information Services. So it's an affiliate of the software lobby group that was started in the US. And they've been trying to create that's that's actually would be a very interesting study group study case for you to talk to.

00:39:00 Interviewee

Because they've for the last 20 years, they've been trying to find ways to make data open and their members are all the producers of financial data and all the consumers of financial.

00:39:13 Interviewee

And you can see if they've tried to create standard paper contracts, standard terms for usage rights, IP rights, so that if I'm looking at data from Stock Exchange A and looking at data from Stock Exchange B.

00:39:33 Interviewee

Then the conditions to use it are not radically different. They've been trying to create standards as to how to access the data. There's other organizations that have been trying to create standards as to how like like FIX and others.

00:39:50 Interviewee

And I'd say very, very difficult to try to get organizations with different agendas to take the same.

00:39:59 Interviewee

So it will become inspiration, but usually nothing more.

00:40:04 Interviewer

Yeah, we're going to the end of the interview. Is there any other governance mechanism related to platform openness that we haven't discussed yet?

00:40:15 Interviewee

So we've we've talked about the technology, the actual governance of of who gets access to what.

00:40:22 Interviewee

The licensing we we've touched upon and and that's actually an important part of the making the data open.

00:40:35 Interviewee

Packaging and the actual fulfillment. How you, as I say physically, but how you virtually go and get access to the data is important. The business models around the data is also important and same as if I go back to stock exchanges.

00:40:53 Interviewee

You can see one Stock Exchange that will charge crazy amounts of data and crazy amounts of money like Tokyo Stock Exchange. If I want to get access to one piece of data from them, if it's less than 24 hours, my organization has to pay \$40,000 a month.

00:41:10 Interviewee

And if I'm an intermediary and you also need to pay down versus one of their competitors, that will just give the data away for free because they see that it will help provide more economic activity with their primary business.

00:41:30 Interviewee

So that's important.

00:41:32 Interviewee

Are there any additional insights or perspectives you would like to share regarding the mechanisms we covered?

00:41:38 Interviewee

No, Probably talked about everything. One of the things.

00:41:42 Interviewee

I see more and more.

00:41:43 Interviewee

Is tools and metadata and.

00:41:49 Interviewee

To help get the data used more so, for example in the business that I'm I'm looking to build, I've done some prototypes of AI engines and AI tools to facilitate integration of data into AI and machine learning.

00:42:09 Interviewee

Because it's one thing to have the data, then you have to say either I already have a use case or I mean very extreme. If you look at hedge funds, I know of 1 hedge fund in the US.

00:42:21 Interviewee

That has a data budget of 2 billion U.S. dollars per year.

00:42:27 Interviewee

But they make a lot of money with their trading activities, but it's to the point where they have teams of people that only function is to try to find data, bring it in, in case somebody might want to use it in house.

00:42:48 Interviewee

And and that's the. Maybe another thing is, is the intermediaries are important in this whole ecosystem?

00:42:56 Interviewee

And that's maybe something we didn't touch too much upon, I mean, whether it's like the bloombergs or the Refinitiv's or the datarade or the Amazon data exchange. They provide central points where you can find different types of data and get access to it. They bring in their own value by imposing standards for similar types of data that come from different sources because they had to adapt that data to suit their distribution mechanisms and their control mechanisms.

00:43:38 Interviewee

And so I think the intermediaries is also something that can both block openness if they decide to block it or make it more open if they agree, if they make their data and their services available at large for different types of users and if they're able to adapt their business terms for the different types of use cases that you might be out there.

00:44:06 Interviewer

Do you have contacts within intermediaries that you can use me to connect to know, to conduct an interview with them.

00:44:16 Interviewee

Yeah, of course.

00:44:17 Interviewer

Hmm yeah.

00:44:18 Interviewer

Or other participants that you think they may be valuable to my research.

00:44:24 Interviewee

Let me let me look at my contacts. I mean, I can give you contacts at stock exchanges. I can give you contacts at at the BLOOMBERGS and the Refinitiv's and some of those intermediaries.

00:44:39 Interviewee

There's the one place where I don't have any contacts and the one place which one of the things we didn't talk about as far as openness is perception of data.

00:44:51 Interviewee

As you know, in the US actually just about everywhere, I mean.

00:44:57 Interviewee

10 years ago the big, big big concern was is my personal information being sold. I mean, is every does everybody have access to what I'm doing in my life? Yes or no? And that came along in part because in the US you had the concept of the data broker.

00:45:17 Interviewee

Which worked behind the scenes and it's very difficult to find information on data brokers.

00:45:25 Interviewee

Except in a few jurisdiction like Vermont, where if somebody is selling any type of data related to an individual that is a citizen of the state of Vermont, then there has there's a registry that actually exists in all the data brokers out there.

00:45:43 Interviewee

And I think it has like 4 or 500 names, but the whole fact that this was done behind the scenes, it's extremely opaque, gave a bad wrap to the concept of exchanging data.

00:46:03 Interviewee

Right. I think we're all we're getting over that. But the perception by individuals is still there when somebody came to me when somebody comes to me and.

00:46:13 Interviewee

Says, well, how do I know if the bank is not selling all of my information well?

00:46:19 Interviewee

If the bank actually had a data store, you'd actually be able to go and see and actually see what they're making available to others. So some of the openness is also as to who has access to what is being made available to others.

00:46:40 Interviewee

And if it's B2B, it's OK if it's B2B, but with some consumer data in there, then that's.

00:46:49 Interviewee

Something for which?

00:46:50 Interviewee

Their standards like GDPR and others that are all known because we've all received zillions of emails on that.

00:47:00 Interviewee

But how that is being perceived is, I think, the one of the blockers to data exchange and openness being more prevalent.

00:47:13 Interviewer

And as an organization, you want to trade data as an asset and do you do you think data should be on the balance sheet balance sheet?

00:47:26 Interviewee

Yeah, so so you've you've read the the book by.

00:47:30 Interviewee

What's his name?

00:47:31 Interviewee

Yes, absolutely.

00:47:37 Interviewee

It should be.

00:47:39 Interviewee

On the balance sheet, it has value and it has proven value. You might have heard of the use case where in Canada because of COVID it, everybody stopped traveling. So the airlines and Air Canada went to the government and say please, please, please loan us some money and we need it to survive, OK?

00:48:08 Interviewee

And the Governor of Canada said yes. What did they ask as collateral? They did not ask for the planes, cause those are leased anyway. They actually ask as collateral you will put all of your data.

00:48:23 Interviewee

Because that is really as an airline West of value, your value is you've created processes to get people around, but all of that is based on the data that you're processing to make it happen. So their data was valued at I think it was 6 billion Canadian dollars

00:48:43 Interviewee

You know, it was. I'm bringing up the example because it was a large amount of money. There is one company. They're actually based in Costa Rica called GULP data. You might have heard of them. If not, they could. I could. You can look them up on on LinkedIn.

00:49:02 Interviewee

And I I could do an intro to one of the the founders, so there there are people that have been very busy in the data world and their business model is we will give you a loan for your business, you will end your collateral. You'll need to put in is the data that you have.

00:49:21 Interviewee

So same as with software, you'll have to put a copy of all of your data in escrow and, yeah, and and that's their whole. And they've built technology to be able to very quickly evaluate what you have as data and to assign it to value.

00:49:40 Interviewer

Interesting. So the so the rent to the money for exchange for the data?

00:49:48 Interviewee

That you basically what they say is we recognize that the data has value.

00:49:54 Interviewee

And we are confident enough that we're actually willing to loan you money against it.

00:50:03 Interviewee

Oh wauw

00:50:05 Interviewer

Do you have any recommendations or suggestions to improve the openness of data platforms based on your expertise?

00:50:15 Interviewee

Yes, education.

00:50:18 Interviewee

I think again right now it's a it's still extremely early. It was only less than two years ago that Gartner, the Consulting Group recognized that data exchanges and marketplaces was part of the data fabric or data infrastructure that an organization should have.

00:50:45 Interviewee

Same as a few years ago, they said, well, you know what, everybody.

00:50:48 Interviewee

Here is the percentage and we estimate in in 10 to 20 years firms that will be using the cloud well they came up with their predictions saying here this is super important and we believe that within 5 to 10 years 20% of all organizations will be operating or doing some sort of data exchange or marketplace, internally, externally with partners or actually selling data.

00:51:18 Interviewee

And uh.

00:51:21 Interviewee

It was the first time it was recognized by one of those large groups, industry groups. If you see how that under that works, then of course all of CIOs said, OK, that's something I now need to look at, but there's not enough expertise out there because there's not enough people that say, yeah, we've set this in house

00:51:43 Interviewee

I have expertise and I have experience and I've worked with with Company_Int_3 and and with this and with that and that's how we link them together so that these people will then go to another company and cross pollinate the knowledge.

00:51:59 Interviewee

And that's that's where we're at and it's very, very easy to just say, well, I don't know enough, we don't want to screw it up. And right now we'll wait. And so I think education in the industry is probably the most important piece today for the openness to happen, because then people will understand what they're getting into.

00:52:25 Interviewee

And they won't be afraid to say. I will now get a project going to open up some or all of my data.

00:52:37 Interviewer

Thank you very much for participating.

00:52:39 Interviewer

This was the end of the interview.

Transcript Interview 4

30 October 2023

00:05:22 Interviewer

Perfect. Do you have any questions before we start the interview?

00:05:27 Interviewee

So will you share the report, the thesis with me later ?

00:05:31 Interviewer

Yeah, for sure. No problem. It's about openness. So why not being open about it?

00:05:37 Interviewee

Sure, sure sounds good. No questions you can start.

00:05:40 Interviewer

Perfect. What is your job function?

00:05:44 Interviewee

So I'm an assistant professor at the Company_Int_4 University in the Netherlands

00:05:46 Interviewer

You told me before that you worked with data platforms, some in the US, and in Europe, can you describe those platforms for me?

00:06:03 Interviewee

So I worked with narrative, which is one of the data marketplaces in the US. What happens on this marketplace is that there are buyers and sellers who come to this marketplace. Sellers, you can think of companies like MasterCard, visa, Uber.

00:06:27 Interviewee

They they have a lot of data with them and they want to monetize it. So they come to this platform on the other hand, on the other side, you have buyers who are interested in such data points. What this form does is essentially facilitates the trade between buyers and sellers.

00:06:47 Interviewee

The sellers have an option to Say that you know, this data is exclusively given to you, or they say that you know or we are sharing it with 5 other vendors as well. They also have the option to blacklist customers. So essentially you can think of it like a stock market, but for data.

00:07:09 Interviewee

Where or you can think of it as Amazon for data right where it is a marketplace and it's just facilitating the trade. So the marketplace itself does not own any data, but it's the ownership lies with the vendors, the sellers and the main selling point of this marketplace was that you know, every firm these days, they have to.

00:07:35 Interviewee

They are in the business of buying data from different sellers, so they have to set up contracts with 10 different vendors. So if that is the case, that's a very difficult thing for companies to do setting up multiple contracts.

00:07:51 Interviewee

So why not just set up a contract with one marketplace where there are multiple vendors so that that eases some of their tasks. So that was the main selling point of setting up this data marketplace.

00:08:07 Interviewer

OK. Thank you.

00:08:08 Interviewer

Uhm

00:08:10 Interviewer

Are there any other data platforms that you researched or worked with the narrative?

00:08:17 Interviewee

So I've sort of worked with Swash as well, which is not a data platform in itself, but they refer to themselves as data union.

00:08:28 Interviewee

And what they're trying to do is essentially provide some so, so individuals like you and me, we browse a lot of websites throughout the day and we generate a lot of data in the process of browsing right, like url, time of visit, did we visit a financial website, did we visit a retail website? All these data points are being generated while you are browsing, so their idea is that you know, why don't we collect such data and we try to monetize it.

00:09:00 Interviewee

So Swash is one of the platforms which is into this business. And I sort of looked into their operations to determine what sort of incentives do you give to the individuals to share high quality data.

00:09:16 Interviewer

How many years of experience do you have with data platforms?

00:09:21 Interviewee

I would say around 4 to 5 years.

00:09:25 Interviewer

And can you explain what kind of experience you have with data?

00:09:30 Interviewee

So mostly what I do is that I talk to these data platforms, I look at their operations that how they are doing their businesses and eventually my goal is to look at academic problems that I can solve. So I sort of motivate my academic research by looking into the operations of these data platforms. So the problem is motivated from these platforms. Then once I am have modeled the problem I go back and I see whether I have I have accurately modelled the problem and once everything is done, I also go back and give them their feedback, give them the feedback that you know what insights I obtained from solving this problem. And so that's sort of an academic industry collaboration that that usually happens, which is there.

00:10:32 Interviewer

What governance mechanisms have you observed being used in data platforms to facilitate openness?

00:10:43 Interviewee

So I have also worked with Open Data Institute in the UK where I was a research fellow and in India also I have I have some connections with public policy think tank.

00:11:01 Interviewee

So I can give you some broad overview here and this is also in the context of India. So just to give you just to set up the context in the last decade, what happened in India.

00:11:15 Interviewee

There was a huge push towards digital, public infrastructure, digital public infrastructure. It comprised of 3 things. One is an identity layer, which means that every individual should have a digital identity so that establish. The second layer was payments layer, which is that how should payments happen in the Internet based economy? So India developed these platform known as UPI which is unified payment interface.

00:11:51 Interviewee

And the idea was that, you know, you can have this interoperable payment systems. Banks can use it any non banking financial institution can use it. And the 3rd layer was the data layer which is still under construction.

00:12:07 Interviewee

So what has happened, at least in the past 9 to 10 years? And this is one of the things that I've been observing very closely is that.

00:12:20 Interviewee

A lot of push from from physical economy to the digital economy has been happening. Companies are getting transformed, there is digital transformation happening.

00:12:31 Interviewee

So and so.

00:12:45 Interviewee

And what happens because of this you have to look at the governance mechanisms right? Which is when IT came, in the early 2000's, there was this IT-governance. And what happened is that when now data is such a big thing, we are looking at data governance and the sort of roughly what has been happening is people have been at least industries have been just blindly adopting the IT governance to data governance.

00:13:06 Interviewee

So but again your question I think was about what mechanisms do I observe one, one thing that I would say that has happened is data sharing agreements.

00:13:18 Interviewee

When you are, when you are setting up these governance mechanism, you need to make sure that you know who is, what are the sort of agreements, who is the owner of data, what sort of role access do you give to different individuals within an organization and then also outside the organization?

00:13:38 Interviewee

Right. So data sharing agreements has to be in place. That's one big governance thing.

00:13:44 Interviewee

The second thing is standardization. So standardization efforts have to be in place. You need to define common formats, protocols and interfaces for data exchanges within the ecosystem.

00:14:00 Interviewee

And that's how you make sure that you know data platforms can seamlessly interact and share data. So this is also closely linked to sharing data, data sharing agreements. So standardization I think is one big thing which I observed. When you have these populations scale Governance mechanism, then standardization is one thing and then finally data licensing. OK, so so 3 things I would say 3 major things, data sharing, data standardization and then data licensing and attribution. So data licensing makes sure that you are using the data in the way that has been specified in the agreement.

00:14:48 Interviewer

Thank you very much.

00:14:56 Interviewer

Can you provide insights into the use of application programming interfaces as a governance mechanisms for enabling openness in data platforms?

00:15:07 Interviewee

Right. So if you see now the economics of application programming interface was that you know firms like stack exchange or this AccuWeather or this weather apps that they have they provide or even other firms you can think of they provide API access right? And I think this API access sort of enforces this all the 3 things that I mentioned right? So the standardization is there. Otherwise let's say our scraping data then there is low standardization but once you provide API access then the data owner can standardize the data format in which all the parties can access data, right? So standardization is taken care off.

00:15:56 Interviewee

Also through API access you can also make sure that the data sharing agreement is in place, so typically these things are specified that you know what, what sort of data will be shared with you, what data will not be shared with you, and then also the licensing is in place, right? So I think API is one of the enablers of data governance.

00:16:21 Interviewer

Thank you. How are technical development boundary resources such as software development kits or tools used as a governance mechanism to promote openness in data platforms?

00:16:39 Interviewee

So first thing is about the architecture, right? When you talk about software, if you're thinking about let's say open data policies, right, so you need to make sure that all these software architecture is in place where you can.

00:16:57 Interviewee

Where you can you, where you can build let's say data products using the software that is there, right. So so open if you're if you're going for open data policies then that needs a separate architecture compared to architecture which is not governed by open data policies, right. So there the software There is a crucial role.

00:17:19 Interviewee

Also also the other thing is about.

00:17:26 Interviewee

You mentioned software, right, so I'm thinking.

00:17:30 Interviewee

Do you have any specific example in mind that how?

00:17:35 Interviewee

Which type of softwares you want to talk about?

00:17:38 Interviewer

Not really like I defined a definition for the software development kit for the development boundary resources.

00:17:53 Interviewee

Yeah. I think one of the critical components in the software development kits, which is happening now.

00:18:00 Interviewee

Is how do you ensure privacy of the consumers, right. So let's say when you're talking about consumer data, that's critical component of privacy as well, right. So let's say if I'm developing an application which is going to collect consumer data.

00:18:20 Interviewee

So how do I first of all, this again ties back to the data governance thing which which means that you need to make sure that you know what are, what you will be using the consumer data for, how you will be storing it, where you will be storing it for how much duration, all those things are tied to the governance mechanism. But then it's the software software development thing that makes sure how you are, how you are procuring data from consumers with this privacy enhancing technologies. Now also the software architecture is changing.

00:19:01 Interviewee

So for example you have this Federated learning which is now happening where the idea is.

00:19:08 Interviewee

Which is sort of a different concept from machine learning. In machine learning, what happens is you collect data from all the consumers you store in a central repository and then you build your model on top of it, right?

00:19:22 Interviewee

In Federated learning, the idea is that the data stays on the devices of the consumers, but you push the algorithm on the, you push the algorithm to consumers devices.

00:19:34 Interviewee

And the algorithm gets trained on consumer devices, so the data never leaves the consumers device. So if this requires a separate type of software architecture, right? So with these privacy enhancing technologies now in place software plays a crucial role in the efficacy of the governance mechanism.

00:20:01 Interviewer

Other differences here between an open standard and commercialized platforms.

00:20:09 Interviewee

I think so because this comes this comes to the licensing arm where commercial licenses are a bit different compared to open licenses. So, that's where the difference comes in. So typically in commercial licenses, the access this is very restrictive.

00:20:30 Interviewee

You are only required to do specific things with the data, whereas in, whereas depending of course on what type of open license do you have, you can freely make changes. Sometimes you can even commercialize.

00:20:49 Interviewee

You can build commercial products using the data so that's why. So the difference comes from the licensing part I would say.

00:20:57 Interviewer

The licensing is the next question actually. So how do intellectual property? We're skipping One questions I come back to boundary resources later, but how do intellectual property rights such as licensing play a role in governing openness within data platforms?

00:21:20 Interviewee

So, oh, so so how does licensing play a role, right. Yeah.

00:21:27 Interviewee

See what happens is when you have this commercial licenses, typically what you can do with data is specified beforehand and it's very clear that let's say I'm providing you data to build a machine Learning model. You're only supposed to do the DAC.

00:21:47 Interviewee

So, so, so the sort of access that you have is restrictive in some way the things that you can do with it is restrictive when it comes to open data policies, open data licensing there.

00:21:59 Interviewee

You have freedom. That also depends. I mean some open source, some open licenses, they only allow you to do certain things. But there is also varying degree of openness that you have in different licenses. Some licenses do not allow you to commercialize, build commercial applications using data, whereas some do allow you to do that. So what happens is with when you have these open data policies in place, open licensing in place, it builds a platform for innovation.

00:22:36 Interviewee

To give you an example, you have in Singapore, in India, in some parts of Europe as well, like in Estonia, they started this thing where you know you collect, you collect public good public data and then you can build applications on top of it.

00:22:56 Interviewee

So let's say for example, the government has CCTV in different places. They also have transit data which is generated by public transactions.

00:23:08 Interviewee

All this data when it is collected and it is made free to use under open licenses, then you can build products on top of it, right? So there is scope of innovation and I think this also helps to achieve that smart city initiatives which most of the countries now have right so.

00:23:28 Interviewee

So I think this plays a crucial role. What openness? I would say that open data licensing, that gives way for innovation as well.

00:23:40 Interviewee

Now, having said that, what has happened is that you don't see much applications being built on top of it, and there are multiple reasons. One is that you know, sometimes open data is not of good quality. There are not many checks happening that you know whether you can rely on such type of data.

00:24:01 Interviewee

Sometimes the resources are not there for quality checks that ideally should be there.

00:24:09 Interviewee

Plus companies, for example, they are not sure whether if we make this data free to use under open licensing, what will be the strategic implications of it?

00:24:20 Interviewee

Right. So so there has been some apprehension related to the use of open licenses as well, I would say.

00:24:28 Interviewer

2 questions about that. What can be done in your opinion to improve that quality?

00:24:38 Interviewee

I think which I think is very doable is rather than starting out as a population scale intervention, you start with a very focused intervention. So let me let me clarify that.

00:24:57 Interviewee

So what happens is when you when you talk about data, data is a digital good, right? So you often think about, hey, let me let me try to build an open data architecture for the city of Amsterdam or for the city of Rotterdam, right. Now when you think about the scale, the scale is huge, right? You are targeting the entire city. You are building the infrastructure for the entire city. Now what would happen is that some parts of the city you might get good quality data, some might have bad quality data.

00:25:32 Interviewee

There can be a lot of issues.

00:25:34 Interviewee

So rather than doing that, I would say start with a very focused intervention where you are able to spend resources on quality control as well, because if you're spending all the you have limited budget with you to start this regular intervention. If you are going to think big and then you are starting with a population scale thing. Then it's going to be difficult for you to manage this quality check, right? So one thing that you can do is start with focused interventions. Start with small interventions that, hey, let's let me focus on this 4 or 5 zip codes together. Let me see how what sort of data I can collect, whether the quality of data is nice, if there is there are, there is low quality of data being generated.

00:26:27 Interviewee

Then you can quickly do the root cause analysis and see what's going wrong? Right. And then you can scale it. So not scaling up too soon I think is the is one of the thing. The other thing is also to understand incentives of open data, right. So if you let's say of course the public data that is being generated, Government looks after that. That's the responsibility of the government. But then also part of open data can also come from businesses, right? So let's say.

00:27:02 Interviewee

There is some public private partnership and then the private firm wants to make the data open. Are there sufficient incentives given to that private firm to make the data public right for open infrastructure. So whether those incentives are in place or not.

00:27:22 Interviewee

That, I think, is something which has not been looked at rigorously of whatever things is that you know, if everyone shares data then we can do a lot with it and.

00:27:34 Interviewee

There are benefits to it, of course. There are benefits to it, but then you have to look at why would anyone share data in the first place, right? So that is what I do with data cooperatives as well.

00:27:46 Interviewee

So those are the.

00:27:46 Interviewee

2 basic things I have at the top of my mind.

00:27:50 Interviewee

In set having incentives in place.

00:27:52 Interviewee

And then starting with focused interventions.

00:27:55 Interviewer

Thank you very much.

00:28:00 Interviewer

Coming back to the boundary resources quickly, how are non technical social boundary resources such as documentation, training or support used as a governance mechanism to promote openness in data platforms?

00:28:17 Interviewee

So non technical. Can you repeat the question again?

00:28:20 Interviewer

How are non technical social boundary resources such as documentation, training or support used as governance mechanisms, promote openness in data platforms?

00:28:33 Interviewee

Right, so so documentation I would say documentation is an important part of standardization as well where if you have the right set of documentation that sort of enables standardization in the processes

00:28:53 Interviewee

So you mentioned documentation and then the other thing was.

00:28:56 Interviewer

Training and support

00:28:59 Interviewee

Right. So so training and support, I think this this comes at a later stage when when you have the basic infrastructure ready, right? So once you have scaled up the operations, that's where I think or if you think about within an organization if you want to give training about, let's say, open data practices, right? Or or data governance in general. If you want to think about practices.

00:29:29 Interviewee

So that that comes when you have the basic infrastructure ready. If you are still in the operational phase where you are trying to set up the policies, then I think this training is not so useful because there was this one firm which mentioned that I was teaching to an executive class and one of the arms, one of the firms that they were talking about was saying that.

00:30:00 Interviewee

Because of this digital transformation, a lot of policies are changing and because of that they started training, but then they had to organize multiple training sessions because the policies were not set. So it's a it's a constant process of setting policies and training.

00:30:19 Interviewee

And then setting policies and training right. So sometimes organizations are also constrained by the budget they have for training. But I would say training comes as a second step in enabling governance once once the basic policies are set.

00:30:38 Interviewer

OK.

00:30:41 Interviewer

Could you elaborate on the use of decision decision rights such as terms and conditions as mechanism to enable openness and data platforms?

00:30:51 Interviewee

So what so do you mean commercial data platforms or any?

00:30:56 Interviewer

Both of them

00:31:04 Interviewee

So, so so the question is about.

00:31:11 Interviewee

Can you? Can you repeat it again the question?

00:31:14 Interviewer

Could you elaborate on the use of decision rights as a mechanism to enable openness in data platforms?

00:31:23 Interviewee

Decision rights or.

00:31:26 Interviewee

Or decision guides.

00:31:28 Interviewer

No rights.

00:31:30 Interviewee

Decision rights, I see. So, which means who gets to make decisions.

00:31:34 Interviewee

OK.

00:31:36 Interviewee

That right. So so again, so again it goes back to the pillar of data sharing agreement and you when you so when in the agreement, it should be clearly described that who is going to make the decisions, who is going to analyze data, who is going to, where the data is coming from, right. So if you look at the.

00:31:55 Interviewee

Flow of data in the flow of data. Typically, data governance mechanisms tell you that which stakeholder has the right to do what.

00:32:04 Interviewee

Usually I see it as a pyramid, so you what happens is that there is a there are base base stakeholders in the lower end of the pyramid they procure data, they analyze data and then they pass over their insights to the top of the pyramid, right.

00:32:25 Interviewee

And that is where usually the decision making happens.

00:32:29 Interviewee

Now now if it is clearly stated that you know how, who is going to make the decision, how the decision is going to be made, whether the decision itself is data driven or whether it is human driven , human in the sense of gut feeling driven.

00:32:48 Interviewee

Right. So you sometimes what happens in these organizations, it's top executives take decision not based on any data or anything, but just by their experience or their gut feeling.

00:32:58 Interviewee

Right. So so if you have these sort of mechanisms in place, how this decision making happens not only data driven decision making but non data driven decision making as well. So that that combined helps in establishing this data governance mechanism.

00:33:17 Interviewee

I would say.

00:33:18 Interviewer

OK. Thank you.

00:33:20 Interviewer

Can you provide insights of gatekeeping practices such as access rights in data platforms and their impact on enabling openness?

00:33:33 Interviewee

Yes, but I think this is a major, challenge for organizations I would say. Because forget about data, but even when it comes to any IT resource, just giving this role based access has been challenging.

00:33:54 Interviewee

And that comes primarily from security reasons. So earlier, what used to happen we is there was this role based access that was given to different individuals in an organization that if you are manager, you have access to these type of data and you don't have access to these type of data.

00:34:15 Interviewee

That sort of became very tricky because what happens in a big organization is that these sort of roles quickly scale up.

00:34:22 Interviewee

So then it is it gets very difficult to do this role based assignment. Then what happened is there was a paradigm shift and then organizations moved towards this and forgetting the name. But the idea was this perimeter based access.

00:34:42 Interviewee

That typically you think of an organization as an onion. You have different layers. Then If, once you are once you are inside the perimeter, you have full access to data.

00:34:54 Interviewee

If you are outside the parameter, you do not have access to data, so that enables this security based access that is given to data as well. So what I would say to specifically to your question is that role based access is sort of quickly getting outdated. At least that's what my view is, and there are other different mechanisms, one only the one which I know is security based access, which is being given to individuals within an organizations that is coming up.

00:35:39 Interviewee

So I would say that although in the early days it was role based access, but now we are moving to other types of mechanisms.

00:35:47 Interviewer

And then if I understand you correctly, it's because it's difficult to maintain like a lot is changing and it's difficult to maintain?

00:35:58 Interviewee

Yes! So so typically as an organization grows, you have multiple roles even. Let's see, even if the organization doesn't grow. So. So today Floris is a quality analyst, then tomorrow he can be a manager of a team as well, right? So then quickly these roles change. So what do?

00:36:15 Interviewee

You do with it.

00:36:15 Interviewer

Yeah, exactly.

00:36:18 Interviewer

How do pricing mechanisms contribute to promoting openness in data platforms?

00:36:25 Interviewee

Yeah, so this is, this is the question I have looked at in quite some detail.

00:36:31 Interviewee

What happens is, so the idea is typically.

00:36:37 Interviewee

So let me let me break this question into 2 parts.

00:36:41 Interviewee

One is when, let's say, when you think about consumer data, OK, when you think about consumer data, typically the idea is that the sort of organizations that you can form based on consumer data is some data cooperatives or data trusts or let's say even within an organization.

00:37:01 Interviewee

What do you do with consumer data right? Now, typically this data cooperatives which are there which are sort of controlled by the members of the cooperatives themselves.

00:37:14 Interviewee

There, the idea is that let's do something good with data. Now the idea is there is no harm in this idea, but they forget the business model of it that any organization they need to have a business model to sustain, right? You need to make a revenue out of whatever operations you are doing .

00:37:33 Interviewee

And that really comes through monetization, but there are these 2 schools of thought that, you know, once some say that, you know, consumer data should not be monetized, there are privacy concerns and this and that so.

00:37:47 Interviewee

So there is this school of thought because of which, you know, all these institutions, which small scale institutions which sprung up, which fought for consumer rights, consumer data, they sort of ignored this thing of data monetization and they did not scale up eventually.

00:38:07 Interviewee

So what I would say is that pricing in general is very important for data. If you are thinking of data as a resource, if you are calling data as oil data is the next oil and oil of the 21st century, all these things are being used, but then there is a price that oil gets traded at right? So if you're thinking of data as a resource, you need to price it, you need to monetize it. I think the separation that should happen is that whether you have obtained for consumer data, whether you have obtained the necessary consent to monetize the data so that should be separate the consent and the privacy concern should be separate and the monetization aspect should be separate. And I think that 2 things can go hand in hand. You can have privacy and you can have monetization as well. So I think if you want to have this governance mechanism, the right governance mechanisms in place.

00:39:09 Interviewee

You also need to think about monetizing data monetization, data, pricing, and which I don't see in the current governance mechanisms yet that you know, you talk about what should be the data flow, who should get access to data. But then firms are also thinking about monetizing.

00:39:29 Interviewee

Right. It's a good stream of, it provides you with a good stream of revenue. Why not to include data monetization practices in the governance mechanism itself?

00:39:40 Interviewee

So that's the broad thought I have. There have been concerns about privacy and all these things, but then that can be separated from the monetization aspect. OK, so pricing should be included in the governance mechanism.

00:39:55 Interviewer

And how do this? How does this mechanism work? Specifically, so the pricing mechanisms to enable openness?

00:40:07 Interviewee

They enable price discovery, I would say so. So openness in a very different way. So let's say if you generate one gigabyte of data every day.

00:40:21 Interviewee

Now the question is what's the value of this data?

00:40:25 Interviewee

Is it worth few cents or is it worth tens and hundreds of dollars? You will only be able to find out if there is some price discovery mechanism in place right? So if you.

00:40:39 Interviewee

If your data is being traded in a very siloed way in a very non open way, I would say that you know it is being passed to a data broker, goes to some other party, and then you don't know where your data has ended up. So that way I think you would not enable price discovery and that way you would not be able to know what is the value of data in itself.

00:41:02 Interviewee

So if you do want to enable what is the right value of data? And I'm talking about openness in this way in terms of price discovery in getting to know about the value of data, good monetization mechanisms have to be in place. I would say without that what would happen is that your data. See your data is going to be traded in one way or the another. If you are enabling transparent mechanisms of trading data that will at least help you in price discovery that what is the actual worth of data.

00:41:39 Interviewee

Right. So then I think all this governance mechanisms can be strengthened as well. If tomorrow if we find out that individuals like you, their data is very valuable, then we can have governance mechanism, more regulatory aspects can be looked into this place. But if your data is not that valuable.

00:41:58 Interviewee

Then the governance mechanism can be appropriately changed, I would say.

00:42:03 Interviewer

And what are the key requirements to set those pricing mechanisms up for data platforms?

00:42:11 Interviewee

There are different ways to price data and it depends. The reason there are multiple ways. Even if you go to the industry, there are multiple ways still being used is because data is of different types. If there is business to business data, B2B data, sometimes all this data is being created through negotiations.

00:42:33 Interviewee

If there is consumer data, sometimes what businesses do is if, let's say your data is being sold for advertisers to marketers.

00:42:44 Interviewee

Then there is a price per record or price per individual that is being set. If let's say there is, if there is some public data which is being monetized or then you have a common marketplace then you have this 2 sided platforms which come into play and then they have a different pricing mechanisms.

00:43:05 Interviewee

So there are a variety of pricing mechanisms being used also.

00:43:12 Interviewee

The type of mechanisms depend on a variety of factors. It's not only the type of data, but it's also on the use of data. So, for example, whether your if your data is being sold for advertising purposes, then just a price per impression or a price per record is the usual commonly understood mechanism. If your data is being sold for machine learning purposes, then you have a different mechanism in place where your where the purchaser of the data, the buyer of data is going to look at the requirements, right, whether the data is diverse enough, it's accurate enough and the price of the data will be based on those metrics, which is usually more complex than.

00:43:57 Interviewee

What I described before so so I would say that it depends. It depends on what type of data, what's the use of data.

00:44:07 Interviewee

Whether it's public data, private data, so there is no set mechanism and it has to allow for different monetization mechanisms to be in place.

00:44:17 Interviewer

What I find strange here is that you say the use of data is a factor that determines the price. But if I buy this pencil.

00:44:41 Interviewer

Whether I buy it to write or to place at the stands as a piece of art, I pay the same price for the pencil. How does it make a difference in data when I buy your for example, your your latest article?

00:45:02 Interviewer

And I place it on the wall or I read it. What is the difference so?

00:45:07 Interviewee

Right.

00:45:08 Interviewee

So, so. So let's let's look at this question. OK. From an economics point of view.

00:45:13 Interviewee

So how you value a pencil or any other product it is based on how much use or how much utility it provides to the buyer of that particular product, right.

00:45:29 Interviewee

Now let's say you bought a pencil. Right now, what are the things that you can do with pencil.

00:45:34 Interviewee

There are very few things that you can do right. You can either write with it, you can place it so there is not much heterogeneity right. You cannot create \$1,000,000 out of pencil by using it in a different way, right?

00:45:48 Interviewee

So, so similarly when it comes to, let's say, a phone, you can there are there are only so many things that you can do with the phone, right?

00:45:57 Interviewee

So the utility, the price of a product is depends on typically one of the factors that determines the price of the product is the how much variation in the utility is there from consuming that particular product.

00:46:13 Interviewee

Now let's say when you talk about data right now, let's say if I have Social Security number of you, right, BSN.

00:46:25 Interviewee

What are the things that I can do with it of?

00:46:26 Interviewee

Course I can just place it, I can do nothing with it. I can obtain health records of you. I can obtain some other information, some tax information about it. I can do a variety of things, right. So what happens with data?

00:46:46 Interviewee

You can do a variety of things with it, plus.

00:46:49 Interviewee

Let's say if I combine your mobile data with your financial data with some other type of data, if I combine all this data.

00:47:03 Interviewee

I get a lot of information and I can do a lot of things by targeting you right. So a lot of heterogeneity in the things that I can do with it and because there is this lot of heterogeneity, that's why you do not have a set price of data.

00:47:22 Interviewee

So this is the answer to your question. But because you have a lot of variation in the utility, that effective price discovery mechanism is usually done in a different way. It depends on what you use it for advertisers, what they are going to use the data for is simply to target you with ads.

00:47:43 Interviewee

Right, so I know what's the utility for it and that's why just the price per record is is the usual way. But when it comes to machine learning, my model has very different requirement. I need the data to be representative of the population. If you're just giving me information about males, that is not going to help me develop a good machine learning model.

00:48:05 Interviewee

Yeah. So that's.

00:48:06 Interviewer

So, and therefore, those intellectual property rights, such as licensing is important in connection with pricing? For this price...

00:48:19 Interviewee

you can do this many things. Right, right. So that's.

00:48:19 Interviewer

OK.

00:48:22 Interviewer

Perfect. Can you discuss how revenue sharing models are implemented in data platforms to enable openness?

00:48:34 Interviewee

So the revenue sharing model is typically, let's say, if you're talking about the platform, typically the seller, the platform charges a percentage fee from the seller, but then usually there are very few data platforms. I would say 2 sided platforms, where such things happen, so these platforms charges fraction from buyers and sellers both sides like Amazon does.

00:48:59 Interviewee

In the retail market, but when you think about?

00:49:06 Interviewee

B2C or B2B selling where there is just one platform, where there is just a firm who is selling data through, let's say an intermediary or somewhere. There usually this revenue sharing does not happen, so revenue sharing

00:49:25 Interviewee

So first of all, for consumer data that is being sold, there is no revenue sharing. Consumers are not given a fraction of the revenue that is being made by monetizing their data. That only happens in cooperatives, data cooperative where data cooperatives. The idea is to pass on the share of the revenue to the data generators, right? But when it comes to data platforms, usually firms who have data with them, whether it's consumer data or whether it's non consumer data.

00:50:04 Interviewee

They do not pass on the revenue to the consumers in general and that sort of that. That was one of the reasons why these data cooperatives started that we should get some fraction of the revenue of the data that is being monetized because it's our data, so why don't we get a fraction of revenue? What happens typically, the idea that forms the philosophy here is.

00:50:32 Interviewee

That firms are going to give you some service in kind. So for example, Google uses your data but it gives you search for free right? So that is how to provide benefits to you. So usually the sort of benefit that is being passed on it is not monetary but it is some kind of service that they do?

00:50:56 Interviewer

Thank you. Can you explain how relational control in terms of the ecosystem identity is utilized to enable openness in data platforms?

00:51:09 Interviewee

Data identity. So all I know is that.

00:51:13 Interviewer

The ecosystem identity.

00:51:16 Interviewer

The data platform ecosystem identity.

00:51:22 Interviewee

So OK, I did not understand the question then.

00:51:25 Interviewer

So I take the definition.

00:51:32 Interviewer

So relational control refers to the degree to which the platform owner relies on norms and value that is shared with app developers to, for example, to influence the behavior and can be divided into self control and clan control. So how you create the ecosystem identity.

