

# MASTER'S THESIS

## Succesvol managen van data science projecten

Kali, A.R. (Aniel)

**Award date:**  
2020

[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](mailto:pure-support@ou.nl)

providing details and we will investigate your claim.

Downloaded from <https://research.ou.nl/> on date: 25. Mar. 2023

**Open Universiteit**  
[www.ou.nl](http://www.ou.nl)



# Succesvol managen van data science projecten

---

## Successful management of data science projects

|                     |  |
|---------------------|--|
| Opleiding:          | Open Universiteit, faculteit Management, Science & Technology<br>Masteropleiding Business Process Management & IT          |
| Programme:          | Open University of the Netherlands, faculty of Management, Science & Technology<br>Master Business Process Management & IT |
| Cursus:             | IM0602 Voorbereiden Afstuderen BPMIT<br>IM9806 Afstudeeropdracht Business Process Management and IT                        |
| Student:            | ANIEL ROYTAAS KALI   |
| Identiteitsnummer:  |  |
| Datum:              | 9 juli 2020  |
| Afstudeerbegeleider | PROF.DR.IR. REMKO (R.W.) HELMS   |
| Meelezer            | DR. IR. HARRY (H.H.) MARTIN  |
| Versie nummer:      | 1.2  |
| Status:             | Definitief   |

## Abstract

Verschillende data science (DS) projecten falen, omdat er bij het managen van DS projecten te weinig aandacht is voor de projectmatige en organisatorische aspecten. Het is belangrijk dat organisaties die DS projecten uitvoeren bekend zijn met de kritieke succesfactoren (KSF) van een DS project. KSF beschrijven immers wat een organisatie nodig heeft om DS projecten succesvol te managen.

Het doel van dit onderzoek is om te achterhalen wat nu deze KSF zijn en daarbij ook extra aandacht te schenken aan de projectmatige en organisatorische aspecten bij DS projecten. Hiervoor is de volgende onderzoeksvraag opgesteld: Hoe kan een DS project uitgevoerd worden, met meer aandacht voor projectmatige en organisatorische aspecten, zodat zij succesvol is? Om een antwoord te kunnen geven op de onderzoeksvraag is enerzijds een literatuuronderzoek gedaan naar KSF, anderzijds is er een casestudy gedaan naar de ervaringen die organisaties hebben met het managen van DS projecten. Dit is gedaan door personen die een belangrijke rol hebben bij het managen van DS organisaties, te bevragen door middel van semigestructureerde interviews en deze data kwalitatief te analyseren.

Organisaties zijn vergeleken door te kijken naar verschillen en overeenkomsten en er is ook gekeken in hoeverre kritieke succesfactoren vanuit de literatuur, die betrekking hebben op niet Nederlandse organisaties, ook voorkwamen bij de geanalyseerde Nederlandse organisaties. Tot slot is er ook gekeken naar de samenhang van verschillende KSF bij organisaties.

## Sleutelbegrippen

Managen data science projecten, casestudy, kritieke succesfactoren, process management methodologie

## Samenvatting

Verschillende data science (DS) projecten falen, doordat er bij het managen van DS projecten te weinig aandacht is voor de projectmatige en organisatorische aspecten. Dit onderzoek is op zoek naar een zo compleet mogelijk beeld van oorzaken, die ervoor kunnen zorgen dat een DS project faalt of slaagt. Hiervoor is er een literatuuronderzoek gedaan naar kritieke succesfactoren (KSF) en is een casestudy gedaan naar de ervaringen van Nederlandse organisaties met DS projecten.

Uit het literatuuronderzoek komen een heleboel KSF naar boven op basis van organisaties, die geen Nederlandse organisaties betreffen. Uit de casestudy zijn vervolgens ook uitdagingen en successen bij Nederlandse organisaties aan het licht gekomen. Deze factoren zijn enerzijds onderling vergeleken, anderzijds ook vergeleken met de KSF die uit het literatuuronderzoek zijn gekomen.

Het is opvallend om te zien, dat veel van de KSF uit het literatuuronderzoek ook een rol spelen bij Nederlandse organisaties, maar er zijn wel degelijk extra KSF gevonden aan de hand van de geanalyseerde organisaties. Daarnaast is het opvallend om te zien, dat de organisaties die onder de loep zijn genomen, veel overeenkomsten vertonen. Zo is een belangrijk resultaat, dat de bedrijfscultuur een belangrijk aspect is voor het succes van een DS project. Dit is dus een kritieke succesfactor op het organisatorische vlak. Deze kritieke succes factor is bovendien het meest aan de orde gekomen in de antwoorden van de verschillende geïnterviewden. Een andere reden dat dit thema veel aan de orde is gekomen, is vanwege het feit dat deze kritieke succesfactor een sterke samenhang vertoont met een tal van andere KSF.