00:51:52 Interviewer

How is that used to enable openness in data platforms?

00:52:01 Interviewee

To be honest, I'm not aware about this a lot, so I do not have much idea about this.

00:52:09 Interviewer

OK.

00:52:14 Interviewer

What are data specific mechanisms specifically to data platforms used to manage data platform openness?

00:52:28 Interviewee

So so one thing that has been recently happening is about identifying the source of data and through how the flow of data in general. So. So if I know that this data originated from Sameer 's mobile.

00:52:47 Interviewee

Am I the first one to use it or has it been passed on through some other things, so tagging data I think that is one of the one of the things which firms are paying attention to and that also creates adds to openness because if data has exchanged many hands.

00:53:02 Interviewee

Then you have sort of lost value of it and then you don't know who has access to data and the security concerns arrive. So so though that is, that is one thing about tagging data, tagging the source of data.

00:53:22 Interviewee

Licensing, which is that if I have 3 terabytes of data and I'm going to share it with my partner firm, I should make sure that what sort of things that I that my partner firm can do with it, if there is a secure breach

00:53:40 Interviewee

I have to report, my partner has to report all these things that should be in place. Right, so all this.

00:53:50 Interviewee

The licensing I think is going under a revamp that because of the regulation that has been there after that there is a lot of attention being paid to the licensing part.

00:54:05 Interviewee

And then in general, the flow of data, right, so that I think comes comes if you if it's within the organization forms in general are not tracking data that you know what data is being used for, what purpose is.

00:54:19 Interviewer

Earlier you spoke about standardization. Can you elaborate a bit on that in terms of data specific mechanisms.

00:54:26 Interviewee

Right. So what happens is that when you when you talk about data, typically you do not obtain data from one source, right. You obtain data from multiple sources. Now if the data is not standardized then it is very it gets very difficult to pull data.

00:54:47 Interviewee

Also, let's say if I'm collecting data from source A on platform A

00:54:53 Interviewee

From source B from another platform. So not only the sources of data that has to be standardized, but also that if you're collecting data on different platforms there should be this interoperability between 2 platforms. So if I'm collecting video data.

00:55:12 Interviewee

From one and storing it on one platform

00:55:15 Interviewee

Then if I want to transform the insights to another platform that has to be interoperable. Similarly, let's say if I'm collecting transit data. So this is a good examples. So let's say information about trams, about trains, other public transport. If that is being collected for mobility, so mobility data in general, there has been pushed. So what do they call it now let me see.

00:55:45 Interviewee

GTFS yeah, General Transit Feed specification. So I'll just paste it here. You can look at, look at it later.

00:56:03 Interviewee

So, so in transit data now you have this that you know all this data should be standardized, that if I'm collecting data from multiple transit systems, if the data is not standardized, then it gets difficult to share it across the systems.

00:56:18 Interviewee

2 different stakeholders as well, and it's also difficult to pull data so that I think is one of the reasons why these systems have not scaled up because standardization has not been there, transit data recently, these things have started up so.

00:56:40 Interviewee

That I think is needed for for these data sharing platforms to scale up and for innovation to happen here.

00:56:48 Interviewer

Thank you very much. We are coming to the end of the interview. Is there any other governance mechanisms related to platform openness that we haven't discussed yet?

00:57:03 Interviewee

I think, at least from what I know, I told you most of it. In general the only thing that I that I will reiterate again is these 3 pillars that I mentioned, which in my view are important.

00:57:20 Interviewee

So that I think and then I think you know better about these governance mechanisms, right? So. So yeah, I think we discussed most of it.

00:57:28 Interviewer

OK.

00:57:31 Interviewer

Are there any additional insights or perspective you would like to share regarding the mechanisms we covered?

00:57:42 Interviewee

Not in particular. I mean, there are there are still.

00:57:47 Interviewer

I think.

00:57:48 Interviewee

I thing is there is not much investment happening in these in these domains people still view data governance are still clubbed under IT governance and that has to be separated.

00:58:01 Interviewee

In my opinion, because the way it resources. So although data comes under data is an information good, but the properties of data are different. So so I think that distinction needs to be made going forward.

00:58:17 Interviewer

OK.

00:58:19 Interviewer

Do you have any recommendations or suggestions to improve the openness of data platforms based on your expertise?

00:58:29 Interviewee

From what I studied, I think the incentive mechanism should be in place. If you're talking about openness, if you're talking about first building a data platform, think about where the data is coming from, who is sharing that data right? If it's, let's say, coming from individuals.

00:58:45 Interviewee

Are you sufficiently compensating individuals for taking their data? If it is coming from forms, are forms ready to share their proprietary or strategic data with you? So if you want to think about building value from these data platforms in the long run, you need to provide sufficient incentives to collect data in the first place.

00:59:09 Interviewee

Is what I had to say yeah, OK.

00:59:13 Interviewer

Do you have any recommendation suggestions for me to improve my research?

00:59:20 Interviewee

Are you interviewing with other experts as well? Who are? Do you know who they are?

00:59:28 Interviewer

So that's my next question. Can you provide me maybe a connection to, I contacted narrative but for until now I didn't got any response.

00:59:42 Interviewer

But I conducted 2 interviews with CEOs of data platforms, One Financial data platform in the United States and another in Canada and uh, I had an interview with principal consultant of TNO

01:00:06 Interviewer

OK, OK.

01:00:08 Interviewee

I see. I think that's a that's a good start. So perhaps.

01:00:17 Interviewee

I can maybe. I'll see. I can. I can refer you to one of one of my colleagues at the Open Data Institute. They would be happy to. They would be happy to do an interview as well and see other other operations are, but I'll see. I'll try to connect you to.

01:00:40 Interviewer

Them, thank you very much.

01:00:43 Interviewee

Well, I would say look at not just private firms, but public firms as well. They have a different perspective. All this public, public, public.

01:00:58 Interviewee

Policy think tanks also have a different perspective. So. So yeah, government entities as well. 23 years ago, the government of Netherlands had this. What do they call Dutch?

01:01:18 Interviewee

Dutch vision on data sharing.

01:01:22 Interviewee

See so and then that led to some touch vision. Yeah. Dutch vision on data sharing between businesses. You can just Google it. They have a document that's digitalization strategy. So perhaps you can connect with people.

01:01:42 Interviewee

Behind these, that might give you some perspective.

01:01:46 Interviewer

OK, I will do that. Thank you very much.

01:01:52 Interviewer

Thank you for your time.

Transcript Interview 5

3 November 2023

00:00:02 Interviewer

Do you have any questions before we start the interview?

00:00:04 Interviewee

No, let's start.

00:00:06 Interviewer

What is your job function?

00:00:09 Interviewee

Well, I am a part-time professor at the Company_int_5, where you are right now. And besides that, I'm the chief science officer of Company_int_5. That's our Research Institute of the Company_int_5.

00:00:26 Interviewer

Right.

00:00:29 Interviewer

Can you describe data platforms organizations that you have been working with?

00:00:37 Interviewee

While in a data platforms actually I am not working with data platforms but we are developing and implementing data platforms for our research, so I'm not using a specific data platform, but one of the things that we are doing, we are trying to design a data platforms and not that you can use it directly, but prototypes of these platforms and then we can see where the bottlenecks are, how the research questions are and if you are trying to develop and to design these kind of platforms for example, well, what are the data sharing issues on such a platform? What are the data quality issues if you bring data from different sources together in one?

00:01:29 Interviewee

So these are the topics that we are working on.

00:01:32 Interviewer

OK. And how many years of experience do you have with data platforms?

00:01:37 Interviewee

Well, I don't know if I can answer that question precisely, but I'm working in the field of databases since 1990, so 20-30 years almost. Yeah. In the field of databases and applications of databases.

00:01:58 Interviewer

Can you explain what kind of experience you have with data platforms?

00:02:05 Interviewee

Well, what we are working on and if you have these data platforms and the idea behind this data platforms is that you have data from different organizations, different sources, you bring them together in one platform, so such that people other people can use the data for their purposes. So what we are working on is saying, OK, if we bring this data together, how can we share this data to other parties in a privacy proof way if we.

00:02:43 Interviewee

At least in governmental data, you cannot share with everybody. You can share it, but you have to obey the rules and regulations. So that's the thing that we are looking at. And beside that we are looking if we bring data from different sources together, sometimes they have some overlap.

00:03:03 Interviewee

How can we exploit this overlap to improve the quality?

00:03:07 Interviewee

These are these kind of topics that we are working if we are working at data platforms and also a little also with more a bit more technical stuff. If you have questions on this platform.

00:03:23 Interviewee

How can we answer them? How can we answer this question in a efficient and effective manner?

00:03:37 Interviewer

You touched already some governance mechanisms as rules and standards. What governance mechanisms have you observed being used in data platform to facilitate openness?

00:03:56 Interviewee

Actually what we see is an we are very Eager to open data, but before you can open data you should respect for example the GDPR.

00:04:10 Interviewee

But that's not only the GDPR. For example, if you look at a police and justice data, there are even more strict and rules that you have to follow.

00:04:24 Interviewee

So actually that's one way.

00:04:32 Interviewee

That we are working on them and to facilitate this type of data what we do, we are looking to new technology for example how to encrypt this data homomorphic encryption. Because if you can encrypt the data then you can do with the data what you like but you cannot go back to the original data.

00:04:58 Interviewer

So that's a really specific example of data specific governance mechanisms used only for data platforms and not for social media platforms?

00:05:08 Interviewee

No, not for social media.

00:05:09 Interviewee

But we are working. We are putting a lot of effort to open the data, but you have to do that very careful.

00:05:17 Interviewee

Because they are very sensible data sometimes.

00:05:19 Interviewer

Exactly. Do you have more examples of data specific governance mechanisms as encrypting?

00:05:27 Interviewee

Yeah, encrypting is one of them, and another one is, and which attribute should you anonymize ? So these are important things.

00:05:37 Interviewee

Looking at the different kinds of technique, for example, well known as k-anonymity and all diversity, these are all kinds of techniques that we apply on such a platform. Because you have to say, OK, Here we have a data set and some of these attributes are very sensible. So what we do first we anomize this data before we share them with others.

00:06:02 Interviewer

Do you have examples of non-technical mechanisms?

00:06:08 Interviewee

Yeah, non-technical mechanisms are quite important because technical mechanisms help you for only is the part of the solution. non-technical mechanism are more the self soft and guidelines that we have for example. Well the main question is if we have the data and we share it.

00:06:30 Interviewee

The core question is, can we de-anomize the data? Can we go back to the original data and can we trace the person behind the data? So that's something that's not allowed. That's something that we would not like to have. So and what we do in this, if we look at the data we look how sensitive it is.

00:06:56 Interviewee

and then and what kind of privacy rules and regulation applied to it.

00:07:04 Interviewee

And that's sometimes you have to make a trade off between utility and privacy.

00:07:12 Interviewee

So if you say well and it's too sensitive and we have to anonymize all these attributes then the disadvantages that you cannot use it very well anymore. So you have to make a trade off between utility and privacy, and that's something that you have to do as a human being. You have to take into account what's the purpose that you have with the data and and and and of course, and the privacy of this data. So the tradeoff between these two and the parameters and that's some things. Yeah, you have to follow some guidelines. I don't know if I'm clear.

00:07:56 Interviewer

Yeah, you're.

00:07:58 Interviewer

There are some synonyms, but it's also about the trade off between the openness and control.

00:08:15 Interviewer

I did preliminary research on openness in digital platforms as different kinds of platforms and what governance mechanisms do they use? And I want to verify if that also applies for data platforms. So I have some questions and I hope that you can elaborate how this applies for data platforms.

00:08:58 Interviewer

How our application programming interface is used as a government governance mechanisms to enable openness in data platforms?

00:09:09 Interviewee

Yeah, I I I'm afraid that I not really Get the questions precisely. You have the API, you have the data.

00:09:22 Interviewer

How does the API function as a governance mechanism to enable openness?

00:09:27 Interviewee

OK. OK. Yeah. Yeah. OK. Well, that can be in a different ways. What you can have is first that you say well, we have the data, we do all kind of preparedness and so on. And you have an API and the API you can do All kind of checks. For example, suppose that. Uh.

00:09:53 Interviewee

Let's give an example. Suppose that you have a question you would like to have the professions of a set of people who are living in Amsterdam.

00:10:06 Interviewee

So then, uh, well, you can ask the API give me the data, the profession of the inhabitants of Amsterdam.

00:10:16 Interviewee

So, well, that's not a problem if you have writers because you have many writers. If you have a software engineers, you have many software engineer living in Amsterdam.

00:10:25 Interviewee

But when it boils down about the mayor of Amsterdam, there's only one mayor. So at that moment there should be a rule in the API that say, well, we can answer this question, but if there is only one or a limited number of persons who are fulfilling that type of job.

00:10:47 Interviewee

We should not give that answer

00:10:52 Interviewee

Did you get it?

00:10:53 Interviewer

Not completely

00:10:58 Interviewee

So what you have you have you ask, for example in an open data environment, give me all the inhabitants of amsterdam With their profession and you aggregate on profession, you wanted to know how many software engineers do we have? How many artists do we have? How many doctors do we have medical doctors?

00:11:25 Interviewer

And there should be a business rule that there is only one mayor?

00:11:28 Interviewee

But yeah, exactly. And if there is only one major, then maybe you should not give an. Those who request for output professions, you can give them the other profession, but you should not give them. Well, we have major one because then everybody knows we are talking about Femke Halsema.

00:11:50 Interviewee

You see, and that in the APA API you have to build this kind of mechanism. So actually it boils down that if you are working on on a GDPR and privacy and this kind of mechanism, it's not only at the backend but also on the front front end. You should have checks.

00:11:55 Interviewer

OK.

00:12:10 Interviewee

And this is an example, yeah.

00:12:11 Interviewer

OK. Thank you very much.

00:12:14 Interviewer

How are technical development boundary resources such as software development kits or tools used as a governance mechanism to promote openness in data platforms?

00:12:26 Interviewee

Well, the technical issues, well they they are fairly useful because if you have the choice, if you have a data set and you have the data set and you say, well maybe some of these Are very sensitive, so that could be a reason not to publish the whole data set.

00:12:50 Interviewee

But if you have technical and mechanisms that helps you to anonymize some of these attributes, then you can still you can still use this data set partly.

00:13:06 Interviewer

OK. Do you have order order examples how how these development boundary resources such as software development kits or tools as used as a governance mechanism?

00:13:19 Interviewee

Yeah, sure. Aggregation, aggregation tools that you say well . We want to know something about and let's say we want to know something about.

00:13:33 Interviewee

Criminals or something like that.

00:13:36 Interviewee

But you cannot give the name and Crime that they have committed.

00:13:48 Interviewee

But if you can aggregate it, you can say, well, we have so many people living in Amsterdam that committed this type of crime, then that might be helpful. So aggregation might be a very useful tool.

00:14:06 Interviewer

OK.

00:14:08 Interviewer

How are non-technical social boundary resources such as the documentation, training or support used as a governance mechanism to promote openness in data platforms?

00:14:20 Interviewee

Once again, because you have many.

00:14:23 Interviewer

How are non-technical social boundary resources such as documentation, training or supports used as a governance mechanism to promote openness in data platforms?

00:14:36 Interviewee

Yeah. Well, what you have documentation and these things are very important because if you have a data set, you have openness, you have open data set, then still, if you do not have the documentation, if you do not know what the semantics are behind the data set.

00:14:55 Interviewee

then it's very hard to understand what it actually means and if you do not, can really understand what such an a set means. You cannot apply it properly. So to give you an example, suppose we have a data set in which you have an A field salary.

00:15:15 Interviewee

And you have the salary and someone is interested in the average salary of people living in a certain neighborhood. And they look at the salary and say, Oh well, it's very low.

00:15:27 Interviewee

They're very poor.

00:15:29 Interviewee

But if you have some documentations that tell you, well, that's the salary that's actually part time salary.

00:15:37 Interviewee

Then the interpretation of the end result will be in a different way. Then you will say, well, maybe they are not so poor. They are very rich. You see. So the documentation is very important. But also talking about the semantics of the data which order, that's quite important. But I know from practice that data sets are very poorly documented.

00:16:03 Interviewer

I heard it before in in my job and in this research as well.

00:16:09 Interviewee

Yeah, but it's quite important and it is important to have the proper interpretation of the data. The semantics is very important to have the proper interpretation of the data.

00:16:24 Interviewer

About platform rules, how do intellectual property rights, such as licensing, play a role in governing openness in within data platforms? So did you understand the question?

00:16:36 Interviewee

Yeah, yeah, I think, but but please.

00:16:39 Interviewer

So how do intellectual property rights, such as licensing, enable openers in data platforms?

00:16:49 Interviewee

Actually I do not have a clear answer on that, but I think intellectual property can help you To improve the quality of the data, because if we know that someone is responsible for the data.

00:17:09 Interviewee

Then I assume if you are, if everybody knows that you are responsible for the data, then probably you will be more careful about the quality of you will take more care of the quality of the data. If it's anonymous then.

00:17:26 Interviewee

Nobody is actually responsible for the data, so in that Sense maybe intellectual properties may be helpful.

00:17:34 Interviewer

And in terms of use, I should look as licensing as you can use this data so you get a license to use it. How does it play a role then?

00:17:44 Interviewee

Well, if you have a license to use it, the data, the more the data is used.

00:17:50 Interviewee

The better it is actually, because if more people are using the data, they will discover the flaws of the data and if we have these flaws then we can use it to improve the data. So and.

00:18:05 Interviewee

Licensing of this data, Will also have something like, well, there is some quality control mechanism behind the data.

00:18:15 Interviewer

OK.

00:18:18 Interviewee

At least that's that's people may think. I don't know whether it's true, but it's a hypothesis.

00:18:24 Interviewer

Can you explain or elaborate on the use of decision rights such as terms and conditions as a mechanism to enable openness in data platforms?

00:18:34 Interviewee

Yeah, they're actually all these things balls down in and gaining the trust of people who are using the data. So if you have all these mechanisms and they are working properly then and people will trust the data better. And if they trust and if they trust the data they will be tempted to use the data, so it it all these mechanisms may be very useful to create trust about the data.

00:19:11 Interviewee

And if if people have trusted something, they will use it. But if you do not have trust in the quality of the data or the data set, well then yeah, I can imagine that people will say, well, the data is not very reliable, so I will not use it.

00:19:29 Interviewer

Can you provide insights of gatekeeping practices such as access rights and data platforms, and the impact on enabling openness.

00:19:38 Interviewee

Yeah, access rights and the relation that's opening the data well.

00:19:52 Interviewee

You know you have two things you can access the data.

00:19:59 Interviewee

And you can provide people all kind of rights about the data, but more interesting is always how people will use the data.

00:20:12 Interviewee

You see access is 1 But use control.

00:20:17 Interviewee

Is even maybe more important.

00:20:20 Interviewee

So I would say well.

00:20:24 Interviewee

Give them access to all the data.

00:20:26 Interviewee

But an Monitor about the use of the data.

00:20:33 Interviewee

So use control is coming in place.

00:20:39 Interviewer

Uhm, how do pricing mechanisms contribute to promoting openness in data platforms?

00:20:45 Interviewee

While I don't think that we are willing to pay for data.

00:20:50 Interviewee

At least from a research point of view. yeah. But maybe in a commercial point of view that might be different.

00:21:00 Interviewer

And this might be well applied to that as well. Then can you explain how revenue sharing models are implemented in data platforms to enable openness?

00:21:10 Interviewee

I have no idea about how they are implemented to be honest, so I skipped that question.

00:21:12 Interviewer

OK.

00:21:14 Interviewer

OK. And then about the ecosystem identity.

00:21:21 Interviewer

Can you explain how relational control and I define that as relational control refers to the degree to which the platform owner relies on norms value values that is shared with users and app developers to influence their behavior and can be divided into self-control and clan control.

00:21:43 Interviewer

So can you explain how this is utilized to foster openness in data platforms?

00:21:48 Interviewee

Once again, because I have missed the definition of out, yeah.

00:21:51 Interviewer

So it's, it's about the ecosystem identity, so.

00:21:56 Interviewee

What do you mean by the ecosystem identity?

00:21:57 Interviewer

The platform data platform ecosystem, so the platform owner, the the data users, they comprise the ecosystem ecosystem.

00:22:07 Interviewee

They are all stakeholders

00:22:16 Interviewer

Yes and how is the platform owner? How he relies on norms and values that they share with the the ecosystem. The total of ecosystem. So how is this relational control by the platform?

00:22:33 Interviewee

Yeah. Yeah, but yeah, yeah, you can say two things about it and one is that you say now we have some a set of rules and values and and the data owner, they should enforce that everybody's obliged to these rules and values. Another one is that you say well.

00:22:48 Interviewee

Another one is that you say well we are all peers.

00:22:56 Interviewee

So what you should have you should say OK if you are all peers then we have all a shared responsibility.

00:23:04 Interviewee

And and actually what you should do is if you look at these ecosystems, then you have a set of roles and responsibilities and you should divide or share these responsibilities to these stakeholders of these ecosystems. That's one of this model that you could follow.

00:23:26 Interviewer

And then to data specific mechanisms, we covered this already briefly, but what are data specific mechanisms used to manage data platform, openness. So do you have more examples instead of encryption and aggregation?

00:23:52 Interviewee

Well, actually utility.

00:23:56 Interviewee

As a measure is also important to say, OK, we have this data data set and what extent is it useful and for what purposes is it useful to give a users or stakeholders the insight for which purposes is this data useful? Because not all data is used for for all type of applications.

00:24:27 Interviewee

Yeah, having insight in for what type of applications or what for type of question which data is useful is also I think important.

00:24:40 Interviewer

We're coming to the end of the interview. Is there any other governance mechanism related to platform openness that we haven't discussed yet?

00:24:50 Interviewee

Well, it's a difficult question. I don't think so. I think we have discussed quite important things. We have discussed privacy. We have discussed security. We have discussed quality and we have discussed, we have discussed a number.

00:25:02 Interviewer

Are there any additional insights or perspectives you would like to share regarding the mechanisms that we covered?

00:25:10 Interviewee

Not really right now.

00:25:11 Interviewer

OK.

00:25:15 Interviewer

Are there any missing elements you'd like to add before we conclude this interview?

00:25:20 Interviewee

Let me see.

00:25:26 Interviewee

Well, not exactly. I think I have bring up a number of these issues that I think it's important for data platforms and data sharing. Data sharing is an important one.

00:25:37 Interviewer

And then do you have any recommendations or suggestions to improve the openness of data platform based on your expertise?

00:25:44 Interviewee

Yeah, and well, the utility because we have now a lot of open data sets, but they are not really used.

00:25:52 Interviewee

And maybe in 2019, we proposed in an article to say, well and to redefine the term open data into semi open data.

00:26:12 Interviewee

Because not all data can be open because some of them are very sensitive, but part of the data set can be open and they can be still useful. So what is important now that we should find the bottlenecks? Why?

00:26:29 Interviewee

While there are so many data sets open, but they are not really used.

00:26:35 Interviewer

And what is your suggestion? How this can be facilitated?

00:26:42 Interviewee

Well, we have, we have brought some of these suggestions on the table is one of them, is the semantics of the data exactly know what the data means.

00:26:54 Interviewee

We have poor documentation, but still maybe via reverse engineering we are able to say what exactly each field of the data means. You see that's quite important.

00:27:09 Interviewer

OK.

00:27:13 Interviewer

If you have nothing more to add

00:27:16 Interviewee

Yeah, maybe one thing one thing, another parameter can be quite important if you are working with data set is well.

00:27:26 Interviewee

Maybe and have something like how complete the data set is completeness of this Data set.?

00:27:32 Interviewer

So then you're talking about data quality.

00:27:35 Interviewee

Quality is one of the, but that's on the level of field on the attribute level. But in general, how complete are this data set and interoperability of our data set? We haven't talked about that, but that's also an important one, because if you have just one data.

00:27:55 Interviewee

That might be interesting, but more interesting is if the data if data sets are interoperable. So interoperability is are quite important and the issue in using this open.

00:28:08 Interviewer

how can that be realized?

00:28:13 Interviewee

Yeah, yeah, yeah. Yeah. Well.

00:28:16 Interviewee

Also about the documentation, it's also about the documentation, but also that you say OK, those who deliver the data.

00:28:26 Interviewee

Maybe they can. They can take into account. Well, how can I make the data such? How can we prepare the data such that it's interoperable with a number of other set of data?

00:28:40 Interviewer

So you're talking about standards, certain standards, standardization.

00:28:43 Interviewee

Yeah. Yeah, it identifiers. Uh, standard identifier. You see that kind of thing, but.

00:28:49 Interviewer

Do you have examples of?

00:28:51 Interviewee

Well, we have example of them, but for example we have the BSN number that's a number that can be used, but OK, that's fairly privacy sensitive, but we actually every technical number That says OK, this set of record is unique May be useful. But within the ecosystem you should define that together. Yeah, you see, that's that's maybe one important thing for such a such a thing. Well, maybe that's one of the important task of such an eco ecosystem to say, well, how can we implement interoperability?

00:29:37 Interviewee

Because that's quite important.

00:29:39 Interviewer

Perfect. Thank you very much.

Transcript Interview 6

8 November 2023

00:00:15 Interviewer

Perfect. Do you have any questions before we start the interview?

00:00:21 Interviewee

No, you can start, that's fine.

00:00:23 Interviewer

Perfect. What is your job function?

00:00:27 Interviewee

Yeah. And which domain of it?

00:00:30 Interviewee

I'm I'm a CEO of some companies, yeah.

00:00:34 Interviewee

Most of the time I'm CEO, but my my best part is the technical parts of the CTO function.

00:00:41 Interviewer

OK.

00:00:42 Interviewee

But most of the time, Floris, it's mixed he, and some companies have operations involved. Some may have executive decisions that's it's mixed, but it's, uh CEO, CTO, yeah.

00:00:53 Interviewer

OK, perfect.

00:00:56 Interviewer

Can you describe the data platforms of the organizations you are working for?

00:01:01 Interviewee

Oh yes, describe in what sense you mean?

00:01:06 Interviewer

As what kind of data platform ecosystem it is.

00:01:09 Interviewee

Yes. So let me think about which one then I need to focus on. Uh, I have. When I when I hear, hear the introduction, I have an open API gateway development that we do as well for UM.

00:01:25 Interviewee

An infrastructure company, but that is mainly providing IT solutions to municipalities. So that's maybe a good reference to start from there. When you look at the context. So one of them is (Client), it's called uh, it's in Belgium. They are in Antwerpen and Gent, they are practically doing I think 70% of the municipalities in Belgium.

00:01:48 Interviewee

And they have a platform and it's called (Platform).

00:01:56 Interviewee

We'll keep that one, I guess as a reference because that's the most prominent, I guess for the introduction.

00:02:03 Interviewee

Because it's it's a platform where we have more than 200 suppliers connected to.

00:02:09 Interviewer

As suppliers, as data providers.

00:02:13 Interviewee

Service providers, yes.

00:02:15 Interviewer

OK. And uh, what's what are the data providers in this? Uh.

00:02:20 Interviewee

Yeah, you have a lot of them. So let me introduce how it works. OK, so (Client). They they started with the infrastructure for municipalities and then they went, they moved towards software. But they quickly have a problem that you have a lot of different domains and different requirements from the customers they have municipalities.

00:02:41 Interviewee

And they need to facilitate all those things together. So they were focusing on creating the creation of building blocks, OK.

00:02:49 Interviewee

When you have building blocks, it means that you scope each requirement or set of requirements to a specific supplier and they have a focus and they are good at it. For example, you have a supplier that's dealing with uh, culture, and so the culture houses access for that one. One that deals with asset management, digital asset management, one that can deal with IoT platforms and how to work with electrical charges for example.

00:03:14 Interviewee

And because you have very broad domain of requirements, they went and they defined or they would like to have and they have it for now. They would want to have an API ecosystem. OK, so microservice ecosystem and their and their in their space and that's that's that comes along with a lot of data.

00:03:34 Interviewee

And a lot of organization and governance. So I think that's a good way of talking about that one, OK.

00:03:41 Interviewer

Perfect. If other examples pop up then just.

00:03:46 Interviewee

Another another example, when I mentioned to you about the transportation business, we have a lot of regulations to follow, which is European Commission regulated stuff, but even worldwide and you typically have also a lot of governments to do because you have different parties with different objectives coming together and they need to work together and that's always a tricky part. That's also a good example. Uh. Different specifications, different focuses, different data or semantics about data. And you need to consolidate them as well. So that's a second good example.

00:04:18 Interviewer

OK.

00:04:19 Interviewee

I can give you even more, but let's stick to those.

00:04:22 Interviewer

OK. Do you maybe have experience with data marketplaces as well?

00:04:27 Interviewee

Yes, yes, yes, of course. Yeah. I built, we have built interest from team all of the companies and API gateway with the developer portal and an API marketplace as well, yeah.

00:04:39 Interviewee

And the core engine is now one of the biggest ones I think in in the world, which is a Kong API gateway that rings a bell.

00:04:47 Interviewer

Doesn't ring a bell

00:04:52 Interviewee

And my involvement involvement was quite high on that one. Even so, yeah.

00:04:56 Interviewer

OK. As in?

00:04:59 Interviewee

As an even writing and reviewing books or literature about the solution.

00:05:04 Interviewer

OK.

00:05:05 Interviewee

Yeah, not only building it, but also teaching it.

00:05:10 Interviewer

Perfect. How many years of experience do you have in data platforms?

00:05:14 Interviewee

Def 14 years, I guess something like that.

00:05:17 Interviewer

14 years, OK. And can you explain what kind of experience you have for data platform?

00:05:22 Interviewee

Yes, I have. I had experience on the chip card market with everything that has to do with yeah, with the traditional chip card applications.

00:05:34 Interviewee

That's one thing. I had to experience in the beginning with ECM's, so content management systems like Documentum and all fair school, those kind of those domains where you have to also uh, create knowledge bases for specific customers and and to create a semantical value on top of the data.

00:05:56 Interviewee

I have experience with time series data a lot where you need to measure stuff and also statistically uh provide insights in data. I have experience in benchmarking and ML where you need to for example benchmark machines like high pressure machines or heat pumps and stuff like that. I have data, experience and governmental data, but also in security documents where you need to provide verifiable claims for a specific attribute or some traits that the individual or a company has. Yeah.

00:06:35 Interviewer

And in general, what what's your role in, in, in in those?

00:06:42 Interviewee

Most of the time I always have a technical architectural role as well in those domains, so most of the times when I did something and the the ones I mentioned to you I had also I was also the technical architect of those.

00:06:54 Interviewer

What governance mechanisms have you observed in data platforms to facilitate openness?

00:07:03 Interviewee

Yeah, open API specification that was coming up in that time, which was typically the I think open it. It was not called the Open API, it was called the Swagger version two. I don't know what transformed to open API when I when I in the beginning when that happened. So the what happened is actually that the monolithic systems they went away.

00:07:25 Interviewee

And everybody was jumping on the train of having event based domain driven systems with microservices and Mm-hmm. And like always, they think, uh, we saw the light and that's that's the golden Grail. But that's not not always the case because a new way of working also means a new set of issues that you potentially have. So the governments techniques to to transfer an organization or more multiple organizations and suppliers from a monolithical principle to a microservice based domain.

00:08:00 Interviewee

That was quite tricky because you have to educate them as well. You have to learn the concepts and the principles.

00:08:07 Interviewee

One of the things that was hitting the wall very fast is the I call it the ubiquitous language. I don't know if it makes sense for you, but that's the that's that's generating or or creating a a language which can be understood and spoken from with the stakeholders and the technical people because most of the time you have discrepancies between it.

00:08:28 Interviewee

So I worked a lot on defining the language so meaning that when I talk about something, the other person in another domain, the domain expert for example, and the technical guy, they understand the same, they have the same meaning of that word.

00:08:42 Interviewee

That's a very stupid thing.

00:08:44 Interviewee

But a very important thing.

00:08:45 Interviewer

You know, I mean in, in data governance terms, so making sure that the definitions are understood and are the OK, perfect.

00:08:52 Interviewee

Exactly because most of the problems when you don't tackle that one in the beginning, it's something that lives, it's continuing, it's gross. But if you don't tackle that, it can be.

00:09:03 Interviewee

Uh, a very big problem.

00:09:05 Interviewee

Yeah, that's one of the things. But then you have a lot of systems where you can go. Government is very broad term, OK? So when I was working on those cases, we typically were already in the era of agile and scrum and stuff like that where we have a separation of the roles. You have the stakeholder. You had some. So we had a stakeholder group most of the time.

00:09:25 Interviewee

you have the development team. They had those kind of cycles with their sprints and retrospectives

00:09:31 Interviewee

And everything that you can imagine about the scrum meetings or Scrum methodology. Most of the time governments means also defining the way how you work. So implementing an agile system to work to tailor it on the team itself and on the use case itself that you're trying to tackle. That's also one of the points so.

00:09:51 Interviewee

Ubiquitous language speaking the same language. Uh, that's one thing. Secondly is how will we work together? So how do we define relationships and the roles and responsibilities? Who will do what's when those kind of things is typically what you also do in the governance setup.

00:10:11 Interviewer

And how does this enable openness in data platform ecosystem?

00:10:17 Interviewee

It's again, openness is defining which kind of protocols, how you work with the the semantics of the data. That's one of the things when you talk about semantics, it has different axes. So we have the axis of what kind of protocols will you use, what kind of interfaces will you use and how will you deal with change or evolutions.

00:10:37 Interviewee

OK, so you have different facets, let's say there which you need to tackle and also define again how you will deal with them.

00:10:47 Interviewer

OK. Uh, you touched them already a bit. Can you provide insights into the use of application programming interfaces as a governance mechanisms for enabling openness in data platforms?

00:11:15 Interviewee

API is how they, but because you are conforming on an interface, so when you when you look at the API or rest APIs or GRPC APIs, whatever the goal is that you have a deterministic interface that can be used by somebody else. That's the theory behind it. But most of the times you have to also follow some rules.

00:11:38 Interviewee

Like idempotency. I don't know if that rings a bell or.

00:11:42 Interviewee

To you need to define how your API.

00:11:44 Interviewee

They react what an API does. Well is that when you document this well and you have a clear API which you manage which you control, it's openness because you I can just choose and say, OK, let's go in the marketplace. I see an API that's interesting. Me. I consider it a black box. I will just use the API. I fill in a I want to have B.

00:12:05 Interviewee

OK and I can manage that one. That's one thing that's purely when you look at the interfacing part, but it's not only about that, it's also about.

00:12:15 Interviewee

Updates documentation. What happens when you have a new version? What happens when you have an impact or you're not backwards compatible? So very fast. An API itself is not sufficient, you need also to have an API ecosystem which you manage. You see. So the openness of an API, in my opinion, it's very hard determined by the ecosystem that you are using around the API's so yeah.

00:12:38 Interviewer

Uh-huh. And about that, how are non-technical social boundary resources such as documentation where you spoke about or training or support used as a governance mechanism to promote openness in data platforms?

00:12:54 Interviewee

But they're just merely examples to get you speed up very quickly with and then again documentation. Big question Mark and you have a lot of. That's something which is really vague. You don't have a really standard way of documenting, to be honest, but documentation can also raise confusion if it's not well done in my opinion.

00:13:11 Interviewee

But if it's straightforward, simple with some code examples with some easy teaching mechanisms inside of it to get you really fast from A to B without knowing all the details, that's good documentation that can help for sure. Yeah, you know what? What? What I like a lot is when you look at a technical example, when you look at the technical example, I don't know if you know the rust language or something which is rather gaining popularity at the moment, rust. No.

00:13:38 Interviewee

You know you have Java and Scala and all those things C++ and C and then go and whatever. But what they like is is how Rust documents stuff which I like a lot. You have a principle in API management which is not tackled all the time in an ecosystem and it's called, I don't know what it's called. It's like how is that?

00:13:58 Interviewee

Wait, let me look. Look, look it up. It's Contract based.

00:14:05 Interviewee

It's from Martin Fowler. Do you know Martin Fowler?

00:14:11 Interviewer

No, I don't know.

00:14:14 Interviewee

That's that's a guy who is basically giving recommendation. He's rather known as well, but recommendation on how to deal with architectural problems and and an API driven microservice. But it's a little bit broader than that. But he's also he is very actively there. I will send you a link here, OK.

00:14:31 Interviewee

Just to give you a stupid example, when you look at this one consumer driven contracts, what it actually means is that when you are implementing my API you have an API with a set of 10 endpoints for example you can do 10 things with my API but you're only using one.

00:14:47 Interviewee

OK. How do you know? I know that you're using my API because I defined a well thought ecosystem around it and I see that my API is used by ten parties and I have this amount of transactions and everything runs fine. But if I talk about impact and maintenance, I go from a new very I go to a new version then it means that your contracts.

00:15:08 Interviewee

The contract I have with you can be broken. It's actually, uh, a derivative of openness. Anyway. You have openness, open API, but more important, what's the impact when I change it?

00:15:21 Interviewee

Consumer driven contracts is a way where you say you integrate my API and the set of things that I use from the API. I will contribute it back to your API. The owners of this API to please run those tests before you release because when you follow that principle that's one of the principles I also put in place when we had more than 200 suppliers, I forced them to send over the integration tests so that I can make it part of the lifecycle for the API. So when I run that I can say.

00:15:54 Interviewee

50% can migrate without issues. This 50% has issues. How can they fix that by doing this and that and that and that's part of my documentation as well you see.

00:16:04 Interviewer

Yeah, I understand.

00:16:06 Interviewee

OK.

00:16:07 Interviewer

And how are technical development boundary resources such as software development kits or tools used as a governance mechanism to promote openness and data platforms?

00:16:20 Interviewee

You have two ways there, right? You have the the technical people who says I can make it all by myself and you have the technical people that's used SDK as the case on the API side to me you know.

00:16:32 Interviewer

Yeah. So as as a platform boundary resource, yeah.

00:16:35 Interviewee

Yeah, yeah, yeah. OK. As an SDK on the API. Yeah, that's it has purpose. In my opinion it has its purpose for backwards compatibility. The thing I just mentioned before with the consumer driven contracts, you only have one place to solve it. It's an SDK because.

00:16:55 Interviewee

When you don't solve it there, it means you hit the the consuming application site and he needs to change it. So that's one way of dealing with that. So it has its use SDK because you can fix a lot of stuff in an SDK like what you have with the BFF kind of principle back and forth front end you can fiddle and fix things or change stuff to uhm for the the compliance with the existing solution you can fix it there in SDK. SDK's also not only to compatibility or backwards compatibility. It also solves the issue of more complicated sequences or transactions and so for example, I just need to authenticate somebody, but then you need to discover a reader. You need to fix check if he has a key or a certificate with the signing attribute in it. You need them to upload a document and then you need to sign.

00:17:48 Interviewee

It becomes complicated. So SDK's can ease the way of integrating the use cases.

00:17:55 Interviewee

By removing all this complicated requests that happens on the API itself, and that's the second point, and a third point for SDK's, which is very good, I think, is that that's the part that you can, uh, it's also open source. You can have a lot of contributions on the SDK's as well, which can be interesting. So you don't forcibly need to. You're not the only one in control of 1 development team or a DevOps team or whatever. While in SDK and most of the time is something that can change faster by contributing.

00:18:28 Interviewer

OK. Yeah. How do intellectual property rights, such as licensing, play a role in governing openness within data platforms?

00:18:40 Interviewee

Very important, but not only licensing. Yeah, so licensing. So dependency management is one of the crucial things that you need to have. You need to be aware of what dependencies do I use and what I'm a little bit strict in those but and.

00:18:57 Interviewee

When I'm technically involved in a project, nobody is allowed to add a dependency without going through an architecture report. Nobody, because that's typically something that's technical guys, they don't know very well the details about. That's one thing. So very important thing, also technical people, they need to sign an additional addendum in the contract saying that when they do something, their IP

belongs to the company or the product and so on and so on, but it's all it all has to do with the governance of your intellectual property and also the value of your product.

00:19:31 Interviewee

When you make mistakes in that section, it's possible that you will have a big problem when you want to sell the company, you don't start a company without willing to sell or do a capital raise or involve investors. So if you don't manage that block, you're not doing good business, that's simple so.

00:19:50 Interviewee

That's a very important thing for a lot of startups, to be honest. They don't look at the dependency management. But today, before it was difficult. Today have so many tools that deals with that and that's great. Maybe to come back on the Rust thing I mentioned you.

00:20:06 Interviewee

it's an interesting one to give another example because the way how they deal with API interfaces not from a rest API point of view, but a library interface for example, they do it very well because they don't only expose the documentation, they have a way of validating the documentation and the code examples in the documentation before pushing the release.

00:20:31 Interviewee

That's the principle which I like a lot because I'm reassured if I use that library that the documentation is still up to date and the reality when you look at the rest API and the service has an evolution that it goes to a new version. But the documentation it stays behind, that's typically what happens so.

00:20:52 Interviewee

If the tools that you are using is enforcing you to work well.

00:20:56 Interviewee

That's a good way of having a governance set in place, so it is a factor for me to say. Let's use that programming language for system libraries because I know when I do that I force the people to follow the conventions you see. So that's a good way of of dealing with.

00:21:16 Interviewee

You cannot assume that technical people, I have 350 developers, for example, you cannot assume that all 350, that they will follow the rules. You cannot do that. So we need to put stuff in place to help your government governance. And that's one of the the tricks where you have a lot of trick.

00:21:34 Interviewee

Yeah, but some of the tricks are.

00:21:40 Interviewer

Could you elaborate on the use of decision rights as such as terms and conditions as a mechanism to enable openness in data platforms?

00:21:50 Interviewee

Yes. And it's also a little bit linked to the definition of your license. It depends on your license and together with your license terms and conditions are aligned with those.

00:22:01 Interviewee

So it it all comes down back to ownership of the codes or or a financial impact for me to be very open on that. That's not my expertise. I just have my lawyers and I push it to them and I said that's what I want. That's how I want to put it in the market. That's following my marketing plan or the commercialization plan and they need to make it work.

00:22:22 Interviewee

You know, because if you have for example a product and you need to choose between an AGPL V2 or an MIT license.

00:22:29 Interviewee

For technical people, they say, well, it's quite the same. It's not. It's a big impact. So terms and conditions, even when you have a selling proposition or an integration project with this, when somebody will use your interface, that's something you need to fix and you need to pull it early to, you need to have it formalized very deterministically, yeah.

00:22:49 Interviewer

OK.

00:22:51 Interviewer

Can provide insights of gatekeeping practices such as access rights in data platforms and that impact on enabling openness?

00:23:01 Interviewee

Uh, yes, that's still a mess sometimes today, but you have good ways of doing it most of the time, everything one of the most important stuff when you want to, even when you use role based access control, you always fall in a situation where you need some kind of roles and some kind of permissions which can correlate down to all those interfaces. If I just look at the API rest interface or the microservice ecosystem, that means that you must have an identity provider set in place in your architecture.

00:23:34 Interviewee

Means one component owning the authentication and the role and the permission part. That's one thing which is important.

00:23:43 Interviewee

We have an evolution, of course, and that domain, you know what I mean with identity provider, I guess, yeah. You have an evolution in that domain when you look at Envoy and side cars and that side cars take over the additional role of enforcing policies or access policies, but still you always need to have one place where you can manage those.

00:24:06 Interviewee

When you look at the microservice ecosystem, that's not so easy because everybody in its own contextual domain, they define things differently. So you need to have a mapping you need to solve the mapping issue. There you see typically how it goes is that in your central identity provider you define.

00:24:26 Interviewee

Yeah, roles and permissions to keep it simple and you inject or you create a JSON web token or a token which is transportable and signed by this issuer and you make sure that everybody can or is enforced that they validate it before they apply the the token.

00:24:43 Interviewee

And that's one of the ways to do it. But I can think of 10 other ones how to do it. So I just stick with that.

00:24:54 Interviewer

How do pricing mechanisms contribute to promoting openness in data platform ecosystems?

00:24:59 Interviewee

Yeah, a lot! Yeah. It depends on how you see it. If I have a product I have a I still have a product, but when I launched the product in the market, it's an open API as well. It has to do with a very complicated domain, of European regulation about adas. I don't know if that rings a bell adas, it's the European legislation about both digital signatures.

00:25:23 Interviewee

When you push something like this open, uh, it means that, uhm, you don't go for the financial thing at the moment. That's something that you postpone, you go for the distribution of that, because it depends on how fast you want to go in the market. If it's an open API, you can say by just copy it over and have it run by yourself.

00:25:45 Interviewee

And that's not the same definition. An open API means I open my knowledge through an interface to have something created with that or put added value on top of that. That's the goal. If you want to have added value on top, lower the price, go for distribution, open your data and let's go for it

00:26:04 Interviewee

Your business model will be tailored then to have a game model. Once the integrations are done, so more in the partnership relation, so it's different.

00:26:11 Interviewee

Yeah. So pricing of open APIs definitely impacts a lot how people will use it or not use it. It's very simple. I see two APIs doing the same stuff. The one is I have to pay for the other one not but it lacks one feature. I will contribute that feature to the API and it will use it for free, so it impacts a lot you know.

00:26:36 Interviewee

Yeah, maybe. I forgot on the thing before there with the security roles. Don't forget we are going towards a culture and technology where you where you can lease your rights. That's important as well. So it's I was talking about roles and permission in a very static way.

00:26:56 Interviewer

We are going away from the static way and going towards a dynamic way of dealing with access rights. For example, when you are working for a company as a consultant and you need to have access as an administrator to 1 server for example.

00:27:10 Interviewee

Uh, today they will grant you that access. You will be able to access it anywhere, anytime, whatever you want to. OK, that's not the way that you need to treat data like this because you have again the legal complexity the GDPR ways. You need to make sure that my employees, for example when the access a server.

00:27:30 Interviewee

They will have a leased token that gives them access for that role and that permission for one hour, for example. You see and that's the way how we need to evolve and not with the static roles and permissions and stuff like that, OK.

00:27:44 Interviewer

Does this also applies for data marketplaces for example?

00:27:48 Interviewee

For marketplaces, it depends on if you're on the publisher side. Maybe yes, because you want to make sure that when you release something, it's something that you control. Imagine you have an API. 20 partners are using it and are depending on it and you have one guy that messes up something and pushes it in production. But you need to manage that as well.

00:28:08 Interviewee

You need to have your life cycles or product life cycles well defined and your access rights as well. I don't I don't do that. I mean, I have people that needs to deploy a new version of production for a bank. They need to receive a grant in the form of a key that gives them the possibility to do that task specifically for that amount of time, and it's not that they're pushed on timing.

00:28:32 Interviewee

They have time to do it, that's OK, but it means that you audit as well your life cycle. It's not only auditing your application and the transaction and stuff like that. It's also auditing the sanity or life cycle of the product or your API. So I know I mentioned a lot of stuff and but that's important. You cannot do everything at one time. It's something that you build up step by step. So yeah, that is important, yeah.

00:28:59 Interviewer

And can you explain how revenue sharing models are implemented in data platform to enable openness?

00:29:06 Interviewee

Yeah. Typically you work with plans and each plan can be attached to some financial model.

00:29:12 Interviewee

A plan typically have the freemium models which were boosting up. I think five years before, but.

00:29:19 Interviewee

Yeah, OK. That serves their purpose as well. But typically when you have a plan and you have a financial model, depending on the functionality that you expose, you can let people pay. The only benefit you have is to again onboard. It's a different way of distributing openly. Yeah. So if I look at a product and I can use it already for free.

00:29:39 Interviewee

But with the limited amount, and once I gain money in my business model, I'm comfortable by paying also to have the full tier of the product. Why not? That's a that's a feasible plan. I think it's opinionated. Again, a lot of people are looking to that from a different perspective.

00:29:53 Interviewee

But in my opinion it's OK, you can say I like to start free and then to pay as long as I know my cost upfront, that's OK.

00:30:03 Interviewee

But that gives you the possibility to try out different APIs or different solutions and see whatever suits you best. So that's a good way of doing it. The freemium model is OK.

00:30:12 Interviewee

It's a good model, yeah.

00:30:14 Interviewer

And in terms of data, for example, if you buy a data set or you obtain real time data.

00:30:25 Interviewer

How does it apply in that way?

00:30:31 Interviewee

I didn't understand the question, so data coming from the service or what?

00:30:35 Interviewer

So in data marketplace for example, and you have several data providers, publishing their data and the data consumers can buy that data, how does revenue sharing models apply in terms of data platform openness?

00:30:59 Interviewee

Yeah, that's the same and that's about sampling and sets and data sets. When you still have the possibility to have small sets rather to check something fast and when you want to train a model for example, and you buy small sets, your accuracy will be low. If you are willing to pay because again, your business model is supporting that.

00:31:17 Interviewee

That's a feasible and a good solution for me you. You need to do something. It also depends where the data comes from it. So when the data comes from an added value from that party that wants to earn money on the data set. It's agreeable if the data comes from people stupidly participating in social events, and that data has been sold as a data set that's more an ethical question, which I'm against too, but that's an ethical thing. That's not the same.

00:31:51 Interviewer

And then in terms of the ecosystem identity.

00:31:54 Interviewer

Can you explain how relational control is utilized to force openness in data platforms?

00:32:03 Interviewee

Like I need to understand it well because identity is my thing. So.

00:32:11 Interviewer

I will. I have a list of definitions. I will provide you with the definition to be sure you have.

00:32:22 Interviewee

Yeah. So wait, you mean, what about sharing or the identity information I contribute? I give towards the system how they need to tackle it or how how I think about it or.

00:32:34 Interviewer

Exactly. So how does this? How does this promote openness? So it's the first two.

00:32:39 Interviewee

It's it's it, it doesn't it. It's my sharing my identity with an API. If it's not from a business context, it doesn't promote anything. In my opinion it's only abusing the information that you sent over. I'm.

00:32:55 Interviewee

I'm more the Robin Hoods only entities, so I I don't like central solution for identities. Let's start with that. I will give you an example.

00:33:04 Interviewee

When I go to, I want to buy alcohol online. Liquor. OK, I need to prove that I am more than 18 or 21 years. And so the guy on the other side will ask me, OK, give me your. Even when it's an interface that I use and so on. Give me your birth dates and show me that you were born.

00:33:30 Interviewer

No, I I manage more in terms of identity as how it relies on norms and values. So the ecosystem identity as in the degree to which the platform owner relies on norms and values that that they share with its data consumers or app developers.

00:33:55 Interviewee

Ah, that's different. Yeah. OK, that's from the other side. It's not to the identity sharing they do in those systems. Yeah.

00:34:00 Interviewee

I I think you have to the point where I mentioned about the identity, the wrong answer on your question is, is even relevant.

00:34:10 Interviewee

Yeah, it's relevant even when you have open interfaces, avoid you need to avoid only for the GDPR already. Yeah, and you need to avoid the need to opt out. Avoid that one at the beginning from your interfaces, especially in open APIs. OK, but that's another thing. It's not in your list of questions.

00:34:30 Interviewee

Yeah. So the the the behavior of the owner of the platform and.

00:34:38 Interviewee

Yeah, I'm not fond of the control that can change the way how you need to act with the platform, but typically when the ownership is not well you will not stick to the same platform you see. So that's a different definition of governance and that's a different definition of how to approach your ecosystem that you built. When you have an open API marketplace, you have several people. You have mashape at the time. I don't know if you know that name, mashape.

00:35:09 Interviewee

It's it was a marketplace API marketplace. You have you have different marketplaces that were wrongly managed.

00:35:16 Interviewee

Uhm, not only on on the the way how the people around it manage it or misuse the platform, but also during the vetting of the API's that were enabled on the platform. If you leave everything open, it's not always a good idea to define quality by onboarding services.

00:35:37 Interviewee

Yeah, it's a very broad topic and you see what I mean.

00:35:41 Interviewee

In my opinion, I'm not, uh, it's best to be neutral in those things, but to to have things carefully vetted, you know, carely, on board, if you need to have a monitoring thing on it, is it from a legal perspective or something else that that's very important microservices, they can also or open API's. They can also lead to what I mean with leaking of personal data. That's typically when you had everybody that needs to opt out. People. That's a very difficult topic with an API or an OPI microservice marketplace. That's very difficult to do.

00:36:16 Interviewee

Because you are enforcing people to actually be able to opt out from each service independently, which today is never tackled or not so much tackled in those ecosystems. So yeah.

00:36:29 Interviewee

I hope I answered your question, but it's a very rough thing to answer on.

00:36:35 Interviewer

It's the question that was most hard to answer for the most participants.

00:36:43 Interviewee

Yeah, It's also a very opinionated answer. So like some other questions before as well, it's very opinionated. I'm against bad management of marketplaces and I'm well aware of how they misuse data.

00:37:03 Interviewee

And I'm well aware of the investors behinds the so told open data platforms and that can I, I don't know a lot of good examples of those to be honest. You know it's also because in one moment you need to earn money, OK?

00:37:20 Interviewee

So the moment you have investors joining typically in a second capital raise or a third one, then things can get really screwed up. Uhm, the example I gave you with (Client) in Belgium, that's a good example of how they need to do it well, because the ownership of the the platform and the governance about it is actually monitored by a government as well.

00:37:45 Interviewee

So it stays a little bit neutral. It's something that you cannot start doing crazy stuff there because they will tick you on the fingers and say, hey, you're going one step too far and I think you you need to have the you have a cup form. You know what's the cup form the.

00:38:01 Interviewee

No, no, it's it's the, it's the forum, it's the, it's the, the organization that decides on the browser requirements. So if you have Microsoft and Google and Safari all sitting together you.

00:38:15 Interviewee

Have a virtual organization that controls that. Why? Because.

00:38:21 Interviewee

The whole weapons, depending on those four or five big parties and the the cup form is there because you have a collision between web browsers and certificate authorities. So they create one vehicle on top that is neutral and try to defend also the other side. So the consumers of those products and that's a good principle of governance.

00:38:42 Interviewee

You need to have a neutral independence. Sometimes politically it can get also.

00:38:48 Interviewee

It's sometimes it's tampered with, but OK, you know? But you the idea of having a neutral position, somebody that observes and controls the quality of the governance is also important. Yeah. In banking you have the same payment cards, Visa, MasterCard or the local payment schemes, for example. They are controlled. They cannot go wild and do whatever they want to do.

00:39:09 Interviewee

Even when the technical implementation is very simple, you see.

00:39:13 Interviewer

Mm-hmm. What are data specific mechanisms used to manage data platform openness?

00:39:22 Interviewer

So specifically, we talked about governance mechanisms that also could apply for other digital platforms.

00:39:37 Interviewer

But what are data specific mechanisms used to manage data platform openness?

00:39:43 Interviewee

Uhm, yeah. Protocol implementations and specifications. So very technically. Uh, typically you have a lot of those are rest APIs, so following HTTPS protocol they have an open API which has no API, which is a generalization of the specification which you can share along. They have tooling around it like.

00:40:04 Interviewee

Typically the API client tooling like postman and insomnia and stuff like that. You have the tooling for generating of client codes independently of the code that's helping as well. So we have. It's typically protocols and specifications that facilitates that, yeah.

00:40:22 Interviewee

When, when did you ever do an integration with an with a rest API, for example?

00:40:27 Interviewer

A simple one like to push data into service now.

00:40:33 Interviewee

Yeah. So what you do is you get the API when it's an open API, you crunch it through tool that generates the client interface. You inject the client interface and you can directly implement on the API without knowing all the details of it. Typically the only thing you need to know is how do I authenticate myself? That's always a little bit different and then you're gone with it. So it's protocol and specifications always, yeah.

00:40:59 Interviewer

Are there any other data specific mechanisms that you have observed, used manage data platform openness?

00:41:27 Interviewee

Where to start? Yeah, you have a lot of mechanisms. Yeah. You mean how you treat the data and which mechanisms you use to open the data up. That's what you mean.

00:41:37 Interviewer

To open the data platform ecosystem.

00:41:44 Interviewee

Event buses, for example, that books call backs. You have different ways of dealing with the same problem, but it's not from a rest interface. For example, we have a lot of. Typically when you have a rest implementation, you always have. When you are working in the the domain driven design approach DDD.

00:42:02 Interviewee

Typically you have an event bus, so means that you can subscribe on the bus and you can collect the data based on the event that happens in the system. That's also an openness of the system.

00:42:12 Interviewee

You see, you can have uh data which is going through an algorithm that exports a PDF document with a report that's also openness of data. OK, not in the same way, but you can consolidate the report. You can ask more stuff to generate the differently and you have it a lot in the in the ML use cases as well

00:42:32 Interviewee

So you can also for example say to open up my data. I can host the algorithms and I can feed you with data sets of myself. You can do the way around you have the data set, I can feed you with the algorithm, but I want to keep my algorithm secured and I don't want to share that block.

00:42:47 Interviewee

That sounds like ocean protocol, by the way.

00:42:50 Interviewee

But OK.

00:42:53 Interviewee

So so you have a lot of different mechanisms to open up your way of work, not only data, but also maybe intelligence. You know, you can open up intelligence without compromising the IP of it.

00:43:06 Interviewee

We have different ways. You have so many ways. Yeah, so.

00:43:10 Interviewer

OK. Thank you for uh for sharing. Is there any other governance mechanisms related to platform openness that we haven't discussed yet?

00:43:20 Interviewee

Probably many of them, but OK.

00:43:22 Interviewer

But it's something that pops up right away?

00:43:29 Interviewee

I mentioned one already which you didn't ask explicitly. It's about the individual. The end customer is an important one, which is always overseen. I'm using your API. I'm also a business but t I use a customer at the end there is a lack of protection and personal data and behavior and profiling, so I know it's a business model on itself to profile people and have.

00:43:53 Interviewee

Have them uhm yeah, profiled. But that's something that I don't like you are missing in the govern in the healthy governance takes care of the end user as well, and make sure that the end user is protected. Uh, I know it's not on your questions, but in my opinion it's one of the most important thing that is happening right now, end consumers are not protected.

00:44:15 Interviewee

Technologies go fast, the laws and regulation go very slow. They get abused without knowing it. It's not the fault of the consumer because they don't know the technical landscape. It's the it's the fault of the of the technical owners and the ownership that does it most of all with the financial benefits. So.

00:44:36 Interviewee

You need to keep the ethical thing up in the air as well again. So if you want to do stuff ethically, yeah, you need to govern this as well.

00:44:51 Interviewer

Perfect.

00:44:54 Interviewee

And is that something you learned as well or not?

00:44:57 Interviewer

Yeah, yeah. I ethics was one course of my program and we learned it's there as well in this course. So how, for example, if you build an Machine learning algorithm. How do you make sure that it's ethical if you yeah, apply it.

00:45:27 Interviewer

OK.

00:45:30 Interviewer

But not in terms of an exactly, yeah.

00:45:33 Interviewee

Accessing technology.

00:45:37 Interviewee

Now. Yeah. OK, very good.

00:45:40 Interviewer

Are there any additional insights or perspectives you would like to share regarding the mechanisms we have covered already?

00:45:49 Interviewee

You have a lot of architectural patterns that are involving fast and you need to comply with those.

00:45:55 Interviewee

That's more from a practical point of view. Uh, you have a lack of legal they need to catch up all the time with a lot of stuff they can kill the business case as well sometimes because of a lack of definition, a lack of legal framework.

00:46:15 Interviewee

You also have the typical balance between security and the usability of stuff, but that correlates directly to API management as well. In my opinion. You can put everything open and then don't protect anything, but then you can have other problems.

00:46:39 Interviewer

Do you have any recommendations or suggestions to improve the openness of data platform ecosystems based on your expertise?

00:46:48 Interviewee

Uh, yes, quite some improvements. But I also mentioned the most important ones already I think.

00:46:57 Interviewee

Well, when you look at the practical side of the documentation side is still lacking in my opinion. I gave you the example of Rust because they have this. They enforce you to from a technology point of view, they force you to comply. They force you to make sure that the ecosystem takes care and that they can rely on you.

00:47:16 Interviewee

Reliability is lacking because you don't have. Yeah, when you have an ecosystem about APIs, for example, it's very important to do the vetting like I mentioned, OK, you need to, you need to keep the quality goods and if the quality is good. The rest can be fine as well. If you don't do that, what you typically will see is that when.

00:47:39 Interviewee

You have one or twice an issue with one of the services that are used on the platform. You will quit the platform. It's very simple and you will start doing the same as before. You will create it yourself or push the ownership to a partner and have a partnership with somebody else.

00:47:53 Interviewee

And then you go away from open APIs and you fall into more, let's say, legacy way of working with data.

00:48:03 Interviewee

I think we, so many things. Whatever I mentioned you about the, security as well, not only from a technical perspective but also uh collecting data, sharding data, making sure that it's not, it's the provenance of data, making sure that you cannot trace back stuff to the actual user, or even to the actual source. It's also important, yeah.

00:48:34 Interviewee

We need to have a drink when you.

00:48:35 Interviewee

Say that. Yeah. OK.

00:48:38 Interviewer

OK. Thank you very much.

00:48:40 Interviewee

it's done already?

00:48:40 Interviewer

Yes

00:48:47 Interviewee

My question to you Floris is that you will catch something new that you didn't have before. Or are you still in some stuff, surprised or?

00:48:56 Interviewer

No, but it was.

00:48:58 Interviewer

Probably one of the most technical information that I got. So before I got three interviews with CEOs of data platforms and mostly data marketplaces and they it was less technical than.

00:49:21 Interviewee

I told you I'm also CTO, he haha

00:49:24 Interviewer

I think it's a really, really good addition. It's aligned more I think with the answers that I got from an assistant professor at the university here in Rotterdam, it was I think, well, the answers were not exactly the same, but.

00:49:44 Interviewee

I'm happy that it helped you then, that you don't have three the same versions. Let's say.

00:49:53 Interviewer

I'll stop the recording.

Transcript Interview 7

29 November 2023

00:00:08 Interviewer

Heb je vragen voordat we beginnen met het interview?

00:00:11 Interviewee

Ja, die heb ik al gesteld om even duidelijk te krijgen, waar hebben we het over? Maar het is me nu enigszins duidelijk, ja.

00:00:17 Interviewer

Top, wat korte introductie vragen, wat is je functie momenteel?

00:00:23 Interviewee

Ik ben nu eigenaar, DGA van Company_Int_7.

00:00:29 Interviewer

Hoeveel jaar ervaring heb je met data platformen?

00:00:34 Interviewee

Ja, Als je dus hebt over waar we waar we het zeg maar net over hadden, ja, dan zou ik gewoon zeggen, 7 plus jaar wel meer wat we hier voor mijn volgende werk heb, laten we daar Laten we 10 jaar over maken Als ik mijn vorige werk erbij pak.

00:00:51 Interviewer

Kun je uitleggen wat voor ervaring je hebt met data platformen?

00:00:55 Interviewee

Ja dus, aan de ene kant heb ik het data architectuur gedaan. Dat is echt het ontwerpen van database systemen, dus bronsystemen, maar ook het ontwerpen van datawarehouses, dus stermodellen op basis van bepaalde bronsystemen. Ik heb niet zoveel gewerkt met standaard stukken software. Ik heb eigenlijk veel meer gewerkt op laag niveau dus direct in databases.

00:01:20 Interviewee

De principes zijn hetzelfde, ik zou durven zeggen dat je dan zelfs iets meer moet kunnen, als iets in elkaar klikken zeg maar dus ja. Dus echt het ontwerpen en bouwen daarvan, maar ook daarnaast ook het werken met. Maar ik denk dat het het ontwerp wel een stapje dieper gaat dan puur alleen kijken naar beter werkt. Maar ik heb ook heel veel van dat soort data platformen systemen gewoon gezien. Daar data uitgetrokken voor Qbot datakwaliteits doeleinden.

00:02:00 Interviewer

Zou u dan het platform kunnen beschrijven voor de organisatie waarvoor je werkt of wat?

00:02:11 Interviewee

Ja, Het is niet een platform, dus Dat is dat zijn Natuurlijk bij heel veel klanten geweest.

00:02:17 Interviewee

In dit geval bij bij (Client), wat ik net vertelde, daarvoor bouwen we dat dan neem bijvoorbeeld, dan doe ik ook de data architectuur dus dus Ik ben dan. Ik zit niet op een project, maar eigenlijk gewoon op vele projecten gezeten, zeg maar.

00:02:33 Interviewer

En Als je dan een een project eruit kan zullen pakken wat geschikt is voor deze, voor dit interview, zeg maar wat, wat zou wat wie wie zijn dan de data, gebruikers, de data providers en een platform owner?

00:02:45 Interviewee

Ik zal (Client) pakken, en dat komt omdat dat even recent is.

00:02:50 Interviewee

Dus de platform owner is in dit geval (Client) zelf. Welke rollen zei je nog meer data gebruikers?

00:02:57 Interviewer

Ja, data providers en data gebruikers.

00:03:00 Interviewee

Ja dus data providers zijn in dit geval een aantal partijen die dus product data aanleveren en productdata dat klinkt wat eenvoudig, maar dat is het niet. Dus het zijn allerlei codes voor producten en die zou je verbazen hoeveel verschillende soorten (Client) je wel niet kan hebben, maar iedere keer, stel je ze voor dat iets veranderd In het fles etiket of bijvoorbeeld hoe heet dat.

00:03:24 Interviewee

Of er opeens wel statiegeld bijkomt of dat er opeens een label in alle getallen op zit, dan is het dan weer een nieuw product. Zeg maar. Dus zij hebben een soort van database met allemaal codes van nou, wat hangt met elkaar Samen. Vervolgens heb je ook laboratorium data die hebben gegevens die testen, dus die producten op die productcodes en die laat zeggen van nou, wat zit er allemaal in? Maal in veel breder als dat je Natuurlijk op de fles zelf vindt, zeg maar, maar ze genereren ook zelf data met een platform, dus ze hebben een soort van survey platform.

00:03:58 Interviewee

Waarin ze allerlei vragen stellen en bij bijvoorbeeld bij (Client) zelf.

00:04:02 Interviewee

Die daarvoor bewijsstukken moeten aanleveren, dus dat soort processen en dingen moeten Laten zien. Dus dan gaat het ook om zeg maar om een unstructured data vaak, dus Dat is voor hun ook een uitdagend willen ze enigszins in gestructureerde vorm in zo'n datawarehouse gaan zetten.

00:04:20 Interviewee

We creëren zelf data en ze kopen het in van dit soort professionele partijen.

00:04:24 Interviewer

Oké, en wie zijn de data gebruikers?

00:04:27 Interviewee

Ja In de eerste instantie, nu zij zelf Het is, zeg maar.

00:04:32 Interviewee

Om die rapportages te maken, Maar we hebben een aantal gebruikersgroepen verder gedefinieerd, dus zij zelf vanuit datakwaliteit optiek dus meer van data management perspectief. Dat is een soort van nieuwe nieuwe rol, dus dan dan gaat het niet zozeer om. Kunnen we er rapport van maken, maar hebben we alles netjes vastgelegd.

00:04:51 Interviewee

Zij hebben dan researchers, dus dat zijn gewoon allerlei partijen wereldwijd, Universiteiten die die data pakken en gaan kijken van. Nou kan ik daar wat mee of controlerende functie, klopt het?

00:05:04 Interviewee

Stroopt het ook met ons onderzoek.

00:05:06 Interviewee

Beleidsmakers dat zijn partijen die dus beslissingen maken, niet Alleen Nederland, maar in heel Europa over bijvoorbeeld heel hot op dit moment is suiker en zout belasting die die kijken naar dit soort gegevens. Die willen vaak rapportages, maar die willen ook kunnen zien van hoe komen jullie tot die bevindingen.

00:05:22 Interviewee

Hun eigen management, dus zeg maar, het zijn allerlei investeerders die stoppen geld als donors in zo'n dingen die willen toch zien van nou, wat wordt daarmee gedaan? Wat komt eruit? Dus die hebben dan wel meer een.

00:05:36 Interviewee

Willen kijken nou, wat levert het op? En zeg Maar ja een soort van governance partijen en controlerende partijen, dus dat zijn bijvoorbeeld auditors en die gaan controleren of de bevindingen die dus waar ze komen of die ook wel enigszins houtsnijden die komen, vaak uit wat meer vervelende hoek, dus die worden door (Client) ingehuurd en die die komen even kijken. Dat zijn altijd de partij die niet goed niet goed er vanaf komen. En dan kon ze toch even poolshoogte nemen van wat hebben jullie al met onze dingen gedaan?

00:06:12 Interviewer

Goed, welke governance mechanismen heb je waargenomen die worden gebruikt om openheid en dataplatformen te faciliteren?

00:06:23 Interviewee

Nou, Misschien dat ik wat beter kan gaan, Ik ga het even beantwoorden in wat We gaan doen, want We zijn nu middenin wat wij willen gaan doen. Is in het datamodel willen we eigenlijk een aantal data. Owners gaan aanwijzen, dus dat zijn die partijen die ik vertelde en die willen ze eigenlijk dat zeg maar datasets In de bron en.

00:06:47 Interviewee

In de zeg maar views willen We gaan definiëren en die willen we eigenlijk allemaal gaan labelen. Dus wil eigenlijk gaan zeggen van je hebt een bepaalde gebruikersgroep die heeft toegang tot die en die en die views in die in die vorm en op die manier controleer je eigenlijk van hoe dat klassiek gebeurt ook in datamarts. Dan wordt er vaak ook gezegd, komt de data mart voor deze personen datamarts, Dat is eigenlijk wat hier ook een beetje gaan doen dat we verschillende groepen gaan Laten subscriben op bepaalde data bronnen.

00:07:17 Interviewee

Maar wat we ook gaan doen en Dat is denk ik wel redelijk innovatief is dat we In de verwerking van de bron naar geaggregeerde data gaan we ook de IP gaan we mee Laten lopen in de lineage. Dus we gaan bijvoorbeeld zeggen van als iets als wat cijfer letterlijk terugkomt In de rapportage, dan willen we dat kunnen waarnemen en Dit is geen statisch iets. Het is heel dynamisch iets, Omdat soms worden gegevens wel gebruikt en soms niet, dus je wil eigenlijk dynamisch kunnen trekken van. Komen er niet letterlijk cijfers van dat en dat bureau op een te laag detailniveau in mijn rapportages of in mijn datasets terecht? Dat wil je eigenlijk kunnen volgen. Om ja, om ook zeg maar te managen dat dat dat dat je niet per ongeluk gegevens stuurt naar een partij die dat niet willen.

00:08:05 Interviewee

Op die manier wordt ook dat IP verhaal wordt afgestemd. Ze hebben wel gebruikersrechten, maar ze hebben niet de rechten om die data door te verkopen en dat gaan ze wel soort van doen. Dus daar moet de hele duidelijke afspraken over gemaakt wordt.

00:08:22 Interviewer

Nou heb ik hier daar opgeschreven. Het aanwijzen van data owners data labelen, Toegang toekennen tot de data en het IP meenemen.

00:08:34 Interviewee

IP trail als het ware een soort van IP trail.

00:08:41 Interviewer

Zou jij voor deze mechanismen aan kunnen geven hoe ze werken en wat voor rol ze spelen bij het mogelijk maken van openheid?

00:08:52 Interviewee

Ja dus bijvoorbeeld tijdens zijn eigenlijk allemaal zijn ze wel bedoeld om door middel van juist door middel van controle en duidelijkheid ervoor te zorgen dat een organisatie comfortabeler is om dingen open te maken. Want wat je dus vaak ziet is dat op het moment als je dit dus niet doet.

00:09:10 Interviewee

Dan ontstaat er gewoon een algehele stuip van we mogen niks delen.

00:09:15 Interviewee

En dat dat, dat zie je, dat zie je In de AVG ook wel eens gebeuren dat dat we dan eigenlijk vanuit onkunde en onwetendheid wordt er eigenlijk gewoon kind met wat badwater weggegooid, dus dan zijn ze gewoon bang om überhaupt iets te gaan delen. Dus al deze die 4 mechanismes die hebben eigenlijk allemaal als doel om dit expliciet te maken, zodat Mensen ook comfortabeler worden om te zeggen van nou, Dit is deze gebruiker valt onder die groep, dit zijn we allemaal goed op orde, dus zijn we, durven we dat openbaar te stellen voor de partij, Dat is eigenlijk het centrale mechanisme waarop dit allemaal werkt is door te controleren en door zichtbaar te maken.

00:09:55 Interviewee

Ja het vertrouwen te geven dat iets iets gedeeld mag worden.

00:10:00 Interviewer

OK duidelijk.

00:10:03 Interviewer

Kun je inzicht geven in het gebruik van API's, Application Programming, interfaces als governance mechanisme om openheid in data platform mogelijk te maken.

00:10:14 Interviewee

Ja, in dit geval gaan we ook werken met APIs, dus bijvoorbeeld.

00:10:19 Interviewee

De de website van hun die gaat een API krijgen naar de backend van dat dataplatform zeg maar en die via die app. Die krijg je een stuk identificatie, dus je hebt je. Je zorgt dat je daar een stuk security of zet dat bij machine to machine interfaces dat dat je dat je kan controleren of nou ben jij ook de persoon bij die groep heb je ook die rechten om via deze APIs te kunnen werken. Hoe wij het oplossen, is dat we eigenlijk voor iedere gebruiker een eigen API set inregelen, maar je zou ook bijvoorbeeld kunnen zeggen, doe maar niet, Maar dat zou wel kunnen dat je gewoon met een API gaat werken en dat je dus bepaalde calls and Request afhankelijk maakt van welke gebruiker je bent en welke calls je mag maken. Wij hebben dat eigenlijk gewoon gesplitst functioneel en dat heeft ook weer een beetje met veiligheid te maken, dus we maken gewoon verschillende poorten, verschillende APIs die je gewoon dynamisch open en dicht kan zetten.

00:11:14 Interviewee

Dat is een beetje een ja, een beetje een keus, dat kan je ook anders doen.

00:11:19 Interviewee

Dus ja API's hebben een gatekeeper rol technische security rol zeg maar om data te ontsluiten. Dat hoeft Natuurlijk niet, maar zeker voor dit soort externe partijen is het gewoon heel fijn Omdat je op die manier ja gewoon een duidelijke deur hebt waardoor dingen naar binnen naar buiten komen.

00:11:40 Interviewer

Ja zijn er nog andere inzichten die je hebt, hoe een API openheid mogelijk maakt buiten dit dit geval om?

00:11:49 Interviewee

Nou, wat ik heel fijn vind aan API's is dat ze ze werken als een soort van dubbele interface, dus als jij als jij een API hebt, dan beheer je het eigenlijk aan twee kanten en je hebt gewoon een afspraak met elkaar waardoor je aan jouw kant intern dingen kan wijzigen zoals je dat wil. Maar zolang de API maar dezelfde functionaliteit blijft behouden, hoeft je de andere andere partij eigenlijk nog geeneens te informeren ervan en andersom. Dus een API is ook een soort van ont koppeling van.

00:12:19 Interviewee

Zodat voorkomen wordt dat dat als jij iets ja de andere partij weet wat hij mag verwachten, dus je kan We kunnen niet opeens dingen omver vallen, dat maakt het makkelijker om continu data te delen Omdat je niet hoeft te iets voor ieder wissel ja met elkaar in gesprek hoeft te gaan. Ik heb het tegenovergestelde meegemaakt met Cegeka. Die hadden dan opeens een nieuw datamodel bij een woningbouwvereniging. En toen vielen opeens rapportages van 6 partijen, inclusief onze datakwaliteit rapportage omver en die niet nodig om even te informeren dat ze een upgrade hadden gedaan.

00:12:53 Interviewee

Dan werk je dus niet met API's, dan zitten dus allemaal dingen direct in je in je database geprikt.

00:12:58 Interviewee

Zelfs met views, maar ja, Dat was ook gewoon. Zij gaven Microsoft weer de schuld en dan valt opeens alles omver en dat vormt dan weer ja, risico's performanceproblemen dat Maar dat helpt allemaal niet om een gevoel te krijgen van controle en zekerheid werkt weer eerder hun gevoel van onzekerheid in de hand.

00:13:17 Interviewee

En dat leidt dan weer tot geslotenheid.

00:13:20 Interviewer

Duidelijk! hoe worden technical development boundary resources zoals software development kits of tools gebruikt als governance mechanisme om openheid in dataplatformen mogelijk te maken?

00:13:39 Interviewee

Even kijken, wat bedoel je precies met die vraag?

00:13:43 Interviewer

Development bounce resources zoals software development kit SDK 's en tools hoe worden die gebruikt om openheid in data platformen te mogelijk te maken?

00:13:56 Interviewee

Heb je het dan bijvoorbeeld over IDA's?

00:14:00 Interviewer

Geen idee wat een IDA's is.

00:14:04 Interviewee

Bij Microsoft heb je dot net Studio kun je dus gewoon code bouwt een sharp of in Als je het echt wil in ja tot net zo ook wel prima en bijvoorbeeld voor voor Java. Heb je bijvoorbeeld clips? Dat is programma's waarmee je dus code bouwt, zeg maar.

00:14:21 Interviewer

Juist en software dus dat zijn de tools en dan daarnaast bijvoorbeeld software Development kits.

00:14:32 Interviewee

Dat zijn, ja sdkb ken ik, Maar dat zijn meer built, mechanismes, bijvoorbeeld dingen als Meven, zeg maar.

00:14:40 Interviewer

Juist en hoe worden deze mechanismen gebruikt om openheid en dataplatformen te verwezenlijken?

00:14:49 Interviewee

Dat is een goeie. Dat gaat waarschijnlijk heel indirect omdat

00:14:57 Interviewee

Je hebt backend code en daarmee bouw je bouw je dingen? Ja, ik zou zeggen de rol beperkt omdat het heel indirect is.

00:15:11 Interviewee

Nou ja je, je bouwt daar Natuurlijk je API's mee en je bouwt daar je bekend mee en je bouwt daar alles mee wat wat wat tegen een database aan praten en wat vervolgens weer dingen aan aan API's gegevens dus.

00:15:23 Interviewee

Dat is de rol, en ja, een goede jou goede backend met goede software en goed gecompileerd zorgt voor zeker ja, robuustheid in je in je systeem staat ook min of meer centraal in in beveiliging en security.

00:15:40 Interviewee

Ja, dat zou ik zeggen.

00:15:43 Interviewee

WatMisschien goed is om te weten, is dat wij dit soort dingen helemaal zelf ontwikkelen, dus wij pakken wij pakken niet zozeer blokken software beet, maar wij gaan zelf in IDE 's gaan bouwen. Dus bij ons is heel veel maatwerk, dus bij ons gaat dat heel indirect, dus zitten daar heel erg zelf op, zeg maar.

00:16:04 Interviewer

Duidelijk! En hoe worden niet technische boundary resources zoals documentatie, training of ondersteuning gebruikt als governance mechanisme om openheid in dataplatformen te verwezenlijken?

00:16:19 Interviewee

Nou ja, wat je ziet is dat.

00:16:23 Interviewee

Wat ik pas meemaak, is dat partijen dat dat dat dat partijen klagen, dat ze bijvoorbeeld een lagere data beschikbaarheid hebben dat ze ergens niet bij kunnen dat ze het niet snappen of niet weten of dat allemaal niet is. En dan ga je feitelijk kijken wat wat Er is en dan kom je erachter dat het allemaal wel is.

00:16:40 Interviewee

Gedocumenteerd vastgelegd, beschikbaar is en wat ik dan wat je dan vaak ziet, is dat er dan niet goed gewoon niet goed gecommuniceerd is of dat er gewoon geen goede informatie is vanuit hier en hier ligt het allemaal en effectief bereik je dan geen openheid Omdat Mensen gewoon simpelweg niet weten dat het bestaat.

00:17:01 Interviewee

Dit soort zachtere dingen. Ja, ja, lijkt me heel logisch dat je dat dat de manieren zijn om Mensen recht wijs te maken, bekend te maken.

00:17:11 Interviewee

Het is pas echt open als Mensen het ook weten te vinden en ze het gebruiken.

00:17:19 Interviewee

Het technische zeg maar daarmee kan je dus de de security regelen. Alle maatregelen die we eerder over hadden en met die niet technische doe je meer het ja zorgen dat mensen het weten te vinden en processen inrichten.

00:17:33 Interviewee

De combinatie van die twee zou ook bij.

00:17:38 Interviewer

Duidelijk. Hoe spelen intellectuele eigendomsrechten, zoals licenties, een rol bij het regelen van openheid in dataplatformen?

00:17:52 Interviewee

Nou IP van de data heb ik het net heel net over gehad. Weet je wel die die IP trail dus? Daarmee voorkom je dat je dingen openbaar maakt die in die vorm openbaar moeten worden, maar liefst maar licenties van software, ja.

00:18:06 Interviewer

Sorry IP In de vorm van Internetprotocol of?

00:18:10 Interviewee

Nee intellectual property dus gewoon eigendomsrechten van data, zeg maar, dus je hebt je. Je koopt een dataset van zo'n partij, zo'n laboratorium en die zegt je mag het gebruiken, maar je mag niet deze cijfers een op een openbaar maken. Nou Dat is belangrijk, want dan ja, dat mag dus niet in die vorm open, maar Misschien wel in een meer geaggregeerde vorm.

00:18:32 Interviewee

Dus dan het tracé van die IP een zorgt ervoor dat je dat je daar controle over hebt.

00:18:39 Interviewee

Maar je hebt als je over licenties hebt, dan heb je dus waarschijnlijk ook over de softwarecomponenten die gebruikt.

00:18:47 Interviewee

Of niet?

00:18:49 Interviewer

Nou of de data zelf over het gebruik van de data.

00:18:54 Interviewee

Nou ja, dus dat is wat ik wat ik net vertelde, is dat voor de data zelf hebt dan, hè, dan doe je dat zo, dan moet je dat gaan labelen om ervoor te zorgen dat je dat je veel kan delen het alternatief is.

00:19:10 Interviewee

Het het meest krampachtige alternatief is dat op het moment dat er ook maar iets van een bepaalde dataset in voorkomt, mag het niet.

00:19:17 Interviewee

Daarmee creëert minder openheid als stadje gestructureerd gaat kijken naar van nou wat, hoe is die data dan verwerkt en wat komt er dan uit? Dan kan je meer openheid bieden, dus Dat is de rol van Van van het trekken daarvan.

00:19:31 Interviewee

Maar ja, ik denk dat het de technologie die je gebruikt zelf niet zo heel erg veel invloed heeft. Je kan je kan je kan een postcast database gebruiken met open licenties of een oracle database. Ik kan nog steeds dezelfde over. Ik denk dat voor de API's wel belangrijk is, maar API's hebben vaak wel een soort van open protocol.

00:19:58 Interviewee

Niet opeens dat je Oracle nodig hebt om bij een Oracle API te komen dat is hoe Oracle het wel zou bedenken en Microsoft misschien ook, wij niet.

00:20:08 Interviewee

Gelukkig de rest van de wereld ook niet.

00:20:10 Interviewer

Nou, kun je inzicht geven hoe het gebruik van beslissingsrechten, zoals algemene voorwaarden, als mechanisme om openheid binnen data platform mogelijk te maken gebruikt wordt?

00:20:22 Interviewee

Ik denk in de basis heel erg beperkt, omdat inkopers hebben de neiging om ook in zo'n stuip te liggen, te zeggen van al jullie inkoopvoorwaarden of algemene voorwaarden verwijzen we nu in zijn geheel naar de prullenbak en we verklaren onze van toepassing en zoek het maar een eind uit. Vooral ook inkopers doen dat dus aan de kant van de klant en wetende dat jij iets wil verkopen, dus jij aardig gaat zijn.

00:20:51 Interviewee

Dus ik denk dat algemene voorwaarden in de vorm zoals ze nu vaak gebruikt worden niet handig zijn, Maar ik denk wel dat ze een kans ligt om daar, zeker als je er meer langdurig samenwerking hebt, om in die voorwaarden een afspraak te maken. Maar ik denk dat je dat dan eerder een aanvullende afspraak doet ter vervanging van algemene voorwaarden dat dat meer zal voorkomen dan dat je je daar in je algemene voorwaarden al rekening mee houdt. En dat is dus vooral omdat.

00:21:22 Interviewee

Kijk, Er zijn als jij die data platform hebt en je biedt het aan, dan zijn de mensen die het afnemen zijn je klanten.

00:21:27 Interviewee

En wat je dan dus krijgt, is dat hun inkoopvoorwaarden gaan prevaleren. Dus jij kan in jouw algemene voorwaarden zeggen van nou, wij staan van alles toe en wij staan al wat dingen niet toe. En vervolgens als je een data klant hebt die het afneemt, die zegt gewoon prullenbak.

00:21:46 Interviewee

Ik denk dat je daartegenaan loopt.

00:21:49 Interviewer

Duidelijk! Kun je inzicht geven in gate Keeping practices zoals toegangsrechten in dataplatformen en de impact op het mogelijk maken van openheid?

00:22:01 Interviewee

Ja nou, Dat is dus één van de belangrijke dingen waar wij dan bij (Client) ook mee werken. Ja, die is Natuurlijk heel belangrijk door te werken met kijk. Je maakt gebruikersgroepen aan met rollen en die rollen die map je over zowel modules waarmee je mag werken. Dus welke stukken software, welke API?

00:22:23 Interviewee

Je mapt het over over data tabellen, dus welk domein welke tabellen wel, welke views wel en niet en je maakt die gebruikersgroepen Misschien zelfs over de data inhoud dus dat je een bepaalde filter toepast van je mag dit in dat geografische domein wel zien en niet zien.

00:22:39 Interviewee

En door dat consequent doen in groepen kun je Natuurlijk. Ja, kun je heel goed controleren wie wat mag zien en juist weer door middel van die controle en dat goed vastleggen, zorgen er weer voor dat je juist comfortabel bent met iemand aan een bepaalde gebruikersgroep te verbinden en hun toegang te geven.

00:22:58 Interviewer

Dus eigenlijk Omdat je juist de controle hebt, kan je meer openheid bieden?

00:23:03 Interviewee

Ja dus wat wij doen is bijvoorbeeld bij de VNG hebben we daar heel ver over nagedacht. Heb je allemaal gemeentes en rollen en de data wordt gelabeld in vertrouwelijk tot en met openlijk. Dat is Misschien een hele goeie trouwens. Helemaal niet over gehad, dus je kan dus data zelf ook een classificatie geven en de gecombineerde classificatie van eigenaar, wie is de data versus hoe heeft hij het gelabeld als bepaalde openheid in combinatie met een gebruikersgroep en welke rechten heeft die? Door dat te combineren met elkaar kan je eigenlijk heel mooi systeem bouwen waarin je precies te zien krijgt wie je wel wat mag zien.

00:23:39 Interviewee

Dan spring ik even naar de VNG. Maar wat we daar bijvoorbeeld hebben is dat bijvoorbeeld een projecten dat Dat is dus heel gevoelig. Sommige dingen moeten openbaar gemaakt worden, sommige dingen juist niet, dus bijvoorbeeld als een Als je een plan hebt om een stuk landbouwgrond te gaan omzetten in een woonwijk, dan mag dat niet openbaar zijn tegen grondspeculatie.

00:23:59 Interviewee

Maar op het moment dat jij dat dus besluit om dat wel te doen, dan moet het weer openbaar gemaakt worden om een level playing field te creëren voor alle bouwpartijen om in te schrijven op ja plannen voor ontwikkeling zeg maar.

00:24:13 Interviewee

Dus daar, daar is die die openheid, die geslotenheid moet heel exact geregeld worden en dat doen we dan dus door al die stukjes data aan een eigenaar te verbinden en die eigenaar bepaalt vervolgens welk level van openheid dat heeft. Dat begint op privé. Alleen ik dan intern uitvoering, dus de uitvoeringsorganisatie dan intern rapportage, dan mag de gemeenteraad het zien, Maar de gemeenteraad wordt al gezien als semi extern. Ik bedoel, in gemeenteraden kan Iedereen zitten en Den Haag zitten er af en toe een paar mafklappers. Dan zet je het op externe overheden, dus dan heb je het gemeente Rijk en dan heb je extern publiek.

00:24:52 Interviewee

Dus dan mag eigenlijk, dan mag Iedereen het zien.

00:24:59 Interviewer

Hoe dragen Pricing Mechanisms bij aan het bevorderen van openheid in data platform ecosystemen?

00:25:11 Interviewee

Ze doen wel wat.

00:25:13 Interviewee

Als je terug naar (Client) gaat, zij willen bepaalde data delen met bijvoorbeeld commerciële partijen, maar ook met onderzoekers.

00:25:21 Interviewee

En wat zij dus doen, is dat ze dezelfde datasets willen ze een ander prijsmechanisme zetten, dus ze zeggen eigenlijk van nou, onze missie is een maatschappelijke, dus maatschappelijke partijen Laten wij minder betalen voor bepaalde datasets als commerciële partijen, Omdat die commerciële belangen hebben. Dus we Laten eigenlijk de markt, Laten we iets meer betalen voor dezelfde data om te Laten bijdragen aan een maatschappelijk doel en juist de partijen die daar iets mee gaan doen wat wij positief vinden juist iets minder, dus universiteit krijgen die data goedkoper dan.

00:25:55 Interviewer

Helemaal duidelijk. Kun je uitleggen hoe revenue sharing models worden geïmplementeerd in data platform ecosysteem om openheid mogelijk te maken?

00:26:07 Interviewee

Ja nou bijvoorbeeld, dit voorbeeld van net, daar zijn we dus naar het kijken om die leveranciers van test data op bijvoorbeeld die laboratoria om hun inderdaad te Laten betrekken in een stukje revenu sharing van wat er mee gedaan wordt en dan tegenover te zetten dat zij akkoord gaan met meer openheid op het gebruik van hun gegevens.

00:26:32 Interviewee

Dat je eigenlijk, zegt Van Nou, je krijgt daar een stukje voor terug. We kopen het niet Alleen in, Maar we geven je ook nog een stukje extra voor de gebruikers. Daar zitten we dan tegenover dat je akkoord bent met een weet je wat Het is met dat soort dingen. Heel De Wereld Draait Altijd om geld, dus op het moment dat er dat daar iets tegenover staat, dan gaan er opeens een hele hoop principiële bezwaren overboord.

00:26:53 Interviewee

Dus Dat is daadwerkelijk wel echt een goed model om Omdat we dus wel. Je moet wel heel erg mee oppassen, Omdat Als je dat verkeerd implementeert, dan kun je hele rare incentives krijgen en hele rare, zeker Als je bijvoorbeeld een maatschappelijke partij bent zoals (Client) dan. Als je dat dus verkeerd doet, dan kom je opeens in de hoek te zitten. Van (Client) die Probeert er gewoon geld aan te verdienen, Als je dat zeg maar out in de media hebt dan ga je wel nat zeg maar.

00:27:15 Interviewer

Dan ben ik iets minder technisch onderlegd dan jij. Maar hoe zou je dit dan moeten implementeren?

00:27:23 Interviewee

Nou ja.

00:27:25 Interviewee

Dat is niet alleen techniek. Techniek is alleen safeguarden dat de afspraken die je maakt goed gaan, zodat zijn hè dat je geen dingen deelt of lekt zeg maar, want effectief wordt het dan lekt die niet zouden moeten.

00:27:39 Interviewee

Dat Safe Guarden, dat moet je. Dat moet je dan ook veel meer vanuit een echt vanuit een governance perspectief doen, is het nauwelijks meer techniek, dan moet je echt gaan nadenken over van. Nou, We gaan nu deze partij dit in het delen. Hoe?

00:27:49 Interviewee

En dat, hoe gaan Mensen daar tegenaan kijken? Wordt dat door middel van een soort kader tegenaan gaan houden? Van ja, past dat nou binnen Wat Mensen acceptabel vinden of niet?

00:28:01 Interviewee

Bijvoorbeeld van (Client) , is bijvoorbeeld dat zij waren heel erg terughoudend hierop, Maar dat ze eigenlijk hebben gezegd van nou, Wij hebben geen winstoogmerk, dus We gaan die prijs ook zodanig neerleggen dat het een een kosten sharing is, dus dat het bedoeld wordt om het zeg maar te accelereren, maar niet te veel geld aan te verdienen. Je wilt een voorbeeld wat waar je dat heel erg fout kan gaan. Ik weet niet ken je pink ribbon nog?

00:28:25 Interviewer

Dat consultancybedrijf of iemand? Nee voor de kanker bedoel je?

00:28:29 Interviewee

Ja pink ribbon verzamelde geld om awareness te creëren voor borstkanker, Maar het was een beetje een soort van soort van financieel perpetuum Mobile. Op een gegeven moment gingen ze ja op gegeven moment waren ze heel erg in beeld. Op een gegeven moment hadden ze doordat ze zoveel bekendheid kregen, kregen ze Alleen maar meer geld om meer bekendheid te creëren om meer geld. Meer bekendheid op een gegeven moment kende heel Nederland ze.

00:28:51 Interviewee

Dus hun missie was Gewoon klaar zeg maar, maar ja, op een gegeven bleven ze maar geld ophalen en dan was de vraag, oké, wat moeten we ermee? En toen begonnen Mensen ook beetje achter de oren te krabben van ja wacht even, maar zijn zij nou daadwerkelijk voor borstkanker iets aan het doen of zijn er gewoon een aantal Mensen nu aan het afromen en geld In de achterzak gaan steken? Dit is zo'n voorbeeld van waar je er heel erg in moet oppassen en dan moet je dus keuzes maken en dat beeld hier ook zo van. Je gaat data verkopen voor maatschappelijk doel, dan moet je heel goed oppassen dat je dat binnen de juiste kaders blijft doen en dat je enige ja winsten die je eruit haalt ook weer juist gaat besteden.

00:29:25 Interviewee

En en Misschien Als het stel je voor dat allemaal heel succesvol is, dan moet je Misschien ook je prijzen Wat omlaag zetten.

00:29:32 Interviewee

Als je maatschappelijk doel hebt.

00:29:37 Interviewer

Kun je uitleggen hoe relationele controle en dan met betrekking tot de identiteit van het ecosysteem, dus hoe je je profileert, bijvoorbeeld hoe dat wordt gebruikt om openheid en dataplatformen te bevorderen?

00:29:54 Interviewer

Dus bij deze zal ik wel een definitie erbij pakken, want hier krijg ik de meeste vragen over.

00:30:24 Interviewer

Relational control refers to the degree to which the platform owner relies on norms and values that are shared with platform users or app developers to influence their behavior what can divided into self control en clan control.

00:30:47 Interviewee

Het gaat hier om een soortement van.

00:30:51 Interviewee

Een soort van je hebt een soort van normen en waarden. Verhaal, wat je?

00:30:56 Interviewee

Op een of andere manier Implementeert in je systeem, zodat gebruikers incentified worden op een bepaalde manier te gedragen, of zo?

00:31:10 Interviewer

Of zich kunnen herkennen in bijvoorbeeld de platform identiteit.

00:31:18 Interviewer

En zou je kunnen uitleggen hoe dat wordt gebruikt om openheid en dataplatvormen te bevorderen?

00:31:32 Interviewee

Die vind ik heel Moeilijk, Ik vind het moeilijk om het relational control aan zich zeg.

00:31:40 Interviewee

Ik, Ik kan het zo snel niet voorstellen. Kijk, je hebt.

00:31:45 Interviewee

Je hebt, je hebt je bepaalde gebruikers, die zet je op een platform en daarbij zeg je eigenlijk van, je hebt een soort van fair use. Dit is hoe je ermee om zou moeten gaan.

00:31:56 Interviewee

Ja, Dat is een beetje soft sturen, zeg maar, dus het zijn soft controls.

00:32:08 Interviewee

Ik denk, Maar dat is een mening hoor, Maar ik denk dat dat wat lastig is, Omdat je dan heel erg uit moet gaan van.

00:32:19 Interviewee

Weet je, weet je, weet je een ding is dat Als je zeg, maar Als je slecht over openheid hebt, kijk als jij bepaalde controles hebt van Van nou bepaalde normen en waarde fair use dan Dat is iets dat je met wat je wel doet Als je je Mensen kent, dus stel je voor dat jij een bepaalde Partij toegang geeft tot je platform, zodat wat klanten van ons, maar dan weet je het, dan ken je de groep en Dat is beperkt. Je weet van nou dat die die Mensen die gedragen zich wel, want Er zijn allerlei checks and balances.

00:32:52 Interviewee

Als je het over openheid hebt, de ultieme openheid is iets gewoon zo tegen het internet aan klappen en je weet precies wat de moraliteit van het internet in z'n Geheel is. Die is allemaal nul. Kijk je had dus dus ja, Als je dan zegt Van ja, doe alsjeblieft dit niet een gedraag je alsjeblieft, zo kan je er donder op zeggen dat het niet gebeurt, dus dat dat hangt heel erg af van je je groep zeg maar, dus Ik denk dat dit dat dit werkt binnen een gesloten gebruikersgroep In de professionele omgeving, waarbij gebruikers ook met name rugnummer kan innen.

00:33:21 Interviewee

Dat je kan zien wat ze gedaan hebben. Ik denk dat het totaal niet werkt in echt echte open systemen, dus draagt het bij aan openheid. Ja, beperkt. Ik denk dat het niet.

00:33:34 Interviewer

Ik kom er straks op terug.

00:33:37 Interviewer

Even kijken. Welke data specifieke mechanismen gebruikt om openheid in dataplatformen te bewerkstelligen, dus specifiek voor dataplatformen en over het delen van data. Welke mechanismen zijn dat?

00:33:52 Interviewee

Welke technische mechanismes worden gebruikt?

00:33:56 Interviewer

Ja, dus welke mechanisme gelden specifiek voor data, dus welke data specifieke mechanismen worden gebruikt om openheid en dataplatformen te creëren?

00:34:09 Interviewee

Ja nou ja, val ik Misschien in herhaling, maar wat ik net al zei, het labelen van data van vertrouwelijkheid, openheid tegenover gebruikers.

00:34:20 Interviewee

Eigenaarschap vastleggen.

00:34:23 Interviewee

De de herkomst dus duidelijkheid geven in data zorgt ervoor dat je weet of je het kan delen, of niet?

00:34:33 Interviewee

Dat is wat ik zo snel kan bedenken, maar dan voel ik een beetje herhaling val nu.

00:34:37 Interviewer

Zijn er wellicht?

00:34:40 Interviewer

Mechanismen die niet gelden voor digitaal platform als Uber op een Facebook of een Alibaba en wel voor data platformen wat openheid mogelijk maakt?

00:35:05 Interviewee

Nou ja, het heet dat.

00:35:07 Interviewee

Nou kijk bij Uber en Facebook.

00:35:11 Interviewee

Is, Het is het gewoon hun data vinden, ze vinden zij.

00:35:17 Interviewee

Waar waar je een mening over mag hebben in het geval van Facebook. Er zijn heel veel dingen waarvan ik iets vind dat het van Facebook zou moeten zijn, maar gewoon van jou.

00:35:33 Interviewee

Wat zou je in een data platform kunnen hebben wat ze niet hebben?

00:35:38 Interviewee

Nou ja.

00:35:39 Interviewee

Ik, waar ik gelijk nu even aan?

00:35:40 Interviewee

Moet denken is, is het privacy aspect dus.

00:35:45 Interviewee

Dat jou gewoon persoonsgegevens delen dat Dat is sowieso moeilijk, dan moet je toestemming voor hebben.

00:35:52 Interviewee

Dus anonimiseren zou al heel erg helpen om dingen te kunnen delen.

00:36:00 Interviewee

Ja, het hangt ook heel erg van het soort gegevens af, hè? Wat je wel en niet kan delen, kijk.

00:36:08 Interviewee

Neem bijvoorbeeld het meest uiterste wat ik ken. Bijvoorbeeld gemeentes die hebben riolerings data waarmee we werken, Dat is openbare data. Dat is net Als de Bach de BRK weet je op percelen in het Kadaster. Die informatie is dusdanig openbaar dat als jij nummertje hebt, dan kan je exact vinden wat Het is. Dat is hele andere heel ander soort data dan elektronische patiëntendossiers, maar een andere uiterste noemt, Maar dat is super beveiligd dus.

00:36:35 Interviewee

Maar dat, dat zit hem allemaal weer in. Dat komt dan weer terug op dat labelen. En dat.

00:36:41 Interviewer

OK.

00:36:41 Interviewer

En dan zet je het delen van data en wat, wat zijn dan Misschien specifieke mechanismen voor het delen van data? Als je dan kijkt naar dataplatformen om openheid te creëren?

00:36:52 Interviewee

Ja, nou ja, je hebt Natuurlijk de de manier waarop dus een API of een UI er tegenaan of gebruiksvriendelijker iets, zodat Mensen dingen kunnen downloaden of Misschien zelfs online kunnen bekijken.

00:37:06 Interviewee

Toegankelijkheid en dat soort dat soort mechanismen. Ja, dat soort technische mechanismen, die helpen Natuurlijk om openheid te creëren, ook voor gebruik Misschien wat minder experts zijn.

00:37:18 Interviewer

Wat zou een mechanisme zijn om toegankelijkheid te creëren?

00:37:24 Interviewee

Nou ja gebruiksvriendelijkheid, gebruiksvriendelijke applicaties.

00:37:29 Interviewee

Als je dus kijkt punt API's zijn Natuurlijk vrij technisch, maar als jij allerlei veelgebruikte platformen ondersteunt. Ik denk als jij als jij nadenkt over bijvoorbeeld gewoon direct met power bi Inklikken en.

00:37:42 Interviewee

En je biedt het ook al gelijk in een vorm aan die je uit hoes fijn vinden. Ja, dan creëer je toegankelijkheid naar mijn openheid. Als jij datascientist hebt, dan kan je data aanbieden in een vorm die je die past. Binnen R of SPSS met de benodigde onderzoeksmethode metadata erbij.

00:38:05 Interviewee

Weet je wel, dus Als je Als je gaat nadenken over de verschillende soorten gebruikers en hoe zijn je data willen gebruiken en je zorgt ervoor dat je het voor die doelgroepen hapklaar aanbiedt In de vorm en op een complexiteit niveau die zij snappen en wij kunnen werken. Ja, dat helpt heel erg voor toegankelijkheid en dus ook bij openheid van data.

00:38:18 Interviewer

Die vorm kan ik dan ook zeggen als standaardisatie je zeggen, je hebt het over een vorm die ze fijn vinden dat standaardiseren of niet per se?

00:38:37 Interviewee

Het kunnen standaarden zijn maar het kunnen ook kunnen, het kan ook. Het kan ook maatwerk zijn, neem bijvoorbeeld ook SPSS. Dat is een heel specifiek programma, maar wordt wel heel veel gebruikt door social Scientists, R is veel meer open. Maar dat is ook meer over experts gebruikers, maar weet je wat? Ja de de gemiddelde social scientist die in Leiden zit zeg maar.

00:39:01 Interviewee

Die weet helemaal niks van IT, maar die moeten iets met de SPSS. Als je het voor elkaar krijgt om iets aan te bieden, zodat zij letterlijk gewoon met SPSS kunnen inprikken op de data. Wat ik weet, dat kan wel creëren je een zekere toegankelijkheid, omdat diegene gelijk wel gewoon met het programma kunnen inprikken op je data.

00:39:20 Interviewee

En dat is dus een standaard ja.

00:39:23 Interviewer

Zijn er ook andere governance mechanismen met betrekking tot de openheid van data platform ecosystemen die we nog niet besproken hebben?

00:39:37 Interviewee

Even kijken hoor, nee denk het niet.

00:39:40 Interviewee

Ik zit gewoon even in de data governance lijstje af te lopen. Van wat voor?

00:39:45 Interviewee

Dingen, heb je eigenaarschap hebben gehad?

00:39:58 Interviewee

Als je het zeg maar.

00:40:00 Interviewee

Als je het wat meer intern bekijkt, dus niet een open naar publiek toe, maar meer bijvoorbeeld naar bepaalde organisaties. Dan is een data definitie belangrijk, dus dat je dat duidelijk is van nou, hoe moet ik die gegevens nou lezen? Wat betekent het?

00:40:18 Interviewee

Maar ook daarbij doelbinding, dus bijvoorbeeld van Waarom. Waarom hebben we deze data nou in deze vorm? Hoe moet je daarmee omgaan? Hoe moet je daar hè? Dat dat dat dat ook duidelijk is, maar ook dat dat je dat je eigenlijk zegt tegenover gebruikers van joh. Deze data is bedoeld om hiervoor te gebruiken, maar ook expliciet niet bedoeld om hiervoor te gebruiken en dat je als gebruiker aan de andere kant dat kan toetsen en dan heb je wat meer over professionele openheid.

00:40:49 Interviewee

Dus dat je dat je een andere kant dat je een afspraak maakt wil dit mag Alleen voor dat doel gebruikt worden.

00:40:55 Interviewee

En en hou je daaraan en dan kom je dan kom je wel weer op die afspraken waar je het over had van het vastleggen van waar is het wel en niet voor bedoeld en zorgen dat dat bekend is en er zorgen dat er ook aan gehouden wordt.

00:41:09 Interviewer

Ja duidelijk we komen tot het eind toch, maar Er zijn nog aanvullende inzichten of perspectieven die je wil delen met betrekking tot de mechanismen die we al besproken hebben?

00:41:22 Interviewee

Kan ik niet zo snel bedenken.

00:41:23 Interviewer

Zijn er ontbrekende elementen die je wilt toevoegen voordat we het interview afsluiten.

00:41:31 Interviewee

Denk het niet, ik heb het gevoel dat de ton redelijk geleegd is.

00:41:35 Interviewer

En dan de laatste vraag, heb je aanbevelingen of suggesties om de openheid van data platform ecosystemen te verbeteren op basis van jouw expertise?

00:41:45 Interviewee

Aanbevelingen heb ik vast altijd. Nou ja, wat ik waar ik zelf een heel erge voorstander van ben is.

00:41:54 Interviewee

Maar Dit is een beetje een soort van oud stokpaardje. We hadden het net over standaarden. Er zijn een hele hoop standaarden en het probleem is vaak dat er dan een meta standaard wordt gemaakt en dan hadden we 100 standaarden en die hebben we 101 standaarden en dan gaat er weer iemand z'n best doen en 102 standaarden en zo zo gaat dat altijd maar door en wat je ziet is dat in de IT gaan dingen zo snel dat op een gegeven moment soort van emergente standaard is. Dus wat HTTP is eigenlijk nooit ontworpen, maar is.

00:42:23 Interviewee

Maar is er gewoon gekomen en dat zijn vaak niet de beste standaarden, maar die komen gewoon toevallig aangewaaid, zeg maar. En met data heb je dat probleem ook dat data standaarden zitten vaak In de Middeleeuwen, dus Als ik een dataset ophaal dan is het zelfs wanneer naar dezelfde gegevens kijk Als het niet heel erg in een bepaald stramien is gevat, dan ga je bij gemeente A en dan bij gemeente B haal je dezelfde data op en kan je gewoon overnieuw beginnen.

00:42:51 Interviewee

En dus ik merk dat data standaardisatie zit echt nog In de Middeleeuwen. Weet je wel, iedere smit heeft zijn eigen manier van zwaarder maken en die der zwaard als anders en waar we eigenlijk naartoe moeten is. We moeten eigenlijk met data moeten we de industriële tijd in, dus er moeten moeten standaard formats komen standaard stramieneen komen en waar Ik denk voor heel erg behoefte aan is.

00:43:12 Interviewee

Een meta standaard waarin standaarden geleverd kunnen worden. Dan hebben we in ieder geval dezelfde taal waarin we standaarden beschrijven en dan kan je in ieder geval zeggen van, Ik ga mijn data delen in ISO nog en nog wat.

00:43:27 Interviewee

En wat ik nu merk is dat Als het over IOS's hebt, is dat de de wereld van de ISO komt heel erg uit compliance en niet uit de techniek en hebben de neiging om alles om alles in procesmatig te willen beschrijven, terwijl wat het waarde denk ik behoefte aan is, is een meer zijn meer technisch gedefinieerde standaarden. Dus als jij Als ik zeg de accountancy is hier trouwens bijzonder ver In de financiële wereld ook.

00:43:48 Interviewee

Die hebben allerlei standaarden. Een mooi voorbeeld is bij gemeentes de IBMO In de jeugdzorg, een initiatief van de VNG.

00:43:56 Interviewee

Die hebben een standaard neergelegd waarmee alle zorgaanbieders alle gemeentes dingen met elkaar delen tussen systemen en dan kunnen ze allemaal andere systemen hebben, maar die kunnen met elkaar praten Omdat zze een standaard hebben.

00:44:09 Interviewee

Dit is dan een monolithische standaard, daar geloof ik ook niet zo in, maar Daarom wil ik ook eens van meta standaard hebben, want dan dan kan je toch wel wat gedistribueerder doen en hoeft je ook niet allemaal altijd via dezelfde standaard te te praten met elkaar. Maar dan heb je in ieder geval een data protocol weet je, je hebt een Internet protocol, HTTP.

00:44:29 Interviewee

Eigenlijk hebben we ook een een meta data protocol nodig, zodat we dus dat we ook gewoon een keer los kunnen van al die irritante lock ins van Oracle en al dat gezeik.

00:44:41 Interviewee

Dan kunnen we ook gewoon Centric gewoon aftanken. Weet je wel dat soort dat soort oude monolieten die gewoon, ja maar hun businessmodel is Dit is vooral niet openmaken, hè? We moeten nu naar een wereld waarin dit wel open is.

00:44:55 Interviewer

Duidelijk. Dit was de laatste vraag, dankjewel.

Transcript Interview 8

5 December 2023

00:00:01 Interviewer

Do you have any questions before we start with the interview?

00:00:04 Interviewee

Nope.

00:00:06 Interviewer

Perfect. What is your job function?

00:00:09 Interviewee

I'm a capability lead for data engineering within Company_int_8 and also partner been within Company_int_8.

00:00:21 Interviewer

How many years of experience do you have with data platforms?

00:00:30 Interviewee

24 years experience in data and I think well, 20 plus years experience with data platforms.

00:00:42 Interviewer

Can you explain what kind of experience you have with data platforms?

00:00:47 Interviewee

Yeah, it's mostly related to collecting data, so creating data platforms for well first collect data and then distribute or provide information to the organisation.

00:01:06 Interviewee

And that was in the early days when I started mostly on premise because cloud was not was not that well developed and later on well especially the last 10 years.

00:01:20 Interviewee

It's more more migrated to the cloud.

00:01:23 Interviewer

Yeah. And what was your role in this?

00:01:27 Interviewee

Yeah, I've done several roles. So first creating actually the data platform, so more on the execution side that was creating the data platform, making sure that data was connected to the data platform then storing the data and then distributing the data, so creating well working actually as an ETL developer or as the data platform engineer and later on I was also more in the design activities.

00:02:04 Interviewee

So creating data platforms, designing it, advising the customer what kind of data platform they should use.

00:02:18 Interviewer

Can you describe a platform of an organisation that you are working for like that you can take for example, who are the data providers? Who is the platform owner? Who are the data users or data consumers?

00:02:35 Interviewee

Yeah. Yeah, so.

00:02:37 Interviewee

What we do for instance, for a large construction company in the Netherlands, we have created modern data stack, where we collect data from all their source systems, but also data from outside the organisation.

00:02:48 Interviewee

Where we collect data from all their source systems, but also data from outside the organisation, so third party.

00:02:58 Interviewee

So the data that is actually not created by them and we combine that data within a central data platform and provide reports on top of it with with well dashboards created in Power BI.

00:03:19 Interviewee

The data platform itself is created with a medallion architecture.

00:03:26 Interviewee

So that means that when we connect the data to the source system, we collect all the data in its raw format, store it in the data lake.

00:03:39 Interviewee

And then yeah, we make sure that there is a sort of a history mechanism on the dead leg itself. And can I be more technical?

00:03:52 Interviewer

No problem.

00:03:53 Interviewee

Yeah. OK. So what we use is.

00:03:56 Interviewee

We use data bricks for that and data bricks is a tooling that works actually within Apache Spark mechanism and what we do is we collect the data, store it in the parquet file, then we use delta lake mechanism to sort of history on that on those files. So with the delta lake mechanism, you're able to actually mimic what a what a SQL database always is using so the asset transactions that stands for automaticity, consistency, isolation, durability and that mechanism is is very strong. So you can.

00:04:45 Interviewee

You can be assured that when you have, when you collect data, then you can you can actually travel in time. On top of that files.

00:04:57 Interviewee

Uh, so it it, it makes sure that you're using very, uh, cheap storage. You can. Uh, you have. Yeah. A very easy way to store your data in large numbers.

00:05:16 Interviewee

In large volumes and then on top of that, we have created a medallion architecture.

00:05:23 Interviewee

So that means that we store the data first, then create a sort of a history on that with the Delta lake and then on top of that we are transforming the data a more business information form and then store it in a in the in the last phase, so that people can report on top of it.

00:05:46 Interviewer

yeah, clear.

00:05:48 Interviewer

Perfect.

00:05:52 Interviewer

Can you give insights in the data providers of the construction company?

00:05:57 Interviewee

Yeah, yeah. So the data providers, you mean the data, the.

00:06:01 Interviewer

The third party data providers so the out outside the company itself.

00:06:10 Interviewee

Yeah. OK. So.

00:06:12 Interviewee

So the data that's not collected by them?

00:06:16 Interviewer

Exactly.

00:06:17 Interviewee

Yeah. So the third party that can be from different sites, so.

00:06:21 Interviewee

What they also use is a lot of data from CBS, where they have some statistics on the environment because they're this construction company, is not only delivering buildings or building actually buildings in the houses and offices, but they also work at the infrastructure of Netherlands, so roads and all kinds of things.

00:06:53 Interviewee

So they must know a little bit about the surroundings. So they they will use CBS data, but they also use data from how do you say it from Tendered. I don't know if you know them so, so so they know what kind of new constructions sites are created and where there's a demand on it and they can write proposals towards it and they use the data also.

00:07:19 Interviewer

Perfect.

00:07:21 Interviewer

OK. And who are the data consumers?

00:07:27 Interviewee

Most of the most of the data consumers are in the in the company itself and they are actually divided in several business units.

00:07:42 Interviewee

So one is residential. So they use the houses and the other one is infrastructure.

00:07:52 Interviewee

And it's also international. So the company is, uh, is actually the same as in the Netherlands, divided in the UK and in Nordics and in in the Netherlands, where in the Netherlands is the head office.

00:08:05 Interviewer

Perfect.

00:08:08 Interviewer

I defined openness as the provision of access to platform resources by, for example, third party entities to an open source or specific interface.

00:08:21 Interviewer

So you you give access to your platform to other parties.

00:08:29 Interviewer

What governance mechanisms have you observed being used in data platforms to facilitate openness?

00:08:39 Interviewee

I think that you must be sure that what kind of users you are you're connecting to your data platform. So you know who is accessing what kind of data.

00:08:55 Interviewee

But you must also be aware of the internal governance. That means that making sure that data that is that is prepared for any external or internal consumption needs to be owned by someone.

00:09:17 Interviewee

So that if there's, uh, if there's a problem with the kind of with the correctness of the data, or somebody asked where how the data is built up that an owner of the data needs to answer those questions.

00:09:34 Interviewer

OK, so assigning roles and responsibilities to the data as well, so classical data governance?

00:09:38 Interviewee

Yeah, classical data governance. But I think that also making sure that the data is always there and always correct. Yeah. And I think that it's best to do that in an disconnected or well, not not disconnected, but I I always say uncoupled, but yeah, it's a little bit of a strange translation.

00:10:07 Interviewee

So it means that APIs are the best way to distribute your data. Because your you can change the data underneath it without breaking the connection to your third parties or the people who you are distributing the data.

00:10:29 Interviewer

Perfect.

00:10:31 Interviewer

OK.

00:10:33 Interviewer

Can you explain how user management or access giving access to the data how this mechanism work to in in relation to openness?

00:10:47 Interviewee

Yeah. So I think that working with APIs you can do that on different levels. You can do that with a sort of a token or with user authentication. Most of the times it's well, anonymous, we always discard that, because that's that. That's not safe.

00:11:05 Interviewee

Because that's not safe to share your data.

00:11:12 Interviewee

So I think that using. O out mechanisms or token access is the best way to make sure that you can distribute your data but without keeping track of all users. So having a very extensive User management from outside.

00:11:40 Interviewee

And this kind of user or this kind of access mechanism makes sure that you can provide the access to the to the outside world but not have the extensive user management.

00:12:02 Interviewer

OK.

00:12:02 Interviewee

Hope that makes a little bit sense. And inside of course there's of course an extensive user management because the data is more accessible, so you need to have more mature access management and most of the times it's done by by internal access management systems like entra the, the, the, the Active Directory from Azure which is which is a good way to use it.

00:12:39 Interviewer

Mostly role based?

00:12:43 Interviewee

Yeah, yeah.

00:12:46 Interviewer

OK.

00:13:01 Interviewer

You touched it already a bit with APIs, out of my preliminary research API's, came as an API's came as an as an platform boundary resource to generate or to enable openness. Can you provide insights into the use of application programming interfaces as a governance mechanism for enabling openness in data platforms?

00:13:31 Interviewer

So how are APIs used to enable openness in data platform ecosystem?

00:13:35 Interviewee

Well, again, referring back to the disconnect. So it means that providing an API to the outside world or to the openers actually, but being, Uh, Flexible enough to to change everything that's underneath it. So actually you, you uh, yeah, you disconnect your system from the way that you want to be exposed in the openness.

00:14:21 Interviewee

But with working with APIs, you make sure that that the the way the data goes into the open is always on the same way is always performant you can be assured that.

00:14:43 Interviewee

Well, again, relating back to the governance, you can be assured that an owner when you connect an owner to the API.

00:14:53 Interviewee

There's some kind of a reference also to the openness of to the outside world.

00:15:01 Interviewee

So I think that APIs are are the best way to distribute your data from a data platform into the open.

00:15:11 Interviewee

Hope that answers a little bit your question.

00:15:14 Interviewer

Answers it perfectly.

00:15:21 Interviewer

How are technical development boundary resources such as sulphur development kits as the case or tools used as a governance mechanism to promote openness in data platform ecosystems?

00:15:42 Interviewee

I don't understand the question.

00:15:48 Interviewer

So how are, for example, software development kits or tools used as a governance mechanism to promote openness in data platforms?

00:15:56 Interviewee

Yeah. So what you can do actually you can very easily with software development kits or with with with packages you can easily connect to governance tools like for instance purview or Collibra, or an Active Directory or entra and being able to to collect the information, the necessary information to apply the correct data governance to your software development. So I think that you don't have to create the whole government yourself, you can rely on the governance tooling that's already available.

00:16:43 Interviewee

And use the SDK's to implement that in your software software application.

00:16:55 Interviewee

So I think I don't think that you have to reinvent the wheel every time you must. It's always good to have a sort of a modular build up of your application and if it's related to an API or.

00:17:14 Interviewee

An end user application by itself always use it. Yeah, use the existing application and connect it via the SDK case as much as possible.

00:17:29 Interviewer

Ok. How are non-technical social boundary resources such as documentation, training or support used as a governance mechanism to promote openness in data platforms?

00:17:39 Interviewee

I think that that, that it's crucial to have at least a good documentation.

00:17:46 Interviewee

I've created a lot of APIs connections myself, but it was more of getting the data from the outside world into the data platform and lacking documentation or how to connect to a specific API It's very frustrating so.

00:18:08 Interviewee

So it helps of course to have a good documentation. Training is not always necessary if you're talking to a developer and he wants to have information from your data platform and he's able to do that through an API or through good documentation, it helps.

00:18:29 Interviewee

And it can. Yeah, it can. Training can be a good way to make the openness and the governance around it more clear, so it supports your documentation that you create on.

00:18:54 Interviewee

On top of your application or API, yeah.

00:19:01 Interviewer

That about platform boundary resources.

00:19:07 Interviewer

How do intellectual property rights such as licencing, play a role in government in governing openness within data platforms?

00:19:16 Interviewee

Yeah. So I'm I'm if you if you talk to to.

00:19:23 Interviewee

My opinion, I think that you should always use as much open source and open source licencing as much as possible, at least from my point of view as a consulting company, because I don't think that well, there there's a there's an extra field or will that that's behind licencing your product because you need to have well you need to have your licencing in place. You need to have a sort of support environment in place that can support your customers and I think that that.

00:20:08 Interviewee

The the things that we are creating by creating openness in your data platform is not nothing new. It's not, not not rocket science. It's not, not things that we have haven't done before so.

00:20:24 Interviewee

If you are uh, making your source uh code or, well, the code that you develop to create the openness, if you provide that to the outside world and of course discard it from all the access rights.

00:20:43 Interviewee

That will that prevent you from creating a whole support, uh organisation to be able to to create your openness if you know if you understand what I mean. So I'm in front, I'm in favour of open sourcing it and and using those kind of licences and that everybody can benefit from it and even improve it if necessary.

00:21:17 Interviewee

Instead of licencing it and keeping it as a closed software.

00:21:22 Interviewer

So if I understand you correctly, intellectual property rights such as licencing does not play, does not enable openness in data platform.

00:21:31 Interviewee

Yeah, correct. Yeah. OK, yeah.

00:21:36 Interviewee

It's good as if you want to sell a product and if you want to sell a A a user ready solution. But I don't think that the data platform, at least from my point of view from a consulting point of view is is a data platform fit for purpose. So for every customer it's it's different.

00:21:58 Interviewee

And it's it's tailor made. So we we cannot, we cannot create that that that kind of mechanism that it needs to be licenced and standardised for every customer.

00:22:12 Interviewer

And from a point of view, if you look from the data platform owner perspective to the data provider, data providers and data data consumers.

00:22:26 Interviewer

Intellectual property rights is not enabling openness in data platforms?

00:22:37 Interviewee

Yeah. Well, if it's an operational phase, then it doesn't matter. Well, I I can. I can imagine that that some kind of software providers are are using that mechanism, but I don't think that.

00:22:56 Interviewee

If you well, at least in the consulting in the consulting world, Umm, it's it's not beneficial.

00:23:05 Interviewer

Could you elaborate on the use of decision rights such as terms and conditions as a mechanism to enable openness in data platforms?

00:23:23 Interviewee

Yeah. Well, I think that we can of course. There's always. Yeah. So if I want to, if I can assure that my customer is having all its intellectual property. So what what we do when when we deliver a data platform and some some APIs or access from the from the outside world.

00:23:51 Interviewee

We always transfer our intellectual property towards the customer and the only firm and conditions that we are the demanding from the customers that they don't copy it 1 to 1 to our competitors. That's the only thing that we ask from them.

00:24:10 Interviewee

That's a little bit of a fair use that we ask. You can use our everything you need. You can copy it for for internal use and for internal improvement, but don't give it away to our competitors.

00:24:28 Interviewer

Can you provide insights of gatekeeping practises such as access rights in data platforms, and the impact on enabling openness?

00:24:41 Interviewee

Yeah, of course. You need to make sure that from a gateway perspective that you always that is always accessible, that it needs to be performed.

00:25:01 Interviewee

So that can be a little a little bit of a a of a burden that.

00:25:08 Interviewee

That especially the performance if if a lot of demand is coming from the outside world into your data platform and your system or your gateway is not performant enough, then you can have have a little bit of problem. So you you always need to have a sort of a monitoring system.

00:25:28 Interviewee

That can signal some non performance things on your on your gateway.

00:25:38 Interviewer

OK.

00:25:45 Interviewer

How do pricing mechanisms contribute to promoting openness in data platforms?

00:25:56 Interviewee

Yeah, it depends also on what kind of things? That the the customer is is well the the the the people the the customer of us asking from us as a as a data platform and the openers towards it. So if they want to have a well performed well documented, user friendly openness to their platform. Yeah, of course. The price goes up. So it depends also on how neat they want to have their openness.

00:26:35 Interviewee

That also depends on on on the price that that's connected to it.

00:26:43 Interviewer

From a platform owner perspective, so how do pricing mechanisms contribute in enabling openness from a platform owner perspective?

00:26:53 Interviewee

Yeah. Yeah, yeah, yeah. OK. Yeah, that can be. I think that they can. They can go about it on different ways. So.

00:27:07 Interviewee

One is that they have a sort of a trusted relation with their with the people who are coming from outside the company and connect to their data platform and they have a sort of a mutual understanding. OK, you can use that data with fair use and without any costs, but they also can monetize their data so.

00:27:32 Interviewee

Making different kind of of data sets.

00:27:36 Interviewee

So one data set that's very simple and they provided they share it with their outside world for for free, but they can also have a sort of a premium data set with more more information towards it and they can, yeah they can ask for money for it.

00:27:58 Interviewee

So there are different ways on on doing that, but it also.

00:28:03 Interviewee

I think it also demands a little bit of a internal system administration system from the data platform owner to be able to facilitate that.

00:28:26 Interviewee

I think there are systems or technologies that can provide with with those kinds of questions. So if you want to have a sort of a freemium or premium subscription model towards access to your to your data platform. Then there are there are technologies that that provided that.

00:28:49 Interviewer

Do you have an example?

00:28:50 Interviewee

Yeah, Snowflake is a good is a good platform that can can be used for data sharing and even being able to ask money or monetize your data sets yeah.

00:29:12 Interviewer

Talking about data monetization, how do you or do you have do you? Can you provide insights in how the data is monetized? For example, for certain data sets and how they come up with the price?

00:29:33 Interviewee

I don't have a have a have a real example for that. I think that that well, as soon as you're sharing your data that's more for instance more sensitive or data that's prepped by AI algorithms that can be data that can be monetized so well, let's say that.

00:30:07 Interviewee

I don't have a real example for this.

00:30:11 Interviewee

Yeah. So you, you, you can have a data set where you where you.

00:30:19 Interviewee

Uh, for instance? Share uh.

00:30:24 Interviewee

No, I don't have an example for this.

00:30:29 Interviewer

Can you explain how revenue sharing models are implemented in data platforms to enable openness?

00:30:37 Interviewee

Revenue sharing models, give an example for this.

00:30:40 Interviewer

So you provide me with data about the the inhabitants of Utrecht and this might be illegal though.

00:30:59 Interviewer

But you provide me with the data and I'm the platform owner. If the data is bought then you get a certain percentage of turnover.

00:31:18 Interviewee

OK, but that's that. That means that the data owner is is paying someone else to for the data.

00:31:27 Interviewer

So if the if the data consumer buys data from the from the data platform.

00:31:33 Interviewee

Yeah, yeah.

00:31:35 Interviewer

How our revenue sharing models implemented to or other revenue sharing models implemented to enable openness in data platforms?

00:31:55 Interviewee

I don't have an example for this. I don't no, I cannot say anything valuable on this.

00:32:06 Interviewee

I'm. I'm just thinking on a on an example on how to.

00:32:14 Interviewee

Well, what? What we are investigating. It's it's not. It's not something that we implemented with our customers, but we we are thinking about it is a way that for instance insurance companies can share data with each other so.

00:32:40 Interviewee

But you can you can imagine that if a client is put on a blacklist and they he wants to apply for another insurance at another company, that he must.

00:33:00 Interviewee

What you can do is actually share that data without even knowing from which insurance company that data is coming from.

00:32:54 Interviewee

The other company must know that this is a blacklisted person and.

00:33:00 Interviewee

What you can do is actually share that data without even knowing from which insurance company that data is coming from.

00:33:08 Interviewee

And by adding new data towards it, you can even well, well, even share the data that is added to it and probably yeah. Come come with come up with a sort of a revenue model.

00:33:28 Interviewee

That you can earn by sharing that data, but more in a collective way.

00:33:39 Interviewer

So you haven't observed revenue sharing models or?

00:33:42 Interviewer

No, no, no, no.

00:33:44 Interviewee

Not not not in in our space.

00:33:48 Interviewer

Then how can you explain how relational control in terms of the ecosystem identity is used to enable openness in data platforms so.

00:34:06 Interviewer

And I define the relational control as the degree to which the platform owner relies on norms and values that is shared with the app developers or the data providers data users to influence the behaviour, and that can be divided into self-control and client control. So it's comes down to the Platform identity of the data platform ecosystem.

00:34:34 Interviewer

Can you explain how this is used to enable openness in data platforms?

00:34:42 Interviewee

The data platform and it's a real difficult question. I I don't know.

00:35:23 Interviewee

Don't think that it's that it can be applied for things that that is that is used in our our business.

00:35:30 Interviewee

Because we are delivering platforms specifically for a customer that we have and that customer will decide if he wants to have openness in their system, and probably it's more of their own of their own clients that will use the data or the third parties that they they connect to it. So I don't think it it's it's for wider public and they that it's created for some kind of an identity, yeah.

00:36:04 Interviewer

What are data specific mechanisms used to manage data platform openness?

00:36:10 Interviewee

Yeah, so the API's I think using technology ready technology like Snowflake is I think a good one, but also push mechanisms like a. But I I don't know if it's openness.

00:36:31 Interviewee

You can have push mechanisms like reporting that's that's pushed towards customers what we.

00:36:38 Interviewee

What we also have developed is a sort of a dashboard that's publicly available for some kind of medicines that are in a clinical phase and people we've developed that for the.

00:36:58 Interviewee

Ministry of public healthcare, it's actually a dashboard where people can check what what clinical phase the medicine is in and and how it's decided by the government to facilitate money in this kind of subsidiary that can be also openness.

00:37:31 Interviewee

Yeah, it's, it's, it's it can be done on different ways, yeah.

00:37:36 Interviewer

But for example, an API is also used in other digital platforms, for example booking.com might have several APIs with clients of them. Do you have more examples of data specific mechanisms?

00:38:04 Interviewee

No, no, no, no. I think no. No, because the the like I said, the customers don't have uh it's not that we have a sort of B2C connection from our customers towards with with our data platforms towards the open world. Mm-hmm. It can be that they they create some kind of APIs, but that's that's more for a specific consumer or third party, yeah.

00:38:44 Interviewer

Is there any governance mechanism related to platform openness that we haven't discussed yet?

00:38:59 Interviewee

Yeah, I think that that if you are talking about privacy.

00:39:09 Interviewee

I think that's that's also a big a big topic. So the GDPR is is I think.

00:39:19 Interviewee

That well, it's it's it's a subject of on its own, but you can also also place it below the the the the governance part.

00:39:31 Interviewee

You must be very careful by what you're what you share in the open and make sure that you're always GDPR proof with with sharing your data. Uh, I think that uh, with the new also looking at the new AI act that's coming, it's also very important to be able to to work in those boundaries. So if you are applying AI to your to your data platform and sharing that data in the openness.

00:40:09 Interviewee

That that, that you must be clear. What kind of algorithms you're using and how this is constructed? What kind of data is used to create to in that algorithm and what kind of data is the output that that is not biased and that it is used on the correct way.

00:40:36 Interviewer

Are there any additional insights or perspectives you'd like to share regarding the mechanisms we have covered?

00:40:42 Interviewee

No, I don't think so.

00:40:45 Interviewee

Yeah. So so I think that that if you're talking about the definition of data platforms that the the the data platforms, you are mostly referring to are bigger data platforms with a B2C connection, I think that if you were looking at Facebook or.

00:41:07 Interviewer

Not necessarily, because those are digital platforms. Yeah. And I referring I'm referring to, for example, a data marketplace or a data space.

00:41:20 Interviewer

That is used for the construction company as well. Yeah, but.

00:41:26 Interviewer

It's I think the depends a bit how the data platform is used. So if you if you use it to monetize data and for the rest you don't have any underlying processes behind it, that's a bit.

00:41:47 Interviewer

It narrows down the scope of the case.

00:41:56 Interviewee

You make distinction between the different kinds of data platforms? So you have a data platform that's really customer owned, which you can have also where they share some data. You can have a data platform that's that's also customer owned, but it's more shared more into the open and you can have a data platform really towards b2c.

00:42:20 Interviewer

I I don't make a distinction in in my thesis between the data platforms, but the the focal point of the platform should be trading data. Wwat you see for other digital platform is that data is a by product.

00:42:46 Interviewer

Are there any missing elements you'd like to add before we conclude this interview?

00:42:51 Interviewee

No, I think that's that. No, it's.

00:42:55 Interviewer

OK. Do you have any recommendations or suggestions to improve the openness of data platforms based on your expertise?

00:43:11 Interviewee

Well, I I think that that's. I think for the for the stability and the consistency, that openness should also rely on a lot of metadata driven functionalities, so there's always a lot of well within the data platform and all the processes that are running to collect the data to process the data it's it's already based on metadata driven development, but it can also be on on the distribution side on the yeah. So on the openness and I think that that it benefits from well stability is is is a good a good result.

00:44:07 Interviewee

That, that's always when you change data underneath it, it's automatically changed in your in your your gateway or your API. And yeah, it creates stability, consistency, etc. So yeah.

00:44:25 Interviewer

OK. Thank you for the insights. That was the end of the interview

Transcript Interview 9

8 December 2023

00:00:43 Interviewer

Do you have any questions before we start the interview?

00:01:03 Interviewee

No, go ahead.

00:01:05 Interviewer

Perfect. What is your job function?

00:01:09 Interviewee

My job function is Data platform lead and so I'm the team lead of the data platform team.

00:01:18 Interviewee

My responsibility is partly product ownership, so gathering and maintaining the backlog for the data platform that we manage.

00:01:30 Interviewee

Partly solution architecture, making designs, thinking how the platform can integrate within the broader landscape and also how specific components function within the broader context of the platform.

00:01:43 Interviewee

And uh, the third part is, uh, still a bit of engineering as well. Uh. Making sure ingestion pipelines run accordingly, but also more the people side of engineering, making sure that everyone is on board with the changes that need to be made or for example when a new data source needs to be onboarded, we need to make sure that the approval from the corresponding data owners is in order for example, so that's. It's quite a broad function. It's also not very narrow, it's not very narrowly defined.

00:02:20 Interviewee

Is because that we've only quite recently built a central data analytics department only since 1st of January in this year. So 2023.

00:02:35 Interviewee

And yeah, we're basically just setting up the central data analytics capability. A lot of initiatives are currently happening decentralised, but we're looking to connect all of those decentralised initiatives and all those decentralised teams by building a central hub to facilitate them.

00:02:55 Interviewee

Well, central infrastructure ways of working standards, that kind of thing.

00:02:59 Interviewer

OK. clear, can you describe the data platform of the organisation you are working for?

00:03:07 Interviewee

Uh, yes. Uh how? How broad or how narrow should the definition be?

00:03:15 Interviewer

Well, what is the role of the platform owner? Who are the data providers? Who are the data consumers that kind of information?

00:03:24 Interviewee

OK, alright. Yeah, so.

00:03:29 Interviewee

Both central. So we built a central data platform from IT. So from a central data analytics department within IT.

00:03:43 Interviewee

What we are currently facilitating on the platform is making sure that the BI use cases and the BI needs of the organisation can be fulfilled. So currently there's no capability for centrally ingested. There's no there wasn't a central data warehouse yet for example. So we've built a

data warehouse component on our data platform to make sure that, yeah, the various business teams first focusing on sort of the usual suspects in BI like the finance department, the HR departments make sure that.

00:04:15 Interviewee

They have the right management insights into the data that we already have.

00:04:23 Interviewee

But on the other hand also creating reports on for example, the performance of powerful or fleet.

00:04:36 Interviewee

Well, there's all kinds of activities, of course that occur on a on a vessel, at sea, every day.

00:04:44 Interviewee

And all those activities are locked in a central database on top of it.

00:04:50 Interviewee

Yeah, which provides a lot of insight in how our vessels or our activities are organised like that. Those are also the types of so more the domain specific types of data sources that we are currently ingesting.

00:05:04 Interviewer

OK. And do external parties make use of the data platform?

00:05:12 Interviewee

No, currently not.

00:05:13 Interviewer

And do you get data from external parties?

00:05:20 Interviewee

Not yet.

00:05:21 Interviewer

OK. But you got them from the vessels and via sensors for example.

00:05:28 Interviewee

Yes. And there are, there are already solutions being built within our organisation which make use of open data or expose data to the rest of the world.

00:05:43 Interviewee

But those were those use cases were not yet facilitated by the data platform.

00:05:47 Interviewer

OK.

00:05:50 Interviewer

How many years of experience, do you have for data platforms?

00:06:02 Interviewee

OK, since August 2019.

00:06:06 Interviewee

So that's.

00:06:09 Interviewee

How much is this? It's currently 2023. That's yeah, almost four years.

00:06:12 Interviewer

Four years.

00:06:15 Interviewee

Yes more than four years.

00:06:19 Interviewer

And can you explain what kind of experience you have with data platforms?

00:06:25 Interviewee

Yes, well the 1st 1 1/2 year I started out by using data platforms, querying data and building reports.

00:06:37 Interviewee

And after that I my experience is mainly in building data platforms and building the pipelines that run on the data platforms.

00:06:50 Interviewee

Mainly focusing on the Yeah, the, more, the, the the clouds infrastructure, the clouds engineering part of data platforms and the ingestion part of data platform. So really the things which you can apply engineering to more of the software engineering part of data platforming as opposed to the more data modelling side of of data warehousing. For example, on a data platform.

00:07:22 Interviewer

OK. And and what governance mechanisms have you observed observed being used in data platforms to facilitate openness?

00:07:33 Interviewee

Uhm, how would you define a a governance mechanism? Or can you give some examples?

00:07:40 Interviewer

For example, an an application programming interface, an API.

00:07:48 Interviewee

Yes. Well, yeah, that's.

00:07:53 Interviewee

Well, an API is is an excellent example in these of what I've encountered and built before in data platforms and otherwise governance mechanisms.

00:08:06 Interviewee

Like a few, I would say a data sharing agreement is also a governance mechanism for facilitating openness and then data sharing agreements in the both in the term of and actual documents, as well as the technical implementation of or the possible technical applications of a data contract.

00:08:31 Interviewer

And how does this facilitate openness? Can you specify this?

00:08:38 Interviewee

Well, if you any to be able to share data, like maybe not openness in the sense of deploying it to a public, your web endpoints and everyone can use it with more openness in the sense of sharing it with external parties. But like business to business.

00:09:02 Interviewee

It's always good to have like in with every interface that you define between two systems. It's always important to define sort of the rules of the game in how data gets transferred between two systems. It helps in getting aligning expectations, but also a little bit of accountability.

00:09:24 Interviewee

What data gets to be made available through that interface. So say I would expose a table containing all of my customers to an external party.

00:09:40 Interviewee

There's some. There's always some quality rules that need to be applied or should be applied in order to guarantee the integrity of the data. So in a data sharing agreement, we could, for example, agree.

00:09:54 Interviewee

That's, I don't know the the user ID column is always unique, for example, and if that is not the case then we are. Then we are breaking our contract and then the other party can hold us who can hold me accountable for making sure that the issue is fixed.

00:10:11 Interviewee

Mm-hmm. Because, yeah, this would. This is just a trivial example, of course. But there could be examples in which there is.

00:10:20 Interviewee

Quite a bit more at stake, especially in business to business communication between information systems.00:10:26 Interviewer

Yeah. And besides, uh data sharing agreements have you observed uh, more mechanisms that are used to facilitate openness?

00:10:40 Interviewee

Yes, good question.

00:10:44 Interviewee

And not not really openness in the sense of opening it up to external parties or the outside world.

00:10:52 Interviewee

I don't have a lot of experience with with facilitating openness in data platforms.

00:10:57 Interviewer

OK. And and to internal parties, maybe maybe.

00:11:02 Interviewee

Well, one example of making sure that data gets shared internally.

00:11:13 Interviewee

OK, let's see.

00:11:19 Interviewee

And it's also more on the data contract side.

00:11:23 Interviewee

OK so but one specific mechanism which I've seen in in the current application or which we're currently applying in for van Oort is?

00:11:31 Interviewee

That a business team is consuming data from the central data platform not by directly querying it, but by copying the data from the from the central data platform to the Azure environment of that business team.

00:11:50 Interviewee

And we supplied them with a standard pipeline to make sure that that Data ingestion adheres to the standards that we define centrally within IT.

00:12:05 Interviewee

And so providing standardised infrastructure is in that way sort of a mechanism to ensure data sharing within the organisation.

00:12:23 Interviewee

OK.

00:12:24 Interviewer

So standardisation.

00:12:27 Interviewee

Yes.

00:12:28 Interviewer

OK, perfect. And.

00:12:32 Interviewer

We discussed it already bit, but can you provide insights in how the use of application programming interfaces as a government mechanism enable openness in data platforms?

00:12:50 Interviewee

Yes, well on an API.

00:12:53 Interviewee

Allows you to make sure that downstream systems can easily be coupled to your to the to the source in the sense that.

00:13:10 Interviewee

Get all maybe it's best explains by.

00:13:15 Interviewee

Seeing or considering the alternative. So for example, a downstream system needs to consume a certain data set from central data platform and they would directly query the database for that.00:13:29 Interviewee

As soon as you make changes in the structure of your database then you have the risk of breaking changes.

00:13:38 Interviewee

If you agree upon an abstracted API.

00:13:43 Interviewee

And that allows the downstream systems to query that sort of abstracted interface which doesn't, which shouldn't break in case of certain changes in the underlying source system.

00:14:00 Interviewee

So in that way it facilitates or it enables. It makes it easier to share data between two systems without tightly coupling those systems. Just yeah, tightly coupling systems is not always desirable.

00:14:18 Interviewer

OK. And how are technical development boundary resources such as software development kits as the case or tools used as a governance mechanism to promote openness in data platforms?

00:14:35 Interviewee

Well the most.

00:14:37 Interviewee

The most used example which I've seen is using open API specifications.

00:14:46 Interviewee

Which was used to. It was used to.

00:14:51 Interviewee

Was formerly known as Swagger.

00:14:54 Interviewee

And then yeah, those agreeing on or formulating your or documenting your API according to the Open API specification allows you to easily host documentation online.

00:15:09 Interviewee

But also allows you to easily generate tests, for example based on yeah, the characteristics of your API.

00:15:19 Interviewer

So in short, how do do the technical development boundary resources facilitate openness if you can Explain how to make Anism work for example?

00:15:35 Interviewee

How it works?

00:15:38 Interviewee

Well, you generate an open API. Yeah. OK. There's multiple. There's multiple ways to implement it, but you can generate an open API specification based on your code. Or you can write an open API specification and you can generate your code based on that specification. And then you can also generate documentation and Generates tests based on that specific specific.

00:16:04 Interviewee

And it also makes sure that your API is defined in a with a certain standard, which increases predictability in downstream usage. So people know generally know what to expect when they need to. When they see an open API specification so they know which endpoints to query, like for example, that's putting a post request.

00:16:23 Interviewee

To the the the customer endpoints, that's that gives a certain error when you don't include the user ID for example.

00:16:35 Interviewer

OK.

00:16:37 Interviewer

And how are non-technical social boundary resources such as, documentation, training or support used as a governance mechanism to promote openness and data platforms?

00:16:56 Interviewee

Yeah. Let's see.

00:17:01 Interviewee

What you have?

00:17:05 Interviewee

While data sharing agreements are an example of this, but do not.

00:17:05 Interviewee

While data sharing agreements are an example of this, but do not.

00:17:11 Interviewee

They're not always. Yeah, the the data sharing agreement is an example. Like you can agree in a data sharing agreements form that you consume certain entities through a certain API with a certain frequency for example. But also you can also have.

00:17:32 Interviewee

Licences associated to it or contracts, but the same legal contracts, purchasing agreements, that's the, that's the term I was looking for because you can also, yeah, but making sure that your API is hidden or secured with a with certain logic that's make sure that's Downstream system can only query your API with 200 calls per minute, for example, because that's the purchasing agreements that they made with you.

00:18:04 Interviewee

It's an. It is a mechanism to facilitate openness, because if you wouldn't have the cost agreement in place then you probably wouldn't host an API because people can flood it with all kinds of requests and you're not going to do that For free.

00:18:22 Interviewer

We're coming to those mechanisms in a in a few in a few questions. But if you look to documentation how you document an API for example, or give training about the data platform, uh. Uh. Is that uh mechanism that enables openness? Do you what? What are your insights about that?

00:18:43 Interviewee

Well, for sure I think, yeah, just the the, the especially training as well as always is always important or it's always a useful tool.

00:18:53 Interviewee

It's not always the most scalable tool, but yeah, it's definitely a mechanism which can enable openness. It's also something which we, yeah, that they mentioned. It's also something that we apply ourselves. Like before, people can use our platform also within the within Company_int_8, they need to have a certain baseline of knowledge.

00:19:14 Interviewee

So that they know what they can expect in using the platform.

00:19:19 Interviewer

And how do intellectual property rights such as licencing play a role in governing openness within data platforms?

00:19:33 Interviewee

Yeah, without a licence. Or at least making, even though it's sort of contradictory statements because you're sort of, you're restricting by restricting your API. You can facilitate an openness.

00:19:46 Interviewee

So by restricting it behind licences, restricting the usage behind licences, that enables, uh, it's sort of a minimum barrier for organisation to exposing an API so I don't know if that's if that's somewhat clear.

00:20:05 Interviewer

Not completely, but if you if you and what about intellectual property rights, for example?

00:20:15 Interviewer

With, when you think about third parties, like if you have external data providers or data consumers, how do in intellectual property rights them play a role in governing openness within data platforms or internal use?

00:20:37 Interviewee

Well, they play a role because intellectual property rights should be of course, enforced. Uh. When facilitating openness as in you can't open up your data platform if you expose your uh, yeah, intellectual property through that API or.

00:20:56 Interviewee

You want to define upfront what is the acceptable amount of intellectual property that you want to expose out of the data platform before opening up your platform.

00:21:07 Interviewer

OK. And could you elaborate on the use of decision rights such as terms and conditions as a mechanism to enable openness in data platforms?

00:21:18 Interviewee

Yes, well.

00:21:23 Interviewee

Personally, I haven't seen any implementations or Yeah, which make use of in terms of conditions to open up data platforms. I'm sure there are use cases, but I'm not sure whether they are very common of the feeling that openness in data platforms and like opening up your platform.

00:21:48 Interviewee

Yeah. Like this agreements or like purchasing agreements or terms and conditions, it's something which a lot of organisations want to do, but it's sort of the low priority. I feel like for a lot of organisations.

00:22:04 Interviewee

But yeah, the so the question was using terms and conditions and facilitating openness?

00:22:11 Interviewer

So yes, so how the use of decision rights such as terms and conditions works mechanism to enable openness in data platforms?

00:22:22 Interviewee

Yeah, that is, I think it's. Well, for many organisations, I think it will be. It will be a requirement to put up those kinds of decisions right or like terms and conditions to before they want to open up their data platform like it should be.

00:22:42 Interviewee

Especially in opening up your data platform, you really want to make sure that the legal agreements that you have or that you want to make some agreements with downstream users of your data.

00:22:53 Interviewee

That they don't use it for the for the wrong purposes like it comes with certain conditions otherwise.

00:23:02 Interviewee

Yeah, people will start to misusing your data platform or like stealing your data and selling it elsewhere just to do the that example.

00:23:14 Interviewer

Perfect. Can you provide insights of gatekeeping practises such as access rights and data platforms and the impact on enabling openness?

00:23:24 Interviewee

I think access rights are a must in facilitating openness of platforms, so it's.

00:23:33 Interviewee

Yeah. Otherwise or you have a fully open data platform with no classified data or whatever on there. Then you don't need the specific access rights. But yeah, access rights should always be implemented based on the classification of the data that you want to expose.

00:23:54 Interviewee

Definitely in the data platform you want to make sure that only the people with the right authorization can access certain entities through the API platform.

00:24:03 Interviewer

And how does this work for Company_int_8, for example?

00:24:07 Interviewee

We use for access control. We use Microsoft Entra ID, so you're formerly known as Azure Active Directory to manage access to our platform based on the role that a certain person has and certain roles are then.

00:24:28 Interviewee

Yeah, certain roles have certain technical permissions, so people from the engineering departments automatically have access to the data belonging to the engineering domain, for example, yeah.

00:24:46 Interviewer

I don't know if this applies for an internal data platform, but how do pricing mechanisms contribute to promoting openness in data platforms?

00:25:00 Interviewee

I think it's a factor which makes it more attractive for companies to introduce openness in their platforms, like if.

00:25:11 Interviewee

If you, if outside parties are able to or, if you as an organisation are able to make money out of your the data that you already own, well, that's it's an easy business case. Uh, usually if you have the proper agreements in place and the proper governance mechanisms such that the data is used properly.

00:25:35 Interviewer

OK.

00:25:36 Interviewer

And can you explain how revenue sharing models are implemented in data platforms to enable openness?

00:25:46 Interviewee

What do you mean by revenue sharing mechanisms in this case?

00:25:49 Interviewer

For example, I as a data provider deliver data sets to you as a platform owner and as a platform owner you have revenue sharing models in place.

00:26:04 Interviewer

That I get a part of of of the of the revenue that you earn with trading the data.

00:26:14 Interviewee

It sounds, uh, it sounds pretty nice and, uh, promising model. But I haven't seen it yet.

00:26:19 Interviewee

OK.

00:26:20 Interviewee

Sure, there are a couple of very mature organisations out there who can.

00:26:27 Interviewee

Yeah, who can implement this successfully, but I haven't heard about it yet.

00:26:32 Interviewer

OK.

00:26:36 Interviewer

And can you explain how the ecosystem identity in terms of relational control is used to foster openness in data platforms. So.

00:26:51 Interviewer

For example, make as a platform owner, making sure that.

00:26:57 Interviewer

Uh. Certain norms and values of the ecosystem identity and how this enables openness in data platforms.

00:27:09 Interviewee

Well, what do you mean by ecosystem identity?

00:27:17 Interviewer

The ecosystem of the data platform, so the different stakeholders at the at the at the platform which comprise the the ecosystem. Wait I will provide you the definition.

00:27:34 Interviewer

So the definition of relational control is, uh.

00:27:38 Interviewer

It refers to the degree to which the platform owner relies on norms and values that are shared with, for example, app developers to influence their behaviour, and that can be divided into self-control and clan control.

00:27:56 Interviewee

OK.

00:27:58 Interviewee

I'm not exactly sure how that's.

00:28:04 Interviewee

How this facilitates openness? I can imagine that if there's, like on a relationship basis. If there's a lot of trust in or in the developers, and like in the certain stakeholders and users of the platform that you are more eager to facilitate openness in your platform. But if people are generally very protective of data, and if the culture is very is very siloed, and if people are very protective of data, then that's automatically will be reflected in your data platform as well.

00:28:38 Interviewer

OK. Thank you very much.

00:28:40 Interviewer

And what are data specific mechanisms used to manage data platform openness?

00:28:49 Interviewee

With the data specific mechanisms.

00:28:59 Interviewee

I would say.

00:29:02 Interviewee

Separating your open data from the storage of the of your more confidential data.

00:29:12 Interviewee

I think that's easy way to, or at least it's the way I see how people implement openness in platform, so always separating the or as on the infrastructure level and the data level separating the aspects which can be shared with external parties, always separating that from the things that should be kept internal to keep risks at a minimum.

00:29:39 Interviewer

OK. And how would?

00:29:43 Interviewee

Well by for example, hosting a separate database for data which is just classified as uh as publicly available.

00:29:54 Interviewer

OK. Are there more data specific mechanisms that you're aware of that, facilitate openness within data data platforms.

00:30:06 Interviewee

No, that come to mind at the moment.

00:30:09 Interviewer

OK, no problem.

00:30:12 Interviewer

Is there any other governance mechanism related to platform openness that we haven't discussed yet?

00:30:22 Interviewee

I don't think so. I think there were a lot of aspects covered.

00:30:28 Interviewee

I think openness in data platforms is still a relatively new topic and it's still being discovered.

00:30:35 Interviewee

All across the industry actually, and people are still experimenting with new ways how you can share data.

00:30:41 Interviewee

Not only on people and process level but also on the technology side of things. So it's uh.

00:30:49 Interviewee

It is very interesting to see where where that's headed.

00:30:54 Interviewer

Are there any additional insights or perspectives you'd like to add about the mechanisms that we already covered?

00:31:05 Interviewee

No, I don't think so.

00:31:07 Interviewer

OK. Are there any missing elements you'd like to add before we conclude the interview?

00:31:15 Interviewee

No. Yeah, that's just that, yeah. Just what I just mentioned like I, I I don't have the a lot of experience in facilitating openness to the outside world in data platforms. And it's also partly because I don't think it's already done a lot like there's still a lot to be discovered in how this works and organisations are very careful in sharing their data with outside worlds because it's one or, or at least maybe when they can earn some money from it. But that also that already requires very high data quality and very high predictability of your platform and most organisations are just not there yet.

00:32:00 Interviewee

And yeah, so interesting to see how when the data industry matures and like when organisations mature in their data analytics capabilities, then it's very interesting to see how the the openness in data platforms will evolve.

00:32:18 Interviewer

Perfect. And then the last question. Do you have any recommendations or suggestions to improve the openness of data platforms based on your expertise?

00:32:34 Interviewee

The I think the the the main recommendation is that basically all, but it's more on the people side.

00:32:44 Interviewee

Like and, but also for us as as data specialists, it's.

00:32:50 Interviewee

It requires quite a culture shift to enable more openness in in data platforms, and it's up to us as data specialists to educate the rest of the world, also on the value of data and also when there's value in sharing data, then also yeah. Educate people on the value of sharing data. Like currently there's a lot of data still hidden behind closed doors or like hidden in in different silos.

00:33:23 Interviewee

Yeah, there's a lot of value to be gained when we start sharing data more and because the especially within the organisations, it's people who are so sometimes so protective of their data as in in sharing it with other departments for example.

00:33:40 Interviewee

Even though there's no particular reason to, and so that's why I think it's. Yeah, it's it's really a culture shift, which is necessary to start sharing data more.

00:33:50 Interviewer

And how would you do this?

00:33:54 Interviewee

Yeah. Well first making people aware of what's or what the value is or why you would share data creates.

00:34:07 Interviewee

Yeah, that then inspire people to to actually do it. But make sure you convince them of what's in it for them.

00:34:15 Interviewee

Yeah. Then you would train them more on how possible ways of sharing data or creating more openness and making sure they're able to do it and that's.

00:34:27 Interviewee

Yeah, that's usually how you bring new change into like these kinds of code shifts work. So it's a lot of change management.

00:34:38 Interviewer

Perfect. Thank you very much. This was the end of the interview.

00:34:41 Interviewee

Alright, hope it was useful.

Transcript Interview 10

11 December 2023

00:00:03 Interviewer

Do you have any questions before we start the interview?

00:00:08 Interviewee

No, not really.

00:00:10 Interviewer

Ok. Let's start. What is your job function?

00:00:13 Interviewee

My job function is head of tech for the data platform and uh, that means to build the platform that is used to build a data platform that is used by Albert Heijn, Etos, and Gall and soon 12 other brands across Ahold Delhaize.

00:00:29 Interviewer

OK and Can you describe the platform of the organization you are working for?

00:00:36 Interviewee

Yeah. So basically when you talk about a platform and so building a platform, of course can mean many things, but a data platform, what we're.

00:00:43 Interviewee

Is basically where you can propagate data through the different layers, right? So the the, the, the medallion architecture. Let's say that's what we have chosen where you can take it from source to a clean layer or a goal layer where you're preparing data products. So how do you get your data from the different sources and make it available for consumption is the trick of the data platform.

00:01:07 Interviewee

And how you do this for not only one company, but let's say 14 or 15 brands is the is the tricky part. But yeah, that's what we do in terms of a platform.

00:01:21 Interviewer

And are there any external data providers besides like? I think it's for the Ahold group, right?

00:01:30 Interviewee

Yes, there are. There are quite a number, but that's more suppliers, pricing.

00:01:37 Interviewee

Uh. weather uh for supply chain.

00:01:40 Interviewee

Well, yes there are.

00:01:41 Interviewer

OK, Ahold or Albert Heijn is the data platform owner I believe, who are the data consumers?

00:01:51 Interviewee

99% are within the organization, but we are building the platform within Albert Heijn. So technically it's from Albert Heijn, but let's say we have different solution zones for anybody who wants to consume it, which means then across, Ahold Delhaize, people consume it.

00:02:10 Interviewer

OK, perfectly clear. How many years of experience do you have with data platforms?

00:02:19 Interviewee

Building let's say building them pfoe maybe 12 or 14 years.

00:02:25 Interviewee

14 or 15 years, yeah.

00:02:26 Interviewer

And can you explain what kind of experience you have with data platforms?

00:02:30 Interviewee

So I've done everything from building it to to migrating it or running a team that's building it. I've been an architect, I've been a solution architect or a product architect as you would call it, because you're building across the thing.

00:02:44 Interviewee

I have been a enterprise, integration and data architect, so that was more on the policies and principles around it. Of course I I didn't start there, but I started as a engineer who was actually doing data warehouse migration or let's say SQL Building SQL Up to that level, and also being a what you call a manager or a director or whatever, now who's maintaining or managing a team that is building it.

00:03:20 Interviewer

What governance mechanisms have you observed being used in data platforms to facilitate openness?

00:03:27 Interviewee

DAMA DMBOK, and when I say DAMA DMBOK, of course I can talk about the data platform and what is there and the hat the full, I don't know. Everything in the circle uh by heart, but uh, the main thing there is dama dmbok that we use and it was not called dama dmbok of course before.

00:03:51 Interviewee

But what is important, there is 2 things. One is onboarding people onto the platform, right? So one thing is make sure everybody's on that platform, which means that at least you have governance or oversight over the data.

00:04:06 Interviewee

And the next thing is then you entice the consumers, so it's producers and consumers. So producer will be able to entice your you know you build a platform that people are able to put data on and once you have the data then consumers also come to you.

00:04:21 Interviewee

So sort of connecting the producers and consumers is what is a initial way of doing things.

00:04:28 Interviewee

But now we do more of the DAMA DMBOK, let's say governance principles and where we go through all of the well, there's of course there's there's a, you know about it, I think. But the the, the, the, the whole data governance how we do it, how we go through everything.

00:04:48 Interviewee

From architecture to including the modeling, security, uh, metadata quality, all of those aspects. Sorry, I'm naming only a few of those things in the circle. What I remember.

00:05:02 Interviewer

I I did the exam, I'm aware of them.

00:05:03 Spreker

Ah, OK yeah.

00:05:05 Interviewee

So you know what I mean? Then you know it better than me, right? So how do you, how do you build a platform? That's the first step. How do you make sure the architectures there, how do you model it? Storage. Mm-hmm. Security. And then I'm losing a track of the other ones integration.

00:05:23 Interviewee

Go through the reference data metadata and then last but last but definitely not least the quality.

00:05:30 Interviewee

So I think I named almost everyone. I don't remember what I missed but OK.

00:05:35 Interviewer

And are there any uh, uh platform boundary resources or as governance mechanism or platform rules that that you can tell this is a governance mechanism that enables openness in data platforms?

00:05:53 Interviewee

Yes, definitely. So everything we build, we build with the smart thing to do is everything you build, you need to build it with some guardrails and that guardrails can be for Treating your consumers As people who don't have time to think about, you know, the platform and it should just work, I should have data available and also for the so ease of use right for onboarding people onto the platform. The second thing is FinOps.

00:06:27 Interviewee

You have to keep track of the FinOps part of things is the FinOps is not just a report and a power BI or a dashboard there, but it's really a cultural thing where you need to make it part of everybody's problem to make sure.

00:06:39 Interviewee

You know. That they are thinking about it, when they're building their solutions, as we say, we say it quite often here don't give free Ferraris when they need to drive to the supermarket.

00:06:53 Interviewee

If you're going to the supermarket, you can take a Volkswagen Polo. You don't need to take a Ferrari. There's no such thing as free Ferrari. But don't don't, don't give a Ferrari when they just want to drive it to the supermarket and back, right? It depends on horses for courses. It depends on what people need. Yeah. And the 3rd thing is good engineering. So what you call as software engineering over time, right? So what Google defines as SRE, or let's say uh reliable thing, is to make sure you have engineering over a period of time. It's not just OK, let me build something here. Let me build something and no, you need to have that proper long term vision in mind before you start building a platform.

00:07:31 Interviewee

Because then everything else becomes easy, including governance.

00:07:36 Interviewer

OK. Thank you very much.

00:07:39 Interviewer

Can you provide insights in the use of application programming interfaces, APIs, as a governance mechanism for enabling openness in data platforms?

00:07:52 Interviewee

There's 2 layers in that question, let me try to explain it in a way.

00:07:57 Interviewee

So when you talk about, you talk about inside the platform, right? OK, let me, let me rephrase that. OK. There are 2 ways to look at it. Uh, let us say you want to use a something like open sync API, one of those standards, right? Let's say it can be any standard, but let's say you choose that one.

00:08:13 Interviewee

Now there are 2 ways of looking at APIs. One is in the in the the the the outside circle which is the incoming and outgoing data, incoming is where people with who have a possibility of sending out data broadcasting data are able to easily onboard data onto the platform. So the the producers are able to just send it.

00:08:36 Interviewee

And and the same thing on the other side we have we have APIs being served where anybody can access it and get you know how many bananas we're selling in the supermarket or something like.

00:08:46 Interviewee

Now that is one usage of open of APIs. The second one is the API fication. If I can call it that or actually it a better word is the composable architecture, how do you build a composable architecture of the platform that each component is able to talk to each other and then you can use open sync API?

00:09:07 Interviewee

Which is the good, which is one of the good Ones where you can try and link between the different platforms and how do I link this to governance if if somebody is building a data Mart, let us say we build a in the part one of the data platform components is a data Mart generator.

00:09:24 Interviewee

Or something like.

00:09:26 Interviewee

Now we we build, you know, the proper GDPR compliance. We build the right observability, we build all these NFR into our platform and then we give this data Mark generator now.

00:09:38 Interviewee

Technically, if anybody can use the generator if they have the right authorization and so on, then they're able to create whatever they want there, right? So then you don't apply the GDPR principles, you're not applying data retention. If there's some, you know.

00:09:53 Interviewee

Uh, for example, if you start storing personal uh PII data in that, then you need to have the right principles in place, right? You can't have that. So the governance where the data governance or data management. In that case, what you need to do is make sure that you have these common components also built in for the any of the components. And that's the composable architecture.

00:10:14 Interviewee

Which means if I was to use the same example, let us say somebody comes and says I want to use your platform to build a Data marts or some reports if I. If I say I want this particular feature, I click on it.

00:10:27 Interviewee

Then other features like GDPR written retention right to be forgotten. All of these things should be pre determined. You can't. You can't allow your consumer to decide. OK, I will just use this because then they can make a big mistake in terms of data governance.

00:10:44 Interviewee

So that's what we do.

00:10:47 Interviewer

OK.

00:10:52 Interviewer

And how our technical development boundary resources such as software development kits or tools used as a governance mechanism to promote openness in data platforms?

00:11:05 Interviewee

We use it to not promote the openness part. What do I mean by that? Security? We shift left on security. We have a dev SEC OPS assessment every 3 months where my teams look at what is their maturity level.

00:11:19 Interviewee

And what was the comparison to the previous time? So this is more an operational thing where we go through all dev SEC OPS kind of questionnaires.

00:11:26 Interviewee

But this tool is actually built on backstage, where we have a software development kit or a software development platform, and SDP which is built by our colleagues here and what they do is they allow for good dev ex experience while at the same time keeping it within control.

00:11:45 Interviewee

Right, not too much of only pleasing the developer. So your question is maybe a bit confusing to me, or I don't see it because I think.

00:11:58 Interviewee

If you want to implement security, then the openness part needs to be compromised or let's say needs to be added to your a usage. So in a in a banking environment or even in retail, you can't have them having access to a customer information like PI data, but also let's say bonus cards or you know something like that customer ID.

00:12:22 Interviewee

Uh, you can't have that. You can't have everybody having access to it from a platform and engineering perspective. But let's say engineers need to use that column or those data fields to build something. So that is the level that we can mask data or, you know whatever you want to do. Like we won't talk about the solution, but the governance part. That's why you need to control it.

00:12:42 Interviewee

Control maybe sounds a bad word. You need to have a that's again a guardrail. You know you you have a guardrail in place so that people can't just go wherever they want.

00:12:52 Interviewer

And and if you look to 3rd party data providers for example.

00:13:02 Interviewer

Does it apply there as well? So are.

00:13:06 Interviewee

No, we have very we have very strict DPIA 's in place. Uh, so we follow the demo. So for example, we don't share data even within the departments Albert Heijn I told already told you Albert Heijn, Etos Gall and we use HR. My platform is also used by the HR team for a data lake also used by financials for data like but none of us have access to it.

00:13:26 Interviewee

So we're very, very strict on that part. If they need access. And I've always said no, but in the past they needed help. They did not know what to do with the data and so on even there we did not have access to the data, but we had access to their to their instance of the platform to help them. So before the data was put and then we were all knocked out of the system, right, we were thrown out, which is perfect, but we needed even then we needed a dpi in place.

00:13:51 Interviewee

Before we, before we went anywhere. So we have an internal security audit, internal security team which checks our procedures and everything in a very strict manner. Every 6 months we have an audit every 6 months. Every we can have an audit every month, every week because we're ready now. 2 years ago it was we were still building all of this. When I started this job

00:14:11 Interviewee

Yeah, but now we have a very good. Uh, strict procedure in place.

00:14:16 Interviewer

OK. Coming to those practices in a bit, but how are non technical social boundary resources such as documentation, training or support use as a governance mechanism to promote openness in data platforms?

00:14:34 Interviewee

A big time, but that's only within a prescribed audience. Why do I say that? For example, all our documentation is on GitHub.

00:14:43 Interviewee

All our discussions by people is not on Teams because everybody can have access to teams, but not everybody has access to GitHub because we allow only people in.

00:14:51 Interviewee

So we use GitHub discussions and we use our GitHub documentation for sorry documentation on GitHub to document everything about the platform.

00:15:04 Interviewee

So really we are controlling who is having access to it. We don't just allow it out.

00:15:09 Interviewee

Then of course you have the global Albert Heijn and Ahold Delhaize principles in place, where if somebody is leaving, they have to be out, out, out, checked out of the system. They have to be thrown out of the system and all of that happens. But from our perspective, we have this checks in place. So so there is no openness but it's more controlled openness, right within that GitHub discussion. If you are a team member who's having a problem with the platform or having some missing data or whatever, you can readily access anybody in my team and raise an incident, open a discussion, start a new thread, call out to people, do everything.

00:15:30 Interviewee

But it's more controlled openness, right within that GitHub discussion. If you are a team member who's having a problem with the platform or having some missing data or whatever, you can readily access anybody in my team and raise an incident, open a discussion, start a new thread, call out to people, do everything.

00:15:49 Interviewee

And plan your meetings or whatever. But we don't just say anybody can have access to the data, right? You can't do that.

00:15:57 Interviewee

We used to use confluence before, but we moved to GitHub specifically about documentation because we wanted to have a little more well governance in place.

00:16:09 Interviewer

And would you agree that? If the governance is good, so you are working with the case with documentation, training, support.

00:16:18 Interviewer

And it's not, uh, or if it's used wisely, it would create openness. So if the governance is good, so you know what you are doing, you know you have the right roles and responsibilities in place. Uh, does this allow for more openness because you have this controlled environment?

00:16:38 Interviewee

Yes, because then it's a safe space. Then it becomes a controlled environment where you can.

00:16:45 Interviewee

Safe, safe space. Basically a safe environment for you to experiment to do things and so on. And also you know that you cannot go accidentally go and delete a solution zone or subscription or something like that. It's very, very well controlled. So yes. So that's very important. It's very easy for me to say it's important, but we've done it and it's takes a long time before you get there.

00:17:04 Interviewee

Right. Uh, But yeah, I mean, the more you see right, data is becoming this asset. People talk about valuable asset for organizations and all of this. And then if you mess up the actual accessing it or of it or user access management of it, then really you're really in a bad space but yeah.

00:17:24 Interviewer

Just a question out of curiosity outside of the interview. You say people talk about data as an asset. Organizations wants to manage data as an asset.

00:17:38 Interviewer

Do you think data should be on the balance sheet?

00:17:42 Interviewee

I'd like to see IT spend on a balance sheet first. Not every company, even big, large, large organizations, the IT spent is somewhere hidden. It's not called out.

00:17:56 Interviewee

Or maybe it's very small compared to billions. Then if you talk about 100 million also, maybe it doesn't make the whole balance sheet. So yes, but before I will be long gone before data is making it, I think making it onto a balance sheet I think so I hope I'm wrong, but yeah.

00:18:15 Interviewer

OK.

00:18:18 Interviewer

How do intellectual property rights, such as licensing, play a role in governing openness within data platforms?

00:18:29 Interviewee

What do you mean by sorry? Can you repeat the question?

00:18:32 Interviewer

How do intellectual property rights, such as licensing, play a role in governing openness within data platforms?

00:18:42 Interviewee

So within our organization, we have an open policy in terms of our GitHub repository, our repository repository.

00:18:53 Interviewee

The product as such is known to only let's say it's inner source we have. We have a lot of people inner sourcing.

00:19:01 Interviewee

I set it up about a year ago.

00:19:04 Interviewee

But that means that it is also open repository, so we don't have any restriction there.

00:19:11 Interviewee

And it doesn't. It exists as a product of component many multiple multiple components. So it's not like I can take a CD in or whatever and install it.

00:19:22 Interviewee

I still have. It's not only building blocks because we also provide the data services around it, like set up an ingestion.

00:19:28 Interviewee

And so on.

00:19:29 Interviewee

Mm-hmm. Umm, but to talk about openness that way, I think we're quite open because we quite allow people to. So if somebody, for example, somebody comes along and says, you know, this feature on this component, this would be it would be better if this feature was this plus, you know something else and I say.

00:19:49 Interviewee

Build it, do it yourself, and then they do it. And then of course, we need to have my team is the one that is responsible for maintaining it. So we have to, if we follow the inner sourcing 4 eye principle where we say OK, anybody can build it, my team approves it and then it becomes part of the feature of the new platform, or the new version depends on what you're building.

00:20:10 Interviewee

So yes, we have a lot of openness in that way. Now with that comes a lot of issues because then if you open up and you don't have the right procedures in place or the right team in place.

00:20:21 Interviewee

You can have one of 2 things. One, we have the team approves something which should not have been approved because they'll Oh my God, I have 100 PIA's to get through. Let me just approve it or or.

00:20:33 Interviewee

Even worse, the teams are building something which is not good, and they keep rejecting it. Right? They say no, it's not OK, not OK. And then.

00:20:39 Interviewee

They're like they're stuck.

00:20:41 Interviewee

So this takes a long time to get to that culture and you know, to get to that inner sourcing culture. But luckily within Albert Heijn, we've done a lot of it.

00:20:50 Interviewee

So we have what is called software improvement groups with horizontals and verticals which talk about architecture, security and then here you're talking about the verticals and then we go. So it's quite well structured, let me say.

00:21:04 Interviewer

OK. Could you elaborate on the use of decision, such as terms and conditions as a mechanism to enable openness and data platforms?

00:21:16 Interviewee

We don't. So I got it from the previous question as well, but I'm understanding now more by going we don't.

00:21:24 Interviewee

We cross charge a bit of the usage and the time.

00:21:28 Interviewee

But we don't charge for the product because our consumers are within within our colleagues. So if I was running a company which was building this platform as a whole one, then yes, I would have a whole, let's say.

00:21:44 Interviewee

Setup around this.

00:21:46 Interviewee

Uh, but no, we don't have a.

00:21:50 Interviewee

We don't have that.

00:21:53 Interviewee

We have product mindset, but we don't charge for it. Maybe we should, but we don't yet so.

00:21:58 Interviewer

So back to the the previous question then maybe from previous experience or that you are aware of do intellectual property rights such as licensing play a role in governing openness within data?

00:22:14 Interviewee

Yes, yes, 100%.

00:22:14 Interviewer

Platforms is OK.

00:22:18 Interviewee

In one of the previous examples I I was talking about earlier.

00:22:23 Interviewee

The problem with platform, how the situation with the platforms is who owns the platform, who owns the data. Mm-hmm. With data is very clear. But the platform it's not, especially if you're buying a product. If you're buying a solution that is customized or you're buying a fully customizable solution, then it belongs to the party who's paying the bill.

00:22:44 Interviewee

Well, but if you're buying, well snowflake as well is also when you talk about Snowflake for example, we went, we were tested in the past before we started to build this current one.

00:22:55 Interviewee

There the data is very clear. Who owns it? But who owns the actual, let's say the snowflake bit right the the the platform is also very clear because we're buying a product.

00:23:06 Interviewee

But if you go to a smaller company or somebody else who's building a specific customized platform and so on.

00:23:12 Interviewee

Uh, let us say building a paved road data platform, right? Just for onboarding and uh, consuming data. Then that becomes a discussion. But yes, definitely it makes a difference.

00:23:22 Interviewer

OK, for the upcoming questions, you could could as well take other examples from not only Albert Heijn, but from previous jobs of things that you're aware of. Can you provide insights of gatekeeping practices such as access rights in data platforms and their impact on enabling openness?

00:23:47 Interviewee

Access rights to data is very, very much in place. And what do I mean by in place? That's the whole data governance spot, right. So how do you make sure that data you're talking about it as a valuable asset?

00:24:07 Interviewee

We just spoke about it earlier. You're saying it's a valuable asset. You're saying that you can make insights from it. You're saying that you can make.

00:24:17 Interviewee

Use it to, you know, make data driven decisions and you know, make millions or billions or whatever.

00:24:22 Interviewee

But that means that there's valuable there is a valuable it's a valuable asset, right? That basically means that the core of it is a valuable asset. So you can't just make it accessible to anybody.

00:24:35 Interviewee

So how you make how you expose that data is very important. Now a platform engineer has no need to get access to that data.

00:24:46 Interviewee

Who gets access to their data? Maybe somebody like a data steward.

00:24:50 Interviewee

A data steward or a analyst, or a power BI? Sorry, somebody is building a report or somebody from the business, right, who's really looking at the data. These are the ones that need access, right? So at every level, you need to have molecular column based access to data.

00:25:13 Interviewee

And you can have many tools. So for example we use data bricks as Unity catalog. Yeah for for building it. So that is something that we do.

00:25:24 Interviewee

Uh, but uh, let us say fabric is coming out with a great resolution called one security coming out soon. Or Microsoft is coming up with a new solution called One Security, which really allows you a lot of flexibility and access to data, but we use data bricks.

00:25:46 Interviewee

We use data bricks as Unity catalog which also gives us a granular level because let us say you're rolling out the solution to Gall. Now inside Gall there might be only specific people who should be have access to certain amounts of data and that's what you can control on an AD level, preferably on a group level, but definitely also possible on an individual level.

00:26:07 Spreker

OK. So yes.

00:26:08 Interviewer

Is this yes different than role based access?

00:26:12 Interviewee

It's pretty much role based access. Uh, that's what I mean. Sorry by AD, but when I talk as as your uh thing. But when I talk about it is basically role based access.

00:26:24 Interviewee

There's the, but it could. It may not be an individual access, it could be a group level, but doesn't matter, it's role based.

00:26:31 Interviewer

OK. How do pricing mechanisms contribute to promoting openness in data platforms?

00:26:41 Interviewee

When you say pricing mechanisms, what do you mean?

00:26:44 Interviewer

For example, I am as a data provider, give you access to a data set to sell, and you as or data platform owner, sells the data set to a data consumer?

00:27:12 Interviewee

So so there are multiple, let's say pricing strategies that you can have, right? So, so the usual ones is based on the product or let's say based on our economy or penetration or usage or you know maybe sometimes it's premium you say I give you access to customers. We don't, we're not in the business of doing that, but let's say something like that.

00:27:38 Interviewee

Uh, but the point is that you need to have this synchronicity or mutually beneficial pricing mechanism. And what do I mean by that? Let us say that you are building a a data product. Now let's talk about data products for a minute.

00:27:58 Interviewee

It's not only data, but you need to call it as data products. Let's say for example, if I'm selling information from a telecom to a retailer, some retail to telecom, let's talk about that.

00:28:11 Interviewee

Uh, let's say I want to know everybody in this area who shops at uh Jumbo or Albert Heijn and also has a KPN subscription because I want to target those customers, right? Something like that. So there is a lot of value in potential value in that data, right?

00:28:31 Interviewee

And the value is only generated afterwards, it's only reflected after. It's, not like. OK, I have this data, So what now I'm going to use it and then I'm going to, you know, do that. So this it's a 2 way thing. Can I price it before I use it or do I use it, you know that kind of a thing, right? So there are many, many mechanisms, but I think it needs to be mutually beneficial.

00:28:56 Interviewee

Uh, so that's why I think it's very difficult to set a price like I can say €100 for, you know each person and then you use that and make millions or you know, vice versa, I might say 100,000 and you use it and you get nothing you know.

00:29:13 Interviewee

So I think it should be beneficial. That's my idea and but I still think it should be on a product level.

00:29:21 Interviewee

OK. Why? Because the product is based on the producer data, so the quality comes from there based on the platform because you are passing it through the different layers and making it nice and shiny like gold and on the consumer who's actually using it to do something.

00:29:40 Interviewee

Right. So that's why I think it should be across the chain, let's say.

00:29:44 Interviewer

OK, so it should be, uh, based on the product level, but also on the use of the data consumed.

00:29:52 Interviewee

Yeah. End to end. Let's put it like that.

00:29:56 Interviewee

That's my, that's my view, of course.

00:30:00 Interviewer

Can you explain how revenue sharing models are implemented in data platforms to enable openness?

00:30:11 Interviewee

I would say exactly this. So revenue sharing again would be data product level.

00:30:18 Interviewee

So for me it is a.

00:30:24 Interviewee

And this is where the whole AI and all of this also comes in, right. Do you make your decisions based on millions of data rows of data or do you make your decisions made based on let's say prescribed data sets.

00:30:40 Interviewee

Or do you make a decision based on prescribed data products that are available for you, right. Uh data set might be OK. I'll just take and say how many customers are going to the supermarket every Saturday morning.

00:30:52 Interviewee

But the data product for me is how many people are going there for a for a uh, for in the Saturday morning and buying croissants or buying, you know, one specific thing, which means that if they buy croissants maybe they were and then that's that now one level higher is if they buy croissants maybe they also and that's where machine learning OPS, frame ML OPS frameworks come in. OK they're buying. So they'll be buying eggs and they'll be buying oranges as well, or orange juice as well.

00:31:17 Interviewee

So that's the let's say the the, the, the, the 4 layer or 3 to 4 layers that comes up there. So which means for me still it needs to be an end to end thing. It can't be just in isolation. That's what I'm trying to say.

00:31:31 Interviewer

OK.

00:31:32 Interviewee

So it's a transaction entity. The entity is a transaction.

00:31:37 Interviewee

And the the value of the model or let's say the value of the evaluation or whatever is based on the properties of that whole transaction which means which is there in the product at the end the data product. Sorry at the end.

00:31:53 Interviewer

Perfect.

00:31:56 Interviewer

Can you explain how relational control is utilized to foster openness in data platforms?

00:32:04 Interviewee

And what do you what do you mean?

00:32:06 Interviewer

I shall provide you with the definition of relational control relational control refer refers to

00:32:14 Interviewer

The degree to which the platform owner relies on norms and values that are shared with, for example, app developers or other 3rd party entities to influence their behavior, and that can be divided into self control and clan control.

00:32:42 Interviewer

Can you explain how relational control is utilized to foster openness in data platforms? And I'll refer to relational control as the degree to which the platform owner relies on norms and values that they shared with app developers to influence their behavior, and that can be divided into self control and client client control.

00:33:03 Interviewee

Yes, it's a big it's a big.

00:33:04 Interviewee

It's a big one.

00:33:06 Interviewee

If I understood what you're saying right.

00:33:09 Interviewee

Because I understood something else in terms of relational before.

00:33:13 Interviewee

Now let me give you an example.

00:33:19 Interviewee

We'll talk about bad quality of data.

00:33:24 Interviewee

And you remember one thing, data is always not you. You might know the data is not always, uh, good quality. So we talk about bad quality. Let's take an example where you have customer ID and something like CUSID and you have another one where it is CID did and another one which is C underscore ID. You know whatever.

00:33:46 Interviewee

So the deduplication of data and making sure it's going through the different layers to get to the silver or the gold layer is very important.00:33:54 Interviewee

You need to you need to structure it properly.

00:33:57 Interviewee

But at the same time, you know that CID and all of these things belong in like, a a privacy locker or a PII data locker or data locker or whatever.

00:34:05 Interviewee

So you start, you start dumping it also in the lock, because you're trying now to uh, you're trying now to uh, not allow access to people to get their ID 's or bonus cards or whatever.

00:34:18 Interviewee

Now, if you don't have the right governance mechanisms in place in terms of checks and balances, what will happen is that the people who are using the data now, if they have access to it, people will use it.

00:34:34 Interviewee

Which is, which is very silly, might sound silly, but it's human nature. If I have access to data, I'm going to use it and that may not be the best use of it. Now, do you blame the person who's using it? Or do you blame the platform for allowing it to give access?

00:34:48 Interviewee

I would blame the person using it and the person using it will say no. It's your fault Pritam because you have made me. It was available so I used it. Of course they should think before doing that but.

00:34:57 Interviewee

That's another discussion.

00:34:59 Interviewee

Now they can say oops, I saw it there and it was part of my ML OPS framework and it started to be used and then it was gone.

00:35:06 Interviewee

We used to find things like this, right? We I've been in another client before where we could find the the accidentally. Many things were exposed.

00:35:15 Interviewee

So exposed in terms of address or, you know GDPR compliance. And so luckily that client was not in the not in the EU, but still if they had a EU client and that will not, you know, they cannot work with them or EU application. So it is very important to have that mechanism in place to make sure.

00:35:35 Interviewee

That between the developer and the people, so engineer and user, let's say the user is a data analyst or data steward these 2.

00:35:44 Interviewee

They are linked by the data, so it's very important to have the mechanism in place, yes.

00:35:50 Interviewer

OK. Thank you very much. And now to data specific mechanisms, mechanisms, what are data specific mechanisms used to manage uh, openness in data platforms?

00:36:05 Interviewee

So we use the medallion architecture. We use the we use the principle, but mainly we use the in all my experience we always use the level of medallion. Why? Because it is easily. The governance is good. What do I mean by that?

00:36:24 Interviewee

If you walk into a, if you walk into a the raw layer or the let's say the bronze layer.

00:36:32 Interviewee

Almost nobody should have access to that layer, so we can control the access. The silver layer can be. This is normal. This is a high level, uh classification. The silver layer can be used within the domain, let's say within an organization or a department or something.

00:36:49 Interviewee

The gold layer is used across the domains. Let's say you have a supply chain domain and you have a. So the supply chain domain builds there things were on the silver layer, but when they want to build something to expose to the outside world, they put it on the gold layer, so these are the kind of mechanisms that we have and basically it's based on the medallion architecture and using good principles from Dama DMBOK of course.

00:37:15 Interviewer

OK. Are there any other data specific mechanisms used to enable openness in data platforms that you're aware of?

00:37:25 Interviewee

I think I think when you talk about data specific mechanisms, I mean, no, no, not really.

00:37:39 Interviewee

I think it's a I I'm trying to remember whether it was a urgent mechanism or something like that. Right. There's one. There's one like that.

00:37:52 Interviewee

No, I I'm not sure I.

00:37:53 Interviewee

Think there's there are a few like that around.

00:37:58 Interviewee

Missing urgent. There are a few classifications like that I'm aware of.

00:38:04 Interviewee

Where you can say that something is missing.

00:38:08 Interviewee

Something is missing random missing. Not either. There are a few. Sorry I don't remember all of it, but let's say there are a few types of mechanisms that that are there. But no, we we I'm not fully aware of it, no.

00:38:26 Interviewer

OK. And if I phrase the question differently, what applies for a data platform as a governance mechanism to enable openness, hat does not apply for a digital platform like booking Uber, Facebook, Alibaba or another digital platform?

00:38:46 Interviewee

What applies common is of course, apart from general good engineering is GDPR data.

00:38:56 Interviewee

Privacy data lockers GDPR compliancy data retention right to be forgotten all of these things right? This is very important what does not apply is maybe how you might use the data because as a platform and a digital platform there are different.

00:39:14 Interviewee

As a data platform, I don't care. I want every data I I cannot. I cannot say this is important. That's not because it's not my job. I have to just pass it through to the people consuming and maybe the platform sitting on top of it might use it for some reason and I don't know what the reason might be.

00:39:30 Interviewee

And it my reason might change in the future, right? So I cannot say this is important. I cannot classify data. I can classify private data and so on. But I can't classify how that is and and that's for me the main difference. So it's about importance of data. Let's say in a platform versus a in a digital platform.

00:39:56 Interviewer

OK.

00:39:58 Interviewer

We're going towards the end of the interview. Is there any other government mechanism related to platform openness that we haven't discussed yet?

00:40:06 Interviewee

No, I don't think so.

00:40:09 Interviewer

OK. Are there any other, uh, other any additional insights or perspective you like to share regarding the mechanisms that we have covered?

00:40:23 Interviewee

No, I think you ask all the questions. Maybe availability is one aspect of data governance that we have not touched upon and that means to always make data available, whether you're a producer or a consumer, right? Making sure they're on that. We spoke about it briefly in the beginning where I said, you know, you need to adopt them to get them on the platform and then consumers will come.

00:40:45 Interviewee

Uh, and probably integrity, but I think we covered that part of data governance where we said that when you move through the different layers, you have to make sure the deduplication efforts or whatever doesn't mess with the integrity of the data.

00:41:03 Interviewee

And security mechanisms, security mechanisms of both the the infra or the cloud platform on which it's which it is running and let's say the platform itself in terms of user access management.

00:41:17 Interviewee

So, but yeah, I think those were the 2 or 3 availability, integrity and security are the 3 things that I can say are probably what we touched upon, but not really touched upon, right, that's maybe.

00:41:30 Interviewer

OK. Are there any missing elements you would like to add before we conclude the interview?

00:41:41 Interviewee

No, I think we discussed everything, I mean, for me, uh, for me, governance in a data platform and talk about governance.

00:41:49 Interviewee

It's uh.

00:41:52 Interviewee

It's 4 things right for me. It's 4 main ones and I'll start with uh quality.

00:41:59 Interviewee

Then there is uh uh protection, right or security?

00:42:07 Interviewee

Compliance. I think security and compliance right always that is a that is a painful one that no everybody has to are there too, right security and compliance.

00:42:15 Interviewee

Data management. So how do you move the data through? How do you actually do it?

00:42:20 Interviewee

Do you and the last and maybe not the last I would say there's one more, but let's say there's stewardship, yeah.

00:42:29 Interviewee

That is a good one, and maybe the last one is good engineering practices like. I mean that's a more generic thing. It's not only for data platforms, but if you build, if you have good engineers who are thinking, you know that kind of a thing, then that's always it comes down to that in terms of.

00:42:48 Interviewee

But for me when I talk about data governance, these are the 4 or 5 things or 5 things that.

00:42:53 Spreker

Talk about.

00:42:54 Interviewer

OK. And then the last question, do you have any recommendations or suggestions to improve the openness of data platforms based on your expertise?

00:43:05 Interviewee

Hmm, I need another few hours for that. No, I think apart from what we spoke about, what I spoke about was a lot of learning from experiences where we have either I've done it myself and, you know, tried and failed and then succeeded, or I've seen it happen. I think one aspect which people forget is if you just build a platform, it does not mean that you are a data driven organization.

00:43:33 Interviewee

What is very important to know is that getting to that data driven culture and making data decisions, you know this usual data driven organization, kind of discussions about you here, right. I want to become a data driven organization, people think building a platform is something that OK immediately will become data driven, no.

00:43:53 Interviewee

You need to build a platform. Then you need to have everybody change. There's a there's a big change management aspect and a cultural aspect, so there's change management in terms of, OK, somebody needs to pick up a stick or a A or A or some a bag of sweets or some stroke. Waffle and say hey, please come on my platform. Here you go please.

00:44:14 Interviewee

Or you stick and then hit everybody with it, right? So it depends on what you use. But let's say getting people onboarded on the platform is just the first step after that, how do they use the data? Is the data good? Is it good quality? Is it, is it available? Is it, is it going through all the necessary security and compliance?

00:44:34 Interviewee

All those points I just said that is more important. So the platform itself is the first step.

00:44:41 Interviewee

Actually, I would say it's a prerequisite. It's not even the first step. It's a prerequisite before you start becoming rolling out good governance principles, the platform has to work. I'm a tech guy.

00:44:53 Interviewee

But I'm I'm more a data guy and I say the tech has to work. You will take it for granted. That's it. There's no question about it. It has to work. Beyond that, what you put on top of it and how you build, uh, you know, the people coming on board, getting to data governance, getting to data stewardship and making sure everybody's on board, that is comes.

00:45:14 Interviewee

That is the biggest part of the challenge, because sure, anybody can any good engineer can build a good platform, but the job is only beginning at that point.

00:45:24 Interviewee

So it's very important for that. Uh, for when you talk about ecosystems using data or you talk about data governance across organizations.

00:45:34 Interviewee

The platform is just a prerequisite.

00:45:36 Interviewer

I see this as well in the organization that I'm working for at the moment for the client. What is in your opinion key in this cultural and change management shift?

00:45:53 Interviewee

People. And it's easy to say people. So when I if you talk to so one thing is management.

00:46:02 Interviewee

Buy in or let's say any kind of data governance initiative. Like, you know, if you're trying to make a company data driven organization or something, you need to have the buy in from the board level.

00:46:14 Interviewee

Mm-hmm. Because priorities will defer. All the people will say it's not in my priority. Now. I know. I I know you built a great platform, but I don't have time for it.

00:46:25 Interviewee

I heard a 1,000,000 excuses, right? But the point is somebody says, guys, we're moving towards this. So let's go. So make it part of your data journey or your journey.

00:46:35 Interviewee

Then people will move. That's one. The second thing is some people might just say ohh damn it, my boss asked me to move so I need to do this and move, no.

00:46:46 Interviewee

That's where you need to work with people like that, right? So first thing is a prerequisite. Again, the board has to say yes. We need to move and then. But the second the the the then comes the people part. So somebody says I have to use this new platform. They're building something. They say it's the best thing ever. I need to move. OK. What do you get from it? And that's what you have to explain to them.

00:47:07 Interviewee

Guys by using our platform. It's a very low tech threshold. For example, it's a paved road data platform. All you need to do is press this button and your data is onboarded here. Ohh very nice. That's all I used to spend hours and doing this and now one battle very nice and then the consumer you tell the consumer.

00:47:26 Interviewee

You know, I have all your data that you need. I have it here so you don't have to write a query to SQL query to go between 56 places or create CSV and create data and reports. Everything is here available for you. Is it? Yes. Yeah. One place select staff from tap. Boom. It's.

00:47:43 Interviewee

So this kind of small low hanging fruit or let's say a carrot on the stick is what you need to get. That is the key, right? So the key there is getting that first of all making data available and then getting people to onboard on your platform adoption of the platform because like I said, platform should be ready. That's a prerequisite.

00:48:04 Interviewee

After that, the game begins right the the, the, the, the, the, the, the negotiation game or the strategy game or or or monopoly or whatever you wanna call it, right, you're trying to. You're trying to get real estate or trying to get people there and, you know, playing chess, 3 moves ahead, whatever game it is, the point is that you use all the games maybe to get people on the platform.

00:48:25 Interviewee

Why? Why do I say that? It's not cheating people to get there? No, because after they get on the platform, they realize. Ohh, I can make these insights. Ohh, I can do that. So platform has to be ready in a good way.

00:48:37 Interviewee

Then the data governance principles have to be there already there as a high level from a company perspective with the buy in from the board and then if you start getting people on it, you really start to see the value.

00:48:48 Interviewee

And of course I can say platform should be ready. Doesn't mean it has to be 100% ready. You need to have enough for the people are able to onboard the data and consume from it and you don't break any rules in terms of your processing of data or you know metal and architecture, whatever. So that's the aspect that you need to go from.

00:49:07 Interviewer

OK, very clear. This was the end of the interview.

Appendix 8: Data extraction and analysis

Analysis interview 1

Interview number	Int_1
What is your job function?	Founder and CEO
Can you describe the platform of the organization you are working for?	We sell financial data feeds covering a variety of asset classes ranging from stocks to options to ETFs, mutual funds and a variety of other data like analyst estimates and ESG data, primarily through an API and a web socket
How many years of experience do you have with data platforms?	11 years
Can you explain what kind of experience you have with data platforms?	I have a significant amount of experience with exchange data market data, the flow of real time you know securities information through exchanges and other government regulatory bodies that control that flow of pricing, public market data. I have a lot of experience with data marketplaces.

Question	Transcript	Code	Theme
What governance mechanisms have you observed being used in data platforms to facilitate openness?	XBRL, this is a programming language that's built off of XML and it stands for extensible Business reporting language. Uhm and it is a way to take data that the public needs and is entitled to, but is in a format that historically was filed in kind of PDF form which is not structured, not digital, not easy to access, and so the XBRL mechanism and language. It is a standard that was built to make data more accessible.	XBRL, this is a programming language that's built off of XML and it stands for extensible Business reporting language.	Standardization
How do these mechanisms work, and what specific roles do they play in enabling openness?	Data that that otherwise would be at a PDF and is very are hard and expensive to get access to in a systematic way. So historically a good example of the types of data that's filed this way is just financial statement data, public company financials in the United States. In 2006 the US SEC mandated the use of XBRL publicly or sorry 2009, and now public companies have to file all this information publicly in XBRL. So instead of putting the data on a PDF that nobody can read or access, they have to code their financial statement. That's the first mile from zero to 1 and saying let's take this public data and actually make it structured and digital so it's accessible	Data that that otherwise would be at a PDF and is very are hard and expensive to get access to in a systematic way. So instead of putting the data on a PDF that nobody can read or access, they have to code their financial statement.	Access
How do these mechanisms work, and what specific roles do they play in enabling openness?	The second part though is that the taxonomy for data tags in XBRL has like 75,000 Items, because a financial statement has tons of line items, revenue, operating revenue, expenses and accountants can kind of just make up whatever it's like the, the word extensible in XBRL which means they can't just kind of you can make up your own tags. And so the data is not really it's it's good that it XBRL makes it structured.	The second part though is that the taxonomy for data tags in XBRL has like 75,000 Items. which means they can just kind of you can make up your own tags. And so the data is not really it's it's good that it XBRL makes it structured.	Common terminology

Pillar	Governance Mechanisms	Transcript	Code	Validation
Platform boundary resources	Application programming interfaces	the APIs are really the mechanism to get that data flowing the harder it is to access data the less innovation is gonna happen and moving all of these industries forward, and there's new technologies now that are making it easier to access the data. The API is the easiest possible way. I mean when you come to our website, you get an account, you get API keys and within a minute and a half, you can be pulling data down from our from our databases.	Yes	Positive
	Development boundary resources	Yeah, I think that that when it comes to those resources, it's all about the speed and and openness of access. So API is the first step. You have to have the delivery mechanism, but the education and the usability around the delivery mechanism, if that's not there, it's not really open. If you can't use it easily. That's where I think the SDK's and software development kits come in.	Yes	Positive
	Social boundary resources	Yes, having documentation, tutorials, guides, resources, things like that. There are a lot of developers who just need the API and they're off to the races. They can go, but that that's not not always that easy. Not every API looks the same way. You have to learn the syntax and and learn everything, so that's a huge part of what we do at Company_Int_1 is make that easier.	Yes	Positive
Platform rules	Intellectual property rights	But the the other I guess category in this question in, in our industry, in the finance space is display rights. So that's a, it's a tool that a lot of bigger vendors use to make more money which is are you using the data internally or externally. That's a huge factor in the licensing	Yes	Positive
	decision rights	99% of our customers just accept the terms and conditions on our website, which is normal. We work with very, very, very big organizations. They have their own complete separate contract that like huge huge financial institutions and those are few and far between. Most people just accept the terms and services on our website, terms of services and conditions. I guess kind of the same thing. But they do sign an agreement or a contract also that has more detail on their specific usage of the data depending on what type of data they they buy. Everybody accepts the terms of services and conditions on the website and then also in addition to separate contract for the data.	Yes, with constraints	Positive
	gatekeeping	the access mechanism for us is the API key. It's very simple, it's just you can find it in your account. You can access to the data. I would say that the impact on openness is. Again, a lot of the in the API industry just comes back to the speed and the ease, ease, speed and usability of access. You don't have to get on the phone with a sales person. You can chat with us on the website, get your API keys, boom, your access is open.	Yes	Positive

	pricing	They stop openness a lot because of the complication of the combination of exchange fees, display fees per user fees, and then the technology delivery fee, which would be our fee. And it's confusing because we'll talk to a customer and we'll say this, if you want to buy stock prices data from us, it's \$500 a month.	No	Negative
	revenue-sharing	That and so you're basically putting the player who's willing to do the manual stuff together with the more innovative company that has really good delivery mechanisms and APIs and by working together with the revenue share agreement, you're able to take that data, which is otherwise very difficult to access and terribly supported and with bad technology, and bring it into an ecosystem where it can be delivered more easily. That's why we've established revenue sharing partnerships to unlock more data sets for our customers that we otherwise wouldn't source for them. So it it leads to more comprehensive access to the types of data you need.	Yes	Positive
Ecosystem identity	relational control	It is and I think also helpful that I identify personally as the CEO directly with them because I was an app developer who couldn't afford data and that's why I started a company in the 1st place. So they know the company is being run by people who've had the same problems with open data and we get what they're going through. So we're gonna build the best experience for them and there's a lot of trust in.	Yes	Positive

Question	Transcript	Code	Theme
What are data-specific mechanisms used to manage data platform openness?	Yes, XBRL is the main one. That's huge. I think XBRL should be like everything should be filed that way in the financial industry. It's just a standard that should be global. It works. It's proven to work, it's huge.	It's a standard	Standardization
How do data-specific mechanisms contribute to data platform openness?	We have a lot of proprietary systems in the back end for data quality checks and analysis and things like that.	Data quality checks and analysis	Data Quality
Is there any other governance mechanism related to platform openness that we haven't discussed yet?	We have a kind of a catalog on our website and it's like it's kind of all over the place and it's really tough for people to know where to go to get their data.	We have a catalog on our website for people to know where to go to get their data.	Data catalog
Are there any additional insights or perspectives you would like to share regarding the mechanisms we covered?	Not really	-	-
Are there any missing elements you would like to add before we conclude this interview?	I don't think so. I think we covered a lot of the you know the the XBRL and market data stock exchanges are the the key in the financial data space, the key, the key mechanisms for openness and the business model. So I think we covered covered a lot. I hope it was helpful.	-	-
Do you have any recommendations or suggestions to improve the openness of data platform based on your expertise?	I think XBRL should be a global standard for way more asset classes than just it's applicable to a lot of things other than just financial statements	XBRL should be a global standard	Standardization

Analysis interview 2

Interview number	Int_2
What is your job function?	Principal Consultant at a consultancy firm
Can you describe the platform of the organization you are working for?	So per European project there are different platforms that are being used and different standards that are that are being used. And that said in the the whole point of the Data spaces is that we become less dependent of those big three and develop our own data platforms and and they are called data spaces and those built on.
How many years of experience do you have with data platforms?	10 years
Can you explain what kind of experience you have with data platforms?	Performing research. I have worked also mainly on the research into architecture but also into data management and data regulation.

Question	Transcript	Code	Theme
<p>What governance mechanisms have you observed being used in data platforms to facilitate openness?</p> <p>How do these mechanisms work, and what specific roles do they play in enabling openness?</p>	<p>We make data sharing agreements between Parties. So that's one side is quite contractual and the other is really data protection impact assessment, which is kind of generic to say what is this, what it is that we do with our data, where we share it.</p>	<p>We make data sharing agreements between Parties.</p>	<p>Data Sharing Agreements</p>
<p>What governance mechanisms have you observed being used in data platforms to facilitate openness?</p>	<p>Uh the other is maybe more fundamental and that is Really, we could call that ontology mapping, let's say. So I have a particular terminology and I have a particular logic in my data architecture and the other one has a different kind of ontology and and logic and they they need to be mapped. And that's not so easy because they're often. They look, they look similar but they are not because they are particular jargon that is connected to A to a field or there's a particular legacy terminology that doesn't fit. So. So this is quite complicated.</p>	<p>We could call that ontology mapping. So I have a particular terminology and I have a particular logic in my data architecture and the other one has a different kind of ontology and and logic and they they need to be mapped.</p>	<p>Common terminology</p>

Main questions				
Pillar	Governance Mechanisms	Transcript	Code	Validation
Platform boundary resources	Application programming interfaces	so APIs open for sure, but this this can still lead to exclusionary business models. It can still lead to lock in models. It can still lead to lack of sovereignty. It's it's not the solution always there.	Yes, with constraints	Positive
	Development boundary resources	If they are used. Uh, I mean, it's super interesting because yes, it is. I think it's it is a way to promote openness. This I think we should promote that a lot more open SDK's and open use of open toolkits.	Yes	Positive
	Social boundary resources	Ah, that's a very good question. I like that question. It is getting better. I think there's there's quite some education also.	Yes	Positive
Platform rules	Intellectual property rights	They play a huge role at the moment, actually, because one of the reasons why often companies are Hesitant to really engage in open platforms is	Yes	Positive

		<p>exactly that so. There was, of course the famous Creative Commons licencing scheme that allowed for different levels of openness there and and and where you could decide what you want to be reduced, how you want people to reuse it. This this in a particular sector, this really worked, but it hasn't. It hasn't really been taken up for the product or the more or technical side.</p>		
	decision rights	<p>So in the world that I have been working in, this was never yet an issue because we always worked on on research and development that was pre market or not at the at the tier level that was so high that it was ready to be marketed. o then it becomes a bit like like the cookie button, right? Yeah. OK, whatever. Whatever. Whatever. Here we go. But. But of course, terms and conditions are one way to to promote this.</p>	Yes, with constraints	Positive
	gatekeeping	<p>So AAI means authentication, authorization and identification and that is a that is a a place where this gatekeeping , Offer some possibilities for multi sided platforms. So you can say. You are a data supplier or you you send data to us or you make it accessible only these in these parts or only particular metadata or only for a certain amount of time, or only if you have a certain clearance and it's that that is definitely it has an influence on what we discussed before on on, on the levels of openness. So what do you mean with open and how do you, uh? Because there's particular reasons. Also a good reason for some data to not be open and so. Also a good reason for some data to not be open and so. Either for competitive commercial reasons or for national interest reasons, or for how you don't want to just open all your health data to to anyone you want to make sure that you have control over that. And these AAI schemes that are quite technical and quite developed nowadays, and they they they are definitely of a hot topic and and and and much in development and in terms of connecting these these levels of openness and types of collaboration to what that means in practise.</p>	Yes, with constraints	Positive
	pricing	<p>They say the added value is really in the matchmaking. So everybody has data and nobody has data. Often they don't know that they have data. But with this pricing, it's a bit double what the pricing has done on one side. It has made a lot of companies aware that they have the data that actually has a value and that if they open this up or semi open it up or</p>	Yes, with constraints	Positive

		partially open it up or have different levels of access to it that it's actually becomes an asset. The other side is that for some types of data they they they became an asset where it's it, maybe it would have been better if they would have been just open and free for all, so it has a bit of a double-sided effect in that sense.		
	revenue-sharing	They have case studies where they where they proven that it works that and and they're quite different. Revenue streams can help in these data platforms and in this openness by understanding that often you cannot do this alone... It becomes interesting to hire a third party or to pay per service or to pay per amount of data to have a service built on top of your data. So you give this data for free to a third party and they are they are allowed to build a service on top of it and you get a percentage for instance so it.	Yes	Positive
Ecosystem identity	relational control	Now it's very fashionable to say that you're open and that you share your data, so it it definitely helps. The German industry was dealing with this. So the German industry was all in in the, in the in manufacturing and in in machinery. And they realised that they missed the boat a bit on digitization. So now Bosch and Siemens and all of them, they have big programmes on funding startups to play around in their data, to open up their data sets to so to really have this start up kind of accelerating and that they did them really well In terms of branding and tapping into completely different.	Yes	Positive

Question	Transcript	Code	Theme
<p>What are data-specific mechanisms used to manage data platform openness?</p> <p>How do data-specific mechanisms contribute to data platform openness?</p>	Well, there's the the fair data principles, right. And and they come also with protocols, there's data specific.	fair data principles	data principles
	It's a bit difficult because standards of course are one way and that is important, so data standards but more than more than standards are a result of an interest but also of an architecture and the specificness about data and those architectures is that it's, as I said, the data is is endlessly reusable. So it's you need to rethink a bit on what this, what this is and the specifics there are.	standards of course are one way and that is important	Standardization
	It has a bit to do with the types of formats that you use, the metadata and how you describe it. So the openness is really connected to all and these kinds of things. So how do you describe your data. To metadata descriptions that are that are open and readable machine findable.	Metadata	Metadata management

	And another thing that was very popular and then died out, and it's coming back a bit is sticky policies , which means you describe Rules and technical rules that you connect to a data set that automatically allow or disallow particular actions with that data. So that is a thing that is very data specific. That is one way of promoting openness or controlling it at least because you can have trace of what is happened, who can do what with it? There are different names for that now they call it sticky policies.	Sticky policies	Policies
Is there any other governance mechanism related to platform openness that we haven't discussed yet?	In general, data governance models that I've seen start a bit with ethics. That's that's always a bit vague, but then often chuckled in legal data governance models, technical data governance models and business governance models. So that's a bit the three elements that I always see coming back and I think we discussed all three, what's a bit missing is still the societal governance models. Hmm, if you're data platform owner, you could think of A bit more your wider ecosystem and also another thing that comes in more and more now, is societal impact assessment. So what happens to all this data that you're using that maybe also is owned and comes from somewhere else. And another thing is environmental impact assessment. So how much energy does actually your whole data centre consume? Should you be taxed for that? How do you know how much data you're using, how much processing power, where do you store all this data?	So that's a bit the three elements that I always see coming back and I think we discussed all three, what's a bit missing is still the societal governance models. So what happens to all this data that you're using that maybe also is owned and comes from somewhere else.	Data Governance
Are there any additional insights or perspectives you would like to share regarding the mechanisms we covered?	No, I think I could keep going for another day but I keep it at this.	-	-
Are there any missing elements you would like to add before we conclude this interview?	No	-	-
Do you have any recommendations or suggestions to improve the openness of data platform based on your expertise?	It's very important we need to think of business models that work in in that open because open is is not cannot be for free it it has cost. And we need to think about politically if that cost is public or private and or both and how that then works.	business models that work in in that open because open is is not cannot be for free it it has cost.	Open business models

Analysis Interview 3

Interview number	Int_3
What is your job function?	Former CEO of Data marketplace
Can you describe the platform of the organization you are working for?	So the platform of Company_Int_3 is a SaaS solution that provides every tool needed to be able to take content data that you have and use it in one of three use cases. The first one is packaging the data as products and selling it. The second use case is external data sharing. And then by sharing it, there's some additional calculations and derived contents that's created and that feeds back their risk systems for compliance. And then the third use case is for internal data marketplaces.
How many years of experience do you have with data platforms?	30+ years

Can you explain what kind of experience you have with data platforms?	Product management and general management.
---	--

Question	Transcript	Code	Theme
What governance mechanisms have you observed being used in data platforms to facilitate openness?	Let's centralize all of the data to have one common access mechanism, and there it is, then creating the concept of a data lake. Moving it all into one location with a central access with a data lake and Right, it's now there, but only the developers and the data scientists can now use it	Let's centralize all of the data to have one common access mechanism	Access
What governance mechanisms have you observed being used in data platforms to facilitate openness? How do these mechanisms work, and what specific roles do they play in enabling openness?	So the next step is, let's put in a data governance layer. Here's where you document all the data, all the fields And you put in the the attributes to say this can go internally , To this group, This can go internally everywhere, this can be shared externally to these categories and for data openness, that's the most important part that you need to have is to be able to, you need to figure out exactly what data you can make available to. So now I know exactly who can get access to what, but I haven't solved the problem of Actually getting making the data available to the outside world.	So the next step is, let's put in a data governance layer. Here's where you document all the data, all the fields And you put in the the attributes	Data Catalog
What governance mechanisms have you observed being used in data platforms to facilitate openness? How do these mechanisms work, and what specific roles do they play in enabling openness?	So then the next layer on top of it is a way to expose to in an easy fashion what the data is. So the the data store data marketplace is a great way to do it. Then you use other technologies you put in all the contracts and data contracts and the metadata to be able to make it systematically accessible. Now, so now you have an individual that can go find Data you need to add in there the ability to ask for the data and have some sort of approval mechanism in there to to validate who has access to what and that's often time a combination of the entitlement system and the data Governance application and then once that is there based on how the data has been packaged or prepared for delivery, then it's actually the fulfillment.	Then you use other technologies you put in all the contracts and data contracts and the metadata to be able to make it systematically accessible.	Data sharing agreements
What governance mechanisms have you observed being used in data platforms to facilitate openness?	and for me, every piece of data eventually should be and will be packaged and it will become common to exchange data between individuals, between companies, between companies and individuals.	data eventually should be and will be packaged and it will become common to exchange data	Data products

Pillar	Governance Mechanisms	Transcript	Code	Validation
Platform boundary resources	Application programming interfaces	<p>And as an organization architects their data platform, they will usually select one. Sometimes more, but usually it's they're trying to get into one ecosystem to simplify their development life. So I decide that I want to use Snowflake and Snowflake is the most extreme. In trying to not be open, it provides to the business value that it provides is. Here is a very simple mechanism for you to create your data lake and centralized all of the data when you push the data in the technology. We will make the data usable.</p> <p>and here's different ways to be able to work with the data afterwards, but as a company they have for a long, long time had the philosophy that once it's there we don't actually want this to interface with any other technologies. They literally tried to block you and built in mechanisms in there and same thing with Amazon. They want you to adopt their ecosystem to put the data on S3. They'll offer their marketplace if you want to expose it to the outside World. If you want to do it and expose it internally then you have to build all of the tools yourselves or use a tool like Revelate or one of the competitors because there's a few others now out there. Same thing with Google. So they're trying to lock you into their technological ecosystem.</p>	Yes, with constraints.	Positive
	Development boundary resources	I'd say it's it helps intermediaries, but it's not really something that we've seen. It's really more for, for me the openness of the data really goes through having the right governance, knowing exactly where the data can go, and then having a mechanism to be able to find the data by the user, request it and then get some sort of technical access to get APIs.	Possibly	Negative
	Social boundary resources	It's derivation of governance, because that's where you're exposing what is accessible or not. So it is necessary. It is part of what you need to to expose so the user knows what he can get and how he can get it.	Yes	Positive
Platform rules	Intellectual property rights	It plays an important role in how you're exposing the data. But in the platform itself, again, it goes back to you have to have a governance layer of some sort that's implemented with control mechanisms to be able to define and review and politics and processes to say.	Yes, with constraints	Positive
	decision rights	Yes it. When you make data available, it's always under certain conditions	Yes	Positive

		and those can vary greatly. Some firms will say it's internal use only. Here's all the the the conditions for you to to use it, and you don't like that either come back to us and negotiate and put in more money, or if you don't like the conditions then that's it. The conditions can vary greatly. The most extreme that I've seen is is actually one large provider, financial data that claim the rights to anything that was derived using the content that you accessed from them.		
	gatekeeping	Yeah, of course. Well, that, that's it's basically identifying through the whole chain. Who is liable to be perceived as a user versus just a strictly a processor? And in that case, what are all of the commercial terms and variants of the commercial terms for the different types of usage?	Yes	Positive
	pricing	Yeah, well, the pricing is is really based on how as a the owner of the content, I will perceive the value the data that I'm providing, I mean am I providing it first of all because I have to provide it? Am I providing it as a public service? Am I providing it as a support to my operations and am I providing it as a way to literally make money? And the data has this level of freshness, then it's worth so much if it's less than 20 minutes old then you will pay us one cent per quote that you see with a maximum. Ohh you happen to be a professional trader or you've created a a trading group with your friends. Well sorry it's not 1 cent anymore. Now you have to pay us \$36 per month for the privilege of looking at this data.	Yes, with constraints	Positive
	revenue-sharing	That's where I'll say, I don't know if it's related to openness. And of course, if there is no royalties or payments or if the less expensive the data is technically the more open it is because there's a wider audience....	Possibly	Negative
Ecosystem identity	relational control	-	-	-

Question	Transcript	Code	Theme
<p>What are data-specific mechanisms used to manage data platform openness?</p> <p>How do data-specific mechanisms contribute to data platform openness?</p>	<p>Well, there there's a number of formats that exist, so one of the ways to try to provide openness is through lobby groups and organizations. So in the financial world. So let's let's move sideways in the financial world. And central banks years and years ago created a common format called SDNX. That's a standard way of doing it, and there's open source, fixed engines, etcetera, but once you implement it, you will always have</p>	<p>Well, there there's a number of formats that exist, And central banks years and years ago created a common format called SDNX. That's a standard way of doing it, and there's open source, fixed engines, etcetera,</p>	Standardization

	some of your secret sauce that you add to it,		
Is there any other governance mechanism related to platform openness that we haven't discussed yet?	-	-	-
Are there any additional insights or perspectives you would like to share regarding the mechanisms we covered?	No, Probably talked about everything	-	-
Are there any missing elements you would like to add before we conclude this interview?	-	-	-
Do you have any recommendations or suggestions to improve the openness of data platform based on your expertise?	Yes, education. I think again right now it's a it's still extremely early. It was only less than two years ago that Gartner, the Consulting Group recognized that data exchanges and marketplaces was part of the data fabric or data infrastructure that an organization should have.	Education	Education

Analysis Interview 4

Interview number	Int_4
What is your job function?	Assistant Professor at the Erasmus University
Can you describe the platform of the organization you are working for?	So I worked with narrative, which is one of the data marketplaces in the US. What happens on this marketplace is that there are buyers and sellers who come to this marketplace. Sellers, you can think of companies like MasterCard, visa, Uber. They they have a lot of data with them and they want to monetize it. So they come to this platform on the other hand, on the other side, you have buyers who are interested in such data points. What this form does is essentially facilitates the trade between buyers and sellers.
How many years of experience do you have with data platforms?	5 years
Can you explain what kind of experience you have with data platforms?	So mostly what I do is that I talk to these data platforms, I look at their operations that how they are doing their businesses and eventually my goal is to look at academic problems that I can solve. So I sort of motivate my academic research by looking into the operations of these data platforms. So the problem is motivated from these platforms. Then once I am have modeled the problem I go back and I see whether I have I have accurately modelled the problem and once everything is done, I also go back and give them their feedback, give them the feedback that you know what insights I obtained from solving this problem. And so that's sort of an academic industry collaboration that that usually happens, which is there.

Question	Transcript	Code	Theme
What governance mechanisms have you observed being used in data platforms to facilitate openness?	One thing that I would say that has happened is data sharing agreements. When you are, when you are setting up these governance mechanism, you need to make sure that you know who is, what are the sort of agreements, who is the owner of data, what sort of role access do you give to different individuals within an organization and then also outside the organization. So data sharing agreements has to be in place. Making sure that you are using the data in the way that has been specified in the agreement. That's one big governance thing.	One thing that I would say that has happened is data sharing agreements. you need to make sure that you know who is, what are the sort of agreements, who is the owner of data, what sort of role access do you give to different individuals within an organization and then also outside the organization	Data sharing agreements

What governance mechanisms have you observed being used in data platforms to facilitate openness?	The second thing is standardization. So standardization efforts have to be in place. You need to define common formats, protocols and interfaces for data exchanges within the ecosystem. And that's how you make sure that you know data platforms can seamlessly interact and share data. So this is also closely linked to sharing data, data sharing agreements. So standardization I think is one big thing which I observed. When you have these populations scale Governance mechanism, then standardization is one thing	So standardization efforts have to be in place. You need to define common formats, protocols and interfaces for data exchanges within the ecosystem.	Standardization
What governance mechanisms have you observed being used in data platforms to facilitate openness?	and then finally data licensing. So data licensing makes sure that you are using the data in the way that has been specified in the agreement.	Data licensing makes sure that you are using the data in the way that has been specified	Licensing

Main questions				
Pillar	Governance Mechanisms	Transcript	Code	Validation
Platform boundary resources	Application programming interfaces	I think this API access sort of enforces this all the 3 things that I mentioned right? So the standardization is there. Otherwise let's say our scraping data then there is low standardization but once you provide API access then the data owner can standardize the data format in which all the parties can access data, right? So standardization is taken care off. Also through API access you can also make sure that the data sharing agreement is in place, so typically these things are specified that you know what, what sort of data will be shared with you, what data will not be shared with you, and then also the licensing is in place, right? So I think API is one of the enablers of data governance.	Yes	Positive
	Development boundary resources	When you talk about software, if you're thinking about let's say open data policies, right, so you need to make sure that all these software architectures is in place where you can build let's say data products using the software that is there, right. So so open if you're if you're going for open data policies then that needs a separate architecture compared to architecture which is not governed by open data policies, right. So there the software There is a crucial role. I think one of the critical components in the software development kits, which is happening now. Is how do you ensure privacy of the consumers, right. So let's say when you're talking about consumer data, that's critical component of privacy as well, right.	Yes	Positive

	Social boundary resources	Right, so so documentation I would say documentation is an important where if you have the right set of documentation that sort of enables standardization in the processes. So so training and support, I think this this comes at a later stage when you have the basic infrastructure ready, right?	Yes, with constraints	Positive
Platform rules	Intellectual property rights	There is also varying degree of openness that you have in different licenses, where commercial licenses are a bit different compared to open licenses. So, that's where the difference comes in. So typically in commercial licenses, the access this is very restrictive. You are only required to do specific things with the data, whereas in, whereas depending of course on what type of open license do you have, you can freely make changes. Sometimes you can even commercialize it. You can build commercial products using the data so that's why. All this data when it is collected and it is made free to use under open licenses, then you can build products on top of it, right? Now, having said that, what has happened is that you don't see much applications being built on top of it, and there are multiple reasons. One is that you know, sometimes open data is not of good quality. There are not many checks happening that you know whether you can rely on such type of data.	Yes, with constraints	Positive
	decision rights	so again it goes back to the pillar of data sharing agreement and you when you so when in the agreement, it should be clearly described that who is going to make the decisions, who is going to analyze data, who is going to use the data, where the data is coming from, right.	Yes	Positive
	gatekeeping	Yes, but I think this is a major, challenge for organizations I would say. And that comes primarily from security reasons. So earlier, what used to happen we is there was this role based access that was given to different individuals in an organization that if you are manager, you have access to these type of data and you don't have access to these type of data. So then it is it gets very difficult to do this role based assignment. Then what happened is there was a paradigm shift and then organizations moved towards this and forgetting the name. But the idea was this perimeter based access. That typically you think of an organization as an onion.	Yes, with constraints	Positive

	pricing	<p>there are these 2 schools of thought that, you know, once some say that, you know, consumer data should not be monetized, there are privacy concerns and this and that so. So there is this school of thought because of which, you know, all these institutions, which small scale institutions which sprung up, which fought for consumer rights, consumer data, they sort of ignored this thing of data monetization and they did not scale up eventually. So if you're thinking of data as a resource, you need to price it, you need to monetize it. I think the separation that should happen is that whether you have obtained for consumer data, whether you have obtained the necessary consent to monetize the data so that should be separate the consent and the privacy concern should be separate and the monetization aspect should be separate. And I think that 2 things can go hand in hand. You can have privacy and you can have monetization as well. So I think if you want to have this governance mechanism, the right governance mechanisms in place. So that's the broad thought I have. There have been concerns about privacy and all these things, but then that can be separated from the monetization aspect. OK, so pricing should be included in the governance mechanism. They enable price discovery, I would say so. So openness in a very different way. If your data is being traded in a very siloed way in a very non open way, I would say that you know it is being passed to a data broker, goes to some other party, and then you don't know where your data has ended up. So that way I think you would not enable price discovery and that way you would not be able to know what is the value of data in itself. So if you do want to enable what is the right value of data? And I'm talking about openness in this way in terms of price discovery in getting to know about the value of data, good monetization mechanisms have to be in place. I would say without that what would happen is that your data. See your data is going to be traded in one way or the another. If you are enabling transparent mechanisms of trading data that will at least help you in price discovery that what is the actual worth of data.</p>	Yes, with constraints	Positive
--	---------	--	-----------------------	----------

	revenue-sharing	So the revenue sharing model is typically, let's say, if you're talking about the platform, typically the seller, the platform charges a percentage fee from the seller, but then usually there are very few data platforms. I would say 2 sided platforms, where such things happen, so these platforms charges fraction from buyers and sellers both sides like Amazon does. B2C or B2B selling where there is just one platform, where there is just a firm who is selling data through, let's say an intermediary or somewhere. There usually this revenue sharing does not happen. So first of all, for consumer data that is being sold, there is no revenue sharing. Consumers are not given a fraction of the revenue that is being made by monetizing their data. That only happens in cooperatives, data cooperative where data cooperatives. The idea is to pass on the share of the revenue to the data generators, right? But when it comes to data platforms, usually firms who have data with them, whether it's consumer data or whether it's non consumer data. They do not pass on the revenue to the consumers in general and that sort of that.	Yes, with constraints	Positive
Ecosystem identity	relational control	To be honest, I'm not aware about this a lot, so I do not have much idea about this.	-	-

Question	Transcript	Code	Theme
What are data-specific mechanisms used to manage data platform openness? How do data-specific mechanisms contribute to data platform openness?	So so one thing that has been recently happening is about identifying the source of data and through how the flow of data in general. Am I the first one to use it or has it been passed on through some other things, so tagging data I think that is one of the one of the things which firms are paying attention to and that also creates adds to openness because if data has exchanged many hands. Then you have sort of lost value of it and then you don't know who has access to data and the security concerns arrive. So so though that is, that is one thing about tagging data, tagging the source of data.	so tagging data I think that is one of the one of the things which firms are paying attention to and that also creates adds to openness	Data classification
	Licensing, which is that if I have 3 terabytes of data and I'm going to share it with my partner firm, I should make sure that what sort of things that I that my partner firm can do with it, if there is a secure breach. I have to report, my partner has to	Licensing	Licensing

	report all these things that should be in place. Right, so all this.		
	And then in general, the flow of data, right, so that I think comes comes if you if it's within the organization forms in general are not tracking data that you know what data is being used for, what purpose is.	Flow of data	Data Lineage
Is there any other governance mechanism related to platform openness that we haven't discussed yet?	I think, at least from what I know, I told you most of it.	-	-
Are there any additional insights or perspective you would like to share regarding the mechanisms we covered?	Not in particular.		
Are there any missing elements you would like to add before we conclude this interview?	-	-	-
Do you have any recommendations or suggestions to improve the openness of data platform based on your expertise?	From what I studied, I think the incentive mechanism should be in place. If you're talking about openness, if you're talking about first building a data platform, think about where the data is coming from, who is sharing that data right? If it's, let's say, coming from individuals. Are you sufficiently compensating individuals for taking their data? If it is coming from forms, are forms ready to share their proprietary or strategic data with you? So if you want to think about building value from these data platforms in the long run, you need to provide sufficient incentives to collect data in the first place.	I think the incentive mechanism should be in place. you need to provide sufficient incentives to collect data in the first place.	Incentive mechanisms

Analysis interview 5

Interview number	Int_5
What is your job function?	Professor at the Rotterdam University of Life Sciences & Chief Science Officer at a dutch ministry
Can you describe the platform of the organization you are working for?	I am not working with data platforms, we are developing and implementing data platforms for our research, so I'm not using a specific data platform, but one of the things that we are doing, we are trying to design a data platforms and not that you can use it directly, but prototypes of these platforms and then we can see where the bottlenecks are, how the research questions are and if you are trying to develop and to design these kind of platforms for example, well, what are the data sharing issues on such a platform? What are the data quality issues if you bring data from different sources together in one?
How many years of experience do you have with data platforms?	30+ years
Can you explain what kind of experience you have with data platforms?	Well, what we are working on and if you have these data platforms and the idea behind this data platforms is that you have data from different organizations, different sources, you bring them together in one platform, so such that people other people can use the data for their purposes. So what we are working on is saying, OK, if we bring this data together, how can we share this data to other parties in a in a privacy proof way if we at least in governmental data, you cannot share with everybody. You can share it, but you have to obey the rules and regulations. So that's the thing that we are looking at. And beside that we are looking if we bring data from different sources together, sometimes they have some overlap.

Question	Transcript	Code	Theme
----------	------------	------	-------

What governance mechanisms have you observed being used in data platforms to facilitate openness?	Actually what we see is an we are very Eager to open data, but before you can open data you should respect for example the GDPR. So and what we do in this, if we look at the data we look how sensitive it is and then and what kind of privacy rules and regulation applied to it.	what kind of privacy rules and regulation applied to it. you should respect for example the GDPR.	Privacy
What governance mechanisms have you observed being used in data platforms to facilitate openness?	That we are working on them and to facilitate this type of data what we do, we are looking to new technology for example how to encrypt this data homomorphic encryption. Because if you can encrypt the data then you can do with the data what you like but you cannot go back to the original data. Yeah, encrypting is one of them,	and to facilitate this type of data what we do, we are looking to new technology for example how to encrypt this data	Data encryption
What governance mechanisms have you observed being used in data platforms to facilitate openness?	and another one is, and which attribute should you anonymize ? So these are important things...	and which attribute should you anonymize	Anonymization

Main questions				
Pillar	Governance Mechanisms	Transcript	Code	Validation
Platform boundary resources	Application programming interfaces	OK. Well, that can be in a different ways. What you can have is first that you say well, we have the data, we do all kind of preparedness and so on. And you have an API and the API you can do All kind of checks.	Yes	Positive
	Development boundary resources	Well, the technical issues, well they they are fairly useful because if you have the choice, if you have a data set and you have the data set and you say, well maybe some of these Are very sensitive, so that could be a reason not to publish the whole data set. But if you have technical and mechanisms that helps you to anonymize some of these attributes, then you can still you can still use this data set partly.	Yes	Positive
	Social boundary resources	Yeah. Well, what you have documentation and these things are very important because if you have a data set, you have openness, you have open data set, then still, if you do not have the documentation, if you do not know what the semantics are behind the data set then it's very hard to understand what it actually means and if you do not,	Yes	Positive

		can really understand what such an a set means. You cannot apply it properly		
Platform rules	Intellectual property rights	Actually I do not have a clear answer on that, but I think intellectual property can help you To improve the quality of the data, because if we know that someone is responsible for the data then I assume if you are, if everybody knows that you are responsible for the data, then probably you will be more careful about the quality of you will take more care of the quality of the data	Possibly	Negative
	decision rights	Yeah, they're actually all these things balls down in and gaining the trust of people who are using the data. So if you have all these mechanisms and they are working properly then and people will trust the data better. And if they trust and if they trust the data they will be tempted to use the data, so it it all these mechanisms may be very useful to create trust about the data.	Yes	Positive
	gatekeeping	Yeah, access rights and the relation that's opening the data well. You know you have two things you can access the data. And you can provide people all kind of rights about the data, but more interesting is always how people will use the data. You see access is 1 But use control is even maybe more important. So I would say well give them access to all the data but an Monitor about the use of the data. So use control is coming in place.	Yes, with constraints	Positive
	pricing	While I don't think that we are willing to pay for data. At least from a research point of view. yeah. But maybe in a commercial point of view that might be different.	Possibly	Negative
	revenue-sharing	I have no idea about how they are implemented to be honest, so I skipped that question.	-	-
Ecosystem identity	relational control	You can say two things about it and one is that you say now we have some a set of rules and values and the data owner, they should enforce that everybody's obliged to these rules and values. Another one is that you say well. Another one is that you say well we are all peers. So what you	Yes, with constraints.	Positive

		should have you should say OK if you are all peers then we have all a shared responsibility.		
--	--	--	--	--

Question	Transcript	Code	Theme
<p>What are data-specific mechanisms used to manage data platform openness?</p> <p>How do data-specific mechanisms contribute to data platform openness?</p>	<p>Well, actually utility. As a measure is also important to say, OK, we have this data data set and to what extent is it useful and for what purposes is it useful to give a users or stakeholders the insight for which purposes is this data useful? Because not all data is used for for all type of applications. Yeah, having insight in for what type of applications or what for type of question which data is useful is also I think important.</p>	<p>To what extent is it useful and for what purposes is it useful to give a users or stakeholders the insight for which purposes is this data useful?</p>	<p>Utility</p>
	<p>And and actually what you should do is if you look at these ecosystems, then you have a set of roles and responsibilities and you should divide or share these responsibilities to these stakeholders of these ecosystems. That's one of this model that you could follow.</p>	<p>if you look at these ecosystems, then you have a set of roles and responsibilities</p>	<p>Roles and responsibilities</p>
<p>Is there any other governance mechanism related to platform openness that we haven't discussed yet?</p>	<p>I don't think so. I think we have discussed quite important things.</p>	-	-
<p>Are there any additional insights or perspectives you would like to share regarding the mechanisms that we covered?</p>	<p>Not really right now</p>	-	-
<p>Are there any missing elements you would like to add before we conclude this interview?</p>	<p>Well, not exactly</p>	-	-
<p>Do you have any recommendations or suggestions to improve the openness of data platform based on your expertise?</p>	<p>Yeah, and well, the utility because we have now a lot of open data sets, but they are not really used.</p>	<p>the utility because we but they are not really used.</p>	<p>Utility</p>

Analysis interview 6

Interview number	Int_6
What is your job function?	CEO of some companies
Can you describe the platform of the organization you are working for?	<p>So one of them is (Client), it's called uh, it's in Belgium. They are in Antwerpen and Gent, they are practically doing I think 70% of the municipalities in Belgium. And they have a platform and it's called (Platform name). So they were focusing on creating the creation of building blocks. When you have building blocks, it means that you scope each requirement or set of requirements to a specific supplier and they have a focus and they are good at it. For example, you have a supplier that's dealing with uh, culture, and so the culture houses access for that one. One that deals with asset management, digital asset management, one that can deal with IoT platforms and how to work with electrical charges for example. And because you have very broad domain of requirements, they went and they defined or they would like to have and they have it for now. They would want to have an API ecosystem. OK, so microservice ecosystem and their and their in their space and that's that's that comes along with a lot of data.</p>

How many years of experience do you have with data platforms?	14 years
Can you explain what kind of experience you have with data platforms?	I had to experience in the beginning with ECM's, so content management systems like Documentum and all fair school, those kind of those domains where you have to also uh, create knowledge bases for specific customers and and to create a semantical value on top of the data. I have experience with time series data a lot where you need to measure stuff and also statistically uh provide insights in data. I have experience in benchmarking and ML where you need to for example benchmark machines like high pressure machines or heat pumps and stuff like that. I have data, experience and governmental data, but also in security documents where you need to provide verifiable claims for a specific attribute or some traits that the individual or a company has. Most of the time I always have a technical architectural role as well in those domains, so most of the times when I did something and the the ones I mentioned to you I had also I was also the technical architect of those.

Question	Transcript	Code	Theme
What governance mechanisms have you observed being used in data platforms to facilitate openness?	Yeah, open API specification that was coming up in that time, which was typically the I think open it. It was not called the Open API, it was called the Swagger version two...	open API specification	Application programming interfaces
What governance mechanisms have you observed being used in data platforms to facilitate openness?	That was quite tricky because you have to educate them as well.	because you have to educate them	Education
What governance mechanisms have you observed being used in data platforms to facilitate openness?	creating a a language which can be understood and spoken from with the stakeholders and the technical people because most of the time you have discrepancies between it. So I worked a lot on defining the language so meaning that when I talk about something, the other person in another domain, the domain expert for example, and the technical guy, they understand the same, they have the same meaning of that word...	defining the language	Common terminology
What governance mechanisms have you observed being used in data platforms to facilitate openness?	Secondly is how will we work together? So how do we define relationships and the roles and responsibilities? Who will do what's when those kind of things is typically what you also do in the governance setup.	So how do we define relationships and the roles and responsibilities	Roles & responsibilities

Pillar	Governance Mechanisms	Transcript	Code	Validation
Platform boundary resources	Application programming interfaces	API is how they, but because you are conforming on an interface, so when you when you look at the API or rest APIs or GRPC APIs, whatever the goal is that you have a deterministic interface that can be used by somebody else. That's the theory behind it. But most of the times you have to also follow some rules... Well, is that when you document this well and you have a clear API which	Yes	Positive

		you manage which you control, it's openness because you I can just choose and say, OK, let's go in the marketplace.		
	Development boundary resources	<p>You have two ways there, right? You have the the technical people who says I can make it all by myself and you know. As an SDK on the API. Yeah, that's it has purpose. In my opinion it has its purpose for backwards compatibility. The thing I just mentioned before with the consumer driven contracts, you only have one place to solve it. It's an SDK because. When you don't solve it there, it means you hit the the consuming application site and he needs to change it. So that's one way of dealing with that. So it has its use SDK because you can fix a lot of stuff in an SDK like what you have with the BFF kind of principle back and forth front end you can fiddle and fix things or change stuff to uhm for the the compliance with the existing solution you can fix it there in SDK. SDK's also not only to compatibility or backwards compatibility. It also solves the issue of more complicated sequences or transactions and so for example, I just need to authenticate somebody, but then you need to discover a reader. You need to fix check if he has a a key or a certificate with the signing attribute in it. You need them to upload a document and then you need to sign.</p>	Yes, with constraints	Positive
	Social boundary resources	<p>But they're just merely examples to get you speed up very quickly with and then again documentation. Big question Mark and you have a lot of. That's something which is really vague. You don't have a really standard way of documenting, to be honest, but documentation can also raise confusion if it's not well done well in my opinion. But if it's straightforward, simple with some code examples with some easy teaching mechanisms inside of it to get you really fast from A to B without knowing all the details, that's good documentation that can help for sure. Maybe to come back on the Rust thing I mentioned you. it's an interesting one to give another example because the way how they deal with API interfaces not from a rest API point of view, but a library interface for example, they do it very well because they don't only expose the documentation, they have a way</p>	Yes, with constraints	Positive

		of validating the documentation and the code examples in the documentation before pushing the release. That's the principle which I like a lot because I'm reassured if I use that library that the documentation is still up to date.		
Platform rules	Intellectual property rights	<p>Very important, but not only licensing. Yeah, so licensing. So dependency management is one of the crucial things that you need to have. You need to be aware of what dependencies do I use and what I'm a little bit strict in those but and.</p> <p>When I'm technically involved in a project, nobody is allowed to add a dependency without going through an architecture report. Nobody, because that's typically something that's technical guys, they don't know very well the details about. That's one thing. So very important thing, also technical people, they need to sign an additional addendum in the contract saying that when they do something, their IP belongs to the company or the product and so on and so on, but it's all it all has to do with the governance of your intellectual property and also the value of your product.</p>	Yes	Positive
	decision rights	<p>Yes. And it's also a little bit linked to the definition of your license. It depends on your license and together with your license terms and conditions are aligned with those. So it it all comes down back to ownership of the codes or or a financial impact for me to be very open on that. You know, because if you have for example a product and you need to choose between an AGPL V2 or an MIT license. For technical people, they say, well, it's quite the same. It's not. It's a big impact. So terms and conditions, even when you have a selling proposition or an integration project with this, when somebody will use your interface, that's something you need to fix and you need to pull it early to, you need to have it formalized very deterministically, yeah.</p>	Yes	Positive
	gatekeeping	<p>yes, that's still a mess sometimes today, but you have good ways of doing it most of the time, everything one of the most important stuff when you want</p>	Yes	Positive

		to, even when you use role based access control, you always fall in a situation where you need some kind of roles and some kind of permissions which can correlate down to all those interfaces. If I just look at the API rest interface or the microservice ecosystem, that means that you must have an identity provider set in place in your architecture. Means one component owning the authentication and the role and the permission part. That's one thing which is important.		
	pricing	Yeah, a lot! Yeah. It depends on how you see it. If it's an open API, you can say by just copy it over and have it run by yourself. And that's not the same definition. An open API means I open my knowledge through an interface to have something created with that or put added value on top of that. That's the goal. If you want to have added value on top, lower the price, go for distribution, open your data and let's go for it. So pricing of open APIs definitely impacts a lot how people will use it or not use it. It's very simple. I see two APIs doing the same stuff. The one is I have to pay for the other one not but it lacks one feature. I will contribute that feature to the API and it will use it for free, so it impacts a lot you know.	Yes	Positive
	revenue-sharing	Yeah. Typically you work with plans and each plan can be attached to some financial model. But typically when you have a plan and you have a financial model, depending on the functionality that you expose, you can let people pay. The only benefit you have is to again onboard. It's a different way of distributing openly. So when the data comes from an added value from that party that wants to earn money on the data set. It's agreeable if the data comes from people stupidly participating in social events, and that data has been sold as a data set that's more an ethical question, which I'm against too, but that's an ethical thing.	Yes	Positive
Ecosystem identity	relational control	it doesn't.	No, not relevant	Negative

Question	Transcript	Code	Theme
	Uhm, yeah. Protocol implementations and specifications. So very technically. Uh, typically you have a lot of those are rest APIs, so following HTTPS protocol they have an open API which has no API, which is a generalization of the specification which you can share along. They have tooling around it like. Typically the API client tooling like postman and insomnia and stuff like that. You have the tooling for generating of client codes independently of the code that's helping as well. So we have. It's typically protocols and specifications that facilitates that, yeah.	It's typically protocols and specifications that facilitates that, yeah.	protocols and specifications
	Event buses, for example, Typically you have an event bus, so means that you can subscribe on the bus and you can collect the data based on the event that happens in the system. That's also an openness of the system.	Typically you have an event bus, so means that you can subscribe on the bus and you can collect the data based on the event that happens in the system.	Event buses
Is there any other governance mechanism related to platform openness that we haven't discussed yet?	I mentioned one already which you didn't ask explicitly. It's about the individual. The end customer is an important one, which is always overseen. I'm using your API. I'm also a business but t I use a customer at the end there is a lack of protection and personal data and behavior and profiling, so I know it's a business model on itself to profile people and have.	I use a customer at the end there is a lack of protection and personal data and behavior and profiling	Privacy
Are there any additional insights or perspectives you would like to share regarding the mechanisms we covered?	Uh, you have a lack of legal they need to catch up all the time with a lot of stuff they can kill the business case as well sometimes because of a lack of definition, a lack of legal framework.	lack of legal they need to catch up all the time with a lot of stuff they can kill the business case	Rules and regulations
Are there any missing elements you would like to add before we conclude this interview?	-	-	-
Do you have any recommendations or suggestions to improve the openness of data platform based on your expertise?	Well, when you look at the practical side of the documentation side is still lacky in my opinion. They enforce you to from a technology point of view, they force you to comply. They force you to make sure that the ecosystem takes care and that they can rely on you	documentation side is still lacky in my opinion	Documentation

Analysis interview 7

Interview number	Int_7
What is your job function?	Owner of a data engineering company
Can you describe the platform of the organization you are working for?	Bijvoorbeeld voor (Client) daarvoor bouwen we platform, dan doe ik ook de data architectuur. Ja dus data providers zijn in dit geval een aantal partijen die dus product data aanleveren en productdata dat klinkt wat eenvoudig, maar dat is het niet. Dus het zijn allerlei codes voor producten en die zou je verbazen hoeveel verschillende soorten (Client) je wel niet kan hebben, maar iedere keer, stel je ze voor dat iets veranderd In het fles etiket of bijvoorbeeld hoe heet dat. Dus zij hebben een soort van database met allemaal codes van nou, wat hangt met elkaar Samen. Vervolgens heb je ook laboratorium data die hebben gegevens die testen, dus die producten op die productcodes en die laat zeggen van nou, wat zit er allemaal in?
How many years of experience do you have with data platforms?	10 years
Can you explain what kind of experience you have with data platforms?	Ja dus, aan de ene kant heb ik het data architectuur gedaan. Dat is echt het ontwerpen van database systemen, dus bronsystemen, maar ook het ontwerpen van datawarehouses, dus stermodellen op basis van bepaalde bronsystemen. Ik heb niet zoveel gewerkt met standaard stukken software. Ik heb eigenlijk veel meer gewerkt op laag niveau dus direct in databases.

Question	Transcript	Code	Theme
What governance mechanisms have you observed being used in data platforms to facilitate openness?	Misschien dat ik wat beter kan gaan, Ik ga het even beantwoorden in wat We gaan doen, want We zijn nu middenin wat wij willen gaan doen. Is in het datamodel willen we eigenlijk een aantal data Owners gaan aanwijzen.	Data Owners gaan aanwijzen.	Roles & responsibilities
What governance mechanisms have you observed being used in data platforms to facilitate openness?	In de zeg maar views willen We gaan definiëren en die willen we eigenlijk allemaal gaan labelen	die willen we eigenlijk allemaal gaan labelen	Data classification
What governance mechanisms have you observed being used in data platforms to facilitate openness?	Dus wil eigenlijk gaan zeggen van je hebt een bepaalde gebruikersgroep die heeft toegang tot die en die en die views in die in die vorm en op die manier controleer je eigenlijk van hoe dat klassiek gebeurt ook in datamarts.	een bepaalde gebruikersgroep die heeft toegang tot die en die en die views	Access
What governance mechanisms have you observed being used in data platforms to facilitate openness?	Maar wat we ook gaan doen en Dat is denk ik wel redelijk innovatief is dat we In de verwerking van de bron naar geaggregeerde data gaan we ook de IP gaan we mee Laten lopen in de lineage	in de lineage	Data lineage
What governance mechanisms have you observed being used in data platforms to facilitate openness?	Ze hebben wel gebruikersrechten, maar ze hebben niet de rechten om die data door te verkopen	de rechten om die data door te verkopen	Licensing
What governance mechanisms have you observed being used in data platforms to facilitate openness?	Dus daar moet de hele duidelijke afspraken over gemaakt wordt.	duidelijke afspraken	Data sharing agreements

Pillar	Governance Mechanisms	Transcript	Code	
Platform boundary resources	Application programming interfaces	<p>De de website van hun die gaat een API krijgen naar de backend van dat dataplatform zeg maar en die via die app. Die krijg je een stuk identificatie, dus je hebt je. Je zorgt dat je daar een stuk security of zet dat bij machine to machine interfaces dat dat je dat je kan controleren of nou ben jij ook de persoon bij die groep heb je ook die rechten om via deze aapjes te kunnen werken.</p> <p>Nou, wat ik heel fijn vind aan API's is dat ze ze werken als een soort van dubbele interface, dus als jij als jij een API hebt, dan beheer je het eigenlijk aan twee kanten en je hebt gewoon een afspraak met elkaar waardoor je aan jouw kant intern dingen kan wijzigen zoals je dat wil.</p>	Yes	Positive
	Development boundary resources	<p>Je hebt backend code en daarmee bouw je bouw je dingen? Ja, ik zou zeggen de rol beperkt omdat het heel indirect is. Nou ja je, je bouwt daar Natuurlijk je API's mee en je bouwt daar je bekend mee en je bouwt daar alles mee wat wat tegen een database aan praten en wat vervolgens weer dingen aan aan API's gegevens dus. Dat is de rol, en ja, een goede jou goede backend met goede software en goed gecompileerd zorgt voor zeker ja, robuustheid in je in je systeem staat ook min of meer centraal in in beveiliging en security.</p>	Yes, with constraints	Positive
	Social boundary resources	<p>En dan ga je feitelijk kijken wat er is en dan kom je erachter dat het allemaal wel is gedocumenteerd vastgelegd, beschikbaar is. Wat ik dan wat je dan vaak</p>	Yes	Positive

		<p>ziet, is dat er dan niet goed gewoon niet goed gecommuniceerd is of dat er gewoon geen goede informatie is vanuit hier en hier ligt het allemaal en effectief bereik je dan geen openheid omdat Mensen gewoon simpelweg niet weten dat het bestaat. Dit soort zachtere dingen. Ja, ja, lijkt me heel logisch dat je dat dat de manieren zijn om Mensen recht wijs te maken, bekend te maken. Het is pas echt open als Mensen het ook weten te vinden en ze het gebruiken.</p>		
Platform rules	Intellectual property rights	<p>Je koopt een dataset van zo'n partij, zo'n laboratorium en die zegt je mag het gebruiken, maar je mag niet deze cijfers een op een openbaar maken. Nou Dat is belangrijk, want dan ja, dat mag dus niet in die vorm open, maar Misschien wel in een meer geaggregeerde vorm. Dus dan het tracé van die IP een zorgt ervoor dat je dat je daar controle over hebt.</p>	Yes	Positive
	decision rights	<p>Dus Ik denk dat algemene voorwaarden In de vorm zoals ze nu vaak gebruikt worden niet handig zijn, Maar ik denk wel dat ze een kans ligt om daar, zeker Als je er meer langdurig samenwerking hebt, om in die voorwaarden een afspraak te maken. Maar ik denk dat je dat dan eerder een aanvullende afspraak doet ter vervanging van algemene voorwaarden dat dat meer zal voorkomen dan dat je je daar in je algemene voorwaarden al rekening mee houdt.</p>	Yes, with constraints	Positive
	gatekeeping	<p>Dus wil eigenlijk gaan zeggen van je hebt een bepaalde gebruikersgroep die heeft toegang tot die en die en die views in die in die vorm en op die</p>	Yes	Positive

		<p>manier controleer je eigenlijk van hoe dat klassiek gebeurt ook in datamarts. Dan wordt er vaak ook gezegd, komt de data mart voor deze personen datamarts, Dat is eigenlijk wat hier ook een beetje gaan doen dat we verschillende groepen gaan Laten subscriben op bepaalde data bronnen.</p> <p>Maar wat we ook gaan doen en Dat is denk ik wel redelijk innovatief is dat we In de verwerking van de bron naar geaggregeerde data gaan we ook de IP gaan we mee Laten lopen in de lineage</p>		
	pricing	<p>Als je terug naar Adni gaat, zij willen bepaalde data delen met bijvoorbeeld commerciële partijen, maar ook met onderzoekers. En wat zij dus doen, is dat ze dezelfde datasets willen ze een ander prijsmechanisme zetten, dus ze zeggen eigenlijk van nou, onze missie is een maatschappelijke, dus maatschappelijke partijen Laten wij minder betalen voor bepaalde datasets als commerciële partijen, Omdat die commerciële belangen hebben.</p>	Yes	Positive
	revenue-sharing	<p>Ja nou bijvoorbeeld, dit voorbeeld van net, daar zijn we dus naar het kijken om die leveranciers van test data op bijvoorbeeld die laboratoria om hun inderdaad te Laten betrekken in een stukje revenu sharing van wat er mee gedaan wordt en dan tegenover te zetten dat zij akkoord gaan met meer openheid op het gebruik van hun gegevens. Dat je eigenlijk, zegt Van Nou, je krijgt daar een stukje voor terug. We kopen het niet Alleen in, Maar we geven je ook nog een stukje extra voor de gebruikers.</p>	Yes	Positive
Ecosystem identity	relational control	<p>Die vind ik heel Moeilijk, ik vind het moeilijk om het relational control aan zich</p>	-	-

		zeg. Ik, Ik kan het zo snel niet voorstellen		
--	--	--	--	--

Question	Transcript	Code	Theme
<p>What are data-specific mechanisms used to manage data platform openness?</p> <p>How do data-specific mechanisms contribute to data platform openness?</p>	<p>Dat jou gewoon persoonsgegevens delen dat Dat is sowieso moeilijk, dan moet je toestemming voor hebben. Dus anonimiseren zou al heel erg helpen om dingen te kunnen delen. Ja, het hangt ook heel erg van het soort gegevens af, hè? Wat je wel en niet kan delen, kijk. Neem bijvoorbeeld het meest uiterste wat ik ken. Bijvoorbeeld gemeentes die hebben riolerings data waarmee we werken, Dat is openbare data. Dat is net Als de Bach de BRK weet je op percelen in het Kadaster. Die informatie is dusdanig openbaar dat als jij nummertje hebt, dan kan je exact vinden wat Het is. Dat is hele andere heel ander soort data dan elektronische patiëntendossiers, maar een andere uiterste noemt, Maar dat is super beveiligd dus. Maar dat, dat zit hem allemaal weer in. Dat komt dan weer terug op dat labelen. En dat.</p>	<p>Dus anonimiseren zou al heel erg helpen om dingen te kunnen delen</p>	<p>Anonymization</p>
	<p>dan kan je data aanbieden in een vorm die je die past. Binnen R of SPSS met de benodigde onderzoeksmethode metadata erbij. Als je gaat nadenken over de verschillende soorten gebruikers en hoe zijn je data willen gebruiken en je zorgt ervoor dat je het voor die doelgroepen hapklaar aanbiedt In de vorm en op een complexiteit niveau die zij snappen en wij kunnen werken. Ja, dat helpt heel erg voor toegankelijkheid</p>	<p>data aanbieden in een vorm die je die past</p>	<p>Standardization</p>

	en dus ook bij openheid van data.		
Is there any other governance mechanism related to platform openness that we haven't discussed yet?	Even kijken hoor, nee denk het niet	-	-
Are there any additional insights or perspectives you would like to share regarding the mechanisms we covered?	Kan ik niet zo snel bedenken.	-	-
Are there any missing elements you would like to add before we conclude this interview?	Denk het niet, Ik heb het gevoel dat de ton redelijk geleegd is.	-	-
Do you have any recommendations or suggestions to improve the openness of data platform based on your expertise?	We hadden het net over standaarden. Er zijn een hele hoop standaarden en het probleem is vaak dat er dan een meta standaard wordt gemaakt en dan hadden we 100 standaarden en die hebben we 101 standaarden en dan gaat er weer iemand z'n best doen en 102 standaarden en zo zo gaat dat altijd maar door en wat je ziet is dat in de IT gaan dingen zo snel dat op een gegeven moment soort van emergente standaard is. Een meta standaard waarin standaarden geleverd kunnen worden. Dan hebben we in ieder geval dezelfde taal waarin we standaarden beschrijven en dan kan je in ieder geval zeggen van, Ik ga mijn data delen in ISO nog en nog wat	Een meta standaard waarin standaarden geleverd kunnen worden.	Standardization

Analysis interview 8

Interview number	Int_8
What is your job function?	Capability lead for data engineering
Can you describe the platform of the organization you are working for?	for a large construction company in the Netherlands, we have created modern data stack, where we collect data from all their source systems, but also data from outside the organisation. Where we collect data from all their source systems, but also data from outside the organisation, so third party. So the data that is actually not created by them and we combine that data within a central data platform and provide reports on top of it with well dashboards created in Power BI. The data platform itself is created with a medallion architecture. So that means that when we connect the data to the source system, we collect all the data in its raw format, store it in the data lake and make sure that there is a sort of a history mechanism on the lake itself.

How many years of experience do you have with data platforms?	20+ years
Can you explain what kind of experience you have with data platforms?	I've done several roles. So first creating actually the data platform, so more on the execution side that was creating the data platform, making sure that data was connected to the data platform then storing the data and then distributing the data, so creating well working actually as an ETL developer or as the data platform engineer and later on I was also more in the design activities. So creating data platforms, designing it, advising the customer what kind of data platform they should use.

Question	Transcript	Code	Theme
What governance mechanisms have you observed being used in data platforms to facilitate openness?	I think that you must be sure that what kind of users you are you're connecting to your data platform. So you know who is accessing what kind of data	who is accessing what kind of data	Access
What governance mechanisms have you observed being used in data platforms to facilitate openness?	But you must also be aware of the internal governance. That means that making sure that data that is that is prepared for any external or internal consumption needs to be owned by someone.	data that is that is prepared for any external or internal consumption needs to be owned by someone.	Roles & responsibilities
What governance mechanisms have you observed being used in data platforms to facilitate openness?	But I think that also making sure that the data is always there and always correct. So it means that APIs are the best way to distribute your data. Because your you can change the data underneath it without breaking the connection to your third parties or the people who you are distributing the data.	So it means that APIs are the best way to distribute your data.	Application programming interfaces

Main questions				
Pillar	Governance Mechanisms	Transcript	Code	
Platform boundary resources	Application programming interfaces	<p>So it means that APIs are the best way to distribute your data. Because your you can change the data underneath it without breaking the connection to your third parties or the people who you are distributing the data.</p> <p>Yeah. So I think that working with APIs you can do that on different levels. You can do that with a sort of a token or with user</p>	Yes	Positive

		<p>authentication. Most of the times it's well, anonymous, we always discard that, because that's that. That's not safe.</p> <p>So it means that providing an API to the outside world or to the openers actually, but being, Uh, Flexible enough to to change everything that's underneath it. So actually you, you uh, yeah, you disconnect your system from the way that you want to be exposed in the openness.</p> <p>So I think that APIs are the best way to distribute your data from a data platform into the open.</p>		
	Development boundary resources	<p>Yeah. So what you can do actually you can very easily with software development kits or with with packages you can easily connect to governance tools like for instance purview or Collibra, or an Active Directory or entra and being able to to collect the information, the necessary information to apply the correct data governance to your software development. So I think that you don't have to create the whole government yourself, you can rely on the governance tooling that's already available. And use the SDK's to implement that in your software software application.</p>	Yes	Positive
	Social boundary resources	<p>I think that that, that it's it's crucial to have at least a good documentation.</p> <p>And it can. Yeah, it can. Training can be a good way to make the openness and the governance around it more clear, so it supports your documentation that you create on.</p>	Yes	Positive

Platform rules	Intellectual property rights	<p>My opinion, I think that you should always use as much open source and open source licencing as much as possible, at least from my point of view as a consulting company, because I don't think that well, there there's a there's an extra field or will that that's behind licencing your product because you need to have well you need to have your licencing in place. You need to have a sort of support environment in place that can support your customers and I think that that.</p> <p>That will that prevent you from creating a whole support, uh organisation to be able to to create your openness if you know if you understand what I mean. So I'm in front, I'm in favour of open sourcing it and and using those kind of licences and that everybody can benefit from it and even improve it if necessary.</p>	Yes	Positive
	decision rights	<p>And this kind of user or this kind of access mechanism makes sure that you can provide the access to the to the outside world but not have the extensive user management.</p> <p>So you you always need to have a sort of a monitoring system. That can signal some non performance things on your on your gateway.</p>	Yes	Positive
	gatekeeping	<p>I think that you must be sure that what kind of users you are you're connecting to your data platform. So you know who is accessing what kind of data. But you must also be aware of the internal governance. That means that making sure that data that is that is prepared for any external or internal</p>	Yes	Positive

		consumption needs to be owned by someone. So I think that using. O out mechanisms or token access is the best way to make sure that you can distribute your data but without keeping track of all users. So having a very extensive User management from outside.		
	pricing	So one is that they have a sort of a trusted relation with their with the people who are coming from outside the company and connect to their data platform and they have a sort of a mutual understanding. OK, you can use that data with fair use and without any costs, but they also can monetize their data so. Making different kind of of data sets.	Yes, with constraints	Positive
	revenue-sharing	I don't have an example for this. I don't no, I cannot say anything valuable on this.	-	-
Ecosystem identity	relational control	Don't think that it's that it can be applied for things that that is that is used in our our business. Because we are delivering platforms specifically for a customer that we have and that customer will decide if he wants to have openness in their system, and probably it's more of their own of their own clients that will use the data or the third parties that they they connect to it. So I don't think it's it's for wider public and they that it's created for some kind of an identity	No	Negative

Question	Transcript	Code	Theme
What are data-specific mechanisms used to manage data platform openness?	You can have push mechanisms like reporting that's that's pushed towards customers what	You can have push mechanisms like reporting that's	Data products

<p>How do data-specific mechanisms contribute to data platform openness?</p>	<p>we. What we also have developed is a sort of a dashboard that's publicly available for some kind of medicines that are in a clinical phase and people we've developed that for the [client], it's actually a dashboard where people can check what what clinical phase the medicine is in and and how it's decided by the government to facilitate money in this kind of subsidiary that can be also openness.</p>	<p>that's pushed towards customers</p>	
<p>Is there any other governance mechanism related to platform openness that we haven't discussed yet?</p>	<p>I think that's that's also a big a big topic. So the GDPR is is I think. That well, it's it's it's a subject of on its own, but you can also place it below the the the governance part.</p>	<p>GDPR is a subject of on its own, but you can also place it below the the the governance part.</p>	<p>Privacy</p>
<p>Are there any additional insights or perspectives you'd like to share regarding the mechanisms we have covered?</p>	<p>No, I don't think so.</p>	<p>-</p>	<p>-</p>
<p>Are there any missing elements you would like to add before we conclude this interview?</p>	<p>No, I think that's that.</p>	<p>-</p>	<p>-</p>
<p>Do you have any recommendations or suggestions to improve the openness of data platform based on your expertise?</p>	<p>Well, I I think that that's. I think for the for the stability and the consistency, that openness should also rely on a lot of metadata driven functionalities, so there's always a lot of well within the data platform and all the processes that are running to collect the data to process the data it's it's already based on metadata driven development, but it can also be on on the distribution side on the yeah. So on the openness and I think that that it benefits from well stability is is is a a good a good result.</p>	<p>openness should also rely on a lot of metadata driven functionalities,</p>	<p>Metadata management</p>

Analysis interview 9

Interview number	Int_9
What is your job function?	Data platform lead
Can you describe the platform of the organization you are working for?	So we built a central data platform from IT. So from a central data analytics department within IT. What we are currently facilitating on the platform is making sure that the BI use cases and the BI needs of the organisation can be fulfilled. So currently there's no capability for centrally ingested. There's no there wasn't a central data warehouse yet for example. So we've built a data warehouse component on our data platform to make sure that, yeah, the various business teams first focusing on sort of the usual suspects in BI like the finance department, the HR departments make sure that they have the right management insights into the data that we already have but on the other hand also creating reports on for example, the performance of powerful or fleet.
How many years of experience do you have with data platforms?	4 years
Can you explain what kind of experience you have with data platforms?	And after that I my experience is mainly in building data platforms and building the pipelines that run on the data platforms. Mainly focusing on the Yeah, the, more, the, the the clouds infrastructure, the clouds engineering part of data platforms and the ingestion part of data platform. So really the things which you can apply engineering to more of the software engineering part of data platforming as opposed to the more data modelling side of of data warehousing. For example, on a data platform.

Question	Transcript	Code	Theme
What governance mechanisms have you observed being used in data platforms to facilitate openness?	Well, an API is is an excellent example in these of what I've encountered and built before in data platforms and otherwise governance mechanisms.	API is is an excellent example	Application programming interfaces
What governance mechanisms have you observed being used in data platforms to facilitate openness?	I would say a data sharing agreement is also a governance mechanism for facilitating openness and then data sharing agreements in the both in the term of and actual documents, as well as the technical implementation of or the possible technical applications of a data contract.	I would say a data sharing agreement is also a governance mechanism for facilitating openness	Data sharing agreements

Main questions				
Pillar	Governance Mechanisms	Transcript	Code	
Platform boundary resources	Application programming interfaces	Yes, it allows you to make sure that downstream systems can easily be coupled to your to the to the source in the sense that. If you agree upon an abstracted API and that allows the downstream systems to query that sort of abstracted interface which doesn't, which shouldn't break in case of certain changes in the underlying source system. So in that way it facilitates or it enables. It makes it easier to share data between two systems without tightly coupling those	Yes	Positive

		systems. Just yeah, tightly coupling systems is not always desirable.		
	Development boundary resources	The most used example which I've seen is using open API specifications. It was used to was formerly known as Swagger. And then yeah, those agreeing on or formulating your or documenting your API according to the Open API specification allows you to easily host documentation online. But also allows you to easily generate tests, for example based on yeah, the characteristics of your API. There's multiple ways to implement it, but you can generate an open API specification based on your code. Or you can write an open API specification and you can generate your code based on that specification. And then you can also generate documentation and Generates tests based on that specific specific. And it also makes sure that your API is defined in a with a certain standard, which increases predictability in downstream usage. So people know generally know what to expect when they need to. When they see an open API specification so they know which endpoints to query, like for example, that's putting a post request. To the the the customer endpoints, that's that gives a certain error when you don't include the user ID for example.	Yes	Positive
	Social boundary resources	Well, for sure I think, yeah, just especially training as well as always important or it's always a useful tool. It's not always the most scalable tool, but yeah, it's definitely a mechanism which can enable openness. It's also something which we, yeah, that they mentioned. It's also something that we apply ourselves. Like before, people can use our platform also within the within Company_int_8 they need to have a certain baseline of knowledge.	Yes	Positive
Platform rules	Intellectual property rights	Yeah, without a licence. Or at least making, even though it's sort of contradictory statements	Yes	Positive

		<p>because you're sort of, you're restricting by restricting your API. You can facilitate an openness. So by restricting it behind licences, restricting the usage behind licences, that enables, uh, it's sort of a minimum barrier for organisation to exposing an API so I don't know if that's if that's somewhat clear. Well, they play a role because intellectual property rights should be of course, enforced. Uh. When facilitating openness as in you can't open up your data platform if you expose your uh, yeah, intellectual property through that API or You want to define upfront what is the acceptable amount of intellectual property that you want to expose out of the data platform before opening up your platform.</p>		
	decision rights	<p>Personally, I haven't seen any implementations or Yeah, which make use of in terms of conditions to open up data platforms. I'm sure there are use cases, but I'm not sure whether they are very common of the feeling that openness in data platforms and like opening up your platform.</p>	Possibly	Negative
	gatekeeping	<p>I think access rights are a must in facilitating openness of platforms. Otherwise or you have a fully open data platform with no classified data or whatever on there. Then you don't need the specific access rights. But yeah, access rights should always be implemented based on the classification of the data that you want to expose. Definitely in the data platform you want to make sure that only the people with the right authorization can access certain entities through the API platform.</p>	Yes	Positive
	pricing	<p>I think it's a factor which makes it more attractive for companies to introduce openness in their platforms, If you, if outside parties are able to or, if you as an organisation are able to make money out of your the data that you already own, well, that's it's an easy business case. Uh, usually if you have the proper agreements in place and the proper governance mechanisms such that the data is used properly.</p>	Yes	Positive

	revenue-sharing	It sounds, uh, it sounds pretty nice and, uh, promising model. But I haven't seen it yet.	Possibly	Negative
Ecosystem identity	relational control	I'm not exactly sure how that's how this facilitates openness. I can imagine that if there's, like on a relationship basis. If there's a lot of trust in or in the developers, and like in the certain stakeholders and users of the platform that you are more eager to facilitate openness in your platform. But if people are generally very protective of data, and if the culture is very is very siloed, and if people are very protective of data, then that's automatically will be reflected in your data platform as well.	Possibly	Negative

Question	Transcript	Code	Theme
<p>What are data-specific mechanisms used to manage data platform openness?</p> <p>How do data-specific mechanisms contribute to data platform openness?</p>	<p>Separating your open data from the storage of the of your more confidential data. I think that's easy way to, or at least it's the way I see how people implement openness in platform, so always separating the or as on the infrastructure level and the data level separating the aspects which can be shared with external parties, always separating that from the things that should be kept internal to keep risks at a minimum. Well by for example, hosting a separate database for data which is just classified as uh as publicly available.</p>	<p>Separating your open data from the storage of the of your more confidential data.</p>	<p>Data separation</p>
<p>Is there any other governance mechanism related to platform openness that we haven't discussed yet?</p>	<p>I don't think so. I think there were a lot of aspects covered. I think openness in data platforms is still a relatively new topic and it's still being discovered.</p>	-	-
<p>Are there any additional insights or perspectives you'd like to add about the mechanisms that we already covered?</p>	<p>No, I don't think so.</p>	-	-
<p>Are there any missing elements you would like to add before we conclude this interview?</p>	<p>No</p>	-	-
<p>Do you have any recommendations or suggestions to improve the openness of data platform based on your expertise?</p>	<p>It requires quite a culture shift to enable more openness in in data platforms, and it's up to us as data specialists to educate the rest of the world, also on the value of data and also when there's value in sharing data, then also yeah. Educate people on the value of sharing data. Like currently there's a lot of data still</p>	<p>Educate people on the value of sharing data.</p>	<p>Education</p>

	hidden behind closed doors or like hidden in in different silos.		
--	--	--	--

Analysis interview 10

Interview number	Int_10
What is your job function?	Head of tech data platforms
Can you describe the platform of the organization you are working for?	It is basically where you can propagate data through the different layers, right? So the the, the, the, the medallion architecture. Let's say that's what we have chosen where you can take it from source to a clean layer or a goal layer where you're preparing data products. So how do you get your data from the different sources and make it available for consumption is the trick of the data platform.
How many years of experience do you have with data platforms?	14 years
Can you explain what kind of experience you have with data platforms?	So I've done everything from building it to migrating it or running a team that's building it. I've been an architect, I've been a solution architect or a product architect as you would call it, because you're building across the thing. I have been a enterprise, integration and data architect, so that was more on the policies and principles around it. Of course I I didn't start there, but I started as a engineer who was actually doing data warehouse migration or let's say SQL Building SQL Up to that level, and also being a what you call a manager or a director or whatever, now who's maintaining or managing a team that is building it.

Question	Transcript	Code	Theme
What governance mechanisms have you observed being used in data platforms to facilitate openness?	DAMA DMBOK, From architecture to including the modeling, security, uh, metadata quality, all of those aspects. But what is important, there is 2 things. One is onboarding people onto the platform, right? So one thing is make sure everybody's on that platform, which means that at least you have governance or oversight over the data. And the next thing is then you entice the consumers, so it's producers and consumers. So producer will be able to entice your you know you build a platform that people are able to put data on and once you have the data then consumers also come to you.	DAMA DMBOK, all of the aspects	DAMA DMBOK

Main questions				
Pillar	Governance Mechanisms	Transcript	Code	
Platform boundary resources	Application programming interfaces	Now there are 2 ways of looking at APIs. One is in the in the the the the outside circle which is the incoming and outgoing data, incoming is where	Yes, with constraints	Positive

		<p>people with who have a possibility of sending out data broadcasting data are able to easily onboard data onto the platform. So the the producers are able to just send it. And and the same thing on the other side we have we have APIs being served where anybody can access it and get you know how many bananas we're selling in the supermarket or something like.</p>		
	Development boundary resources	<p>So how do you get your data from the different sources and make it available for consumption is the trick of the data platform. And how you do this for not only one company, but let's say 14 or 15 brands is the is the tricky part. But yeah, that's what we do in terms of a platform.</p> <p>But this tool is actually built on backstage, where we have a software development kit or a software development platform, and SDP which is built by our colleagues here and what they do is they allow for good dev ex experience while at the same time keeping it within control.</p>	Yes	Positive
	Social boundary resources	<p>But what is important, there is 2 things. One is onboarding people onto the platform, right?</p> <p>A big time, but that's only within a prescribed audience. Why do I say that? For example, all our documentation is on GitHub.</p>	Yes	Positive
Platform rules	Intellectual property rights	<p>So within our organization, we have a open policy in terms of our GitHub repository, our repository repository. The product as such is known to only let's say it's inner source we have. We have a lot of people inner sourcing. But</p>	Yes	Positive

		that means that it is also open repository, so we don't have any restriction there.		
	decision rights	If you're going to the supermarket, you can take a Volkswagen Polo. You don't need to take a Ferrari. There's no such thing as free Ferrari. But don't don't, don't give a Ferrari when they just want to drive it to the supermarket and back, right? It depends on horses for courses. It depends on what people need.	Yes	Positive
	gatekeeping	<p>99% are within the organization, but we are building the platform within Company_int_10. So technically it's from Company_int_10, but let's say we have different solution zones for anybody who wants to consume it, which means then across, Company_int_10, people consume it.</p> <p>No, we have very we have very strict DPIA 's in place. Uh, so we follow the demo. So for example, we don't share data even within the departments Company_int_10 I told already told you Company_int_10, Sister_Company_int_10 Sister_Company_int_10 and we use HR. My platform is also used by the HR team for a data lake also used by financials for data like but none of us have access to it.</p>	Yes	Positive
	pricing	We cross charge a bit of the usage and the time. But we don't charge for the product because our consumers are within within our colleagues.	Yes, with constraints	Positive
	revenue-sharing	Uh, but the point is that you need to have this synchronicity or mutually	Yes, with constraints	Positive

		beneficial pricing mechanism.		
Ecosystem identity	relational control	Now, do you blame the person who's using it? Or do you blame the platform for allowing it to give access? I would blame the person using it and the person using it will say no. It's your fault Pritam because you have made me [...] They are linked by the data, so it's very important to have the mechanism in place, yes.	Yes	Positive

What are data-specific mechanisms used to manage data platform openness? How do data-specific mechanisms contribute to data platform openness?	I think I think when you talk about data specific mechanisms, I mean, no, no, not really.	-	-
Is there any other governance mechanism related to platform openness that we haven't discussed yet?	No, I don't think so.	-	-
Are there any additional insights or perspective you like to share regarding the mechanisms that we have covered?	No, I think you ask all the questions. Maybe availability is one aspect of data governance that we have not touched upon and that means to always make data available, whether you're a producer or a consumer, right? Making sure they're on that. We spoke about it briefly in the beginning where I said, you know, you need to adopt them to get them on the platform and then consumers will come.	Maybe availability is one aspect of data governance that we have not touched upon	Availability
Are there any missing elements you would like to add before we conclude this interview?	No, I think we discussed everything.	-	-
Do you have any recommendations or suggestions to improve the openness of data platform based on your expertise?	Hmm, I need another few hours for that. No, I think apart from what we spoke about, what I spoke about was a lot of learning from experiences where we have either I've done it myself and, you know, tried and failed and then succeeded, or I've seen it happen	-	-