Een ander resultaat is dat gebruikte proces management methodologieën nog aangepast kunnen worden en meer rekening kunnen houden met de data scientisten en bijvoorbeeld het verkennende karakter van data science. Echter ondanks de bevestiging dat er veel CSF een rol spelen op het projectmatige en organisatorische vlak, is uit dit onderzoek ook gekomen dat CSF in het technische spectrum ook nog steeds een aandachtspunt zijn. In het bijzonder valt op dat organisaties moeite hebben met het toepassen van nieuwe technieken die een rol spelen bij de implementatie van DS en big data oplossingen. De implementatie fase in het DS project is bij alle onderzochte organisaties een aandachtspunt. Momenteel duurt de implementatie gemiddeld 3 maanden (voor alle organisaties) voor de langere projecten. Tot slot kan de kennis die uit dit onderzoek tot stand is gekomen, gebruikt worden om eigen DS projecten te beoordelen en te verbeteren.

## Summary

Several data science (DS) projects fail because data management projects pay too little attention to project and organizational aspects. This research is looking for the most possible complete list of causes that can cause a DS project to fail or succeed. To this end, a literature study was conducted into critical success factors (CSF) and a case study was conducted based on the experiences of Dutch organizations with DS projects.

The literature search reveals a lot of CSF based on organizations that do not concern Dutch organizations. This case study subsequently also revealed challenges and successes at Dutch organizations. On the one hand, these factors have been compared with one another, and on the other hand they have also been compared with the CSF that have emerged from the literature study.

It is striking to see that many of the CSF from the literature study also play a role in Dutch organizations, but additional CSF have indeed been found on the basis of the organizations analyzed. It is also striking to see that the organizations that have been analysed have many similarities. An important result, for example, is that the corporate culture is an important aspect for the success of a DS project. This is therefore a critical organizational success factor. Moreover, this critical success factor was most discussed in the answers of the various interviewees. Another reason that this topic has been discussed a lot is due to the fact that this critical success factor has a strong connection with many other CSF. Another result is that used process management methodologies can be adapted and take more account of the data scientists and, for example, the exploratory nature of data science.

However, despite the confirmation that many CSF play a role at the project and organizational level, this study also revealed that CSF in the technical spectrum are also still a point for attention. In particular, it is striking that organizations have difficulty applying new techniques that play a role in the implementation of DS and big data solutions. The implementation phase in the DS project is a point of attention for all researched organizations. Implementation currently takes an average of 3 months (for all organizations) for longer projects. Finally, the knowledge gained from this research can be used to assess and improve own DS projects.

# Inhoudsopgave

|   |    |
|---|----|
| Abstract .....  | 2  |
| Sleutelbegrippen.....   | 2  |
| Samenvatting.....   | 2  |
| Inhoudsopgave .....   | 4  |
| 1. Inleiding.....   | 1  |
| 1.1. Aanleiding .....   | 1  |
| 1.2. Probleemstelling.....  | 1  |
| 1.3. Onderzoeksvragen.....  | 2  |
| 1.4. Relevantie .....   | 2  |
| 1.5. Onderzoeksaanpak .....   | 3  |
| 2. Theoretisch kader .....  | 3  |
| 2.1. Onderzoeksaanpak .....   | 3  |
| 2.2. Uitvoering .....   | 5  |
| 2.3. Resultaten en conclusies.....  | 6  |
| 2.4. Doel van het vervolgonderzoek .....                                    | 9  |
| 3. Methodologie .....   | 9  |
| 3.1. Conceptueel ontwerp: keuze van onderzoeksmethode(n) .....              | 9  |
| 3.2. Technisch ontwerp: uitwerking van de methode.....                      | 10 |
| 3.3. Gegevensanalyse.....   | 11 |
| 3.4. Reflectie t.a.v. validiteit, betrouwbaarheid en ethische aspecten..... | 11 |
| 4. Resultaten.....  | 13 |
| 4.1. Beschrijving uitvoering onderzoek .....                                | 13 |
| 4.2. Algemene resultaten .....  | 14 |
| 4.3. Profiel en resultaten organisatie 1 .....                              | 15 |
| 4.4. Profiel en resultaten organisatie 2 .....                              | 19 |
| 4.5. Profiel en resultaten organisatie 3 .....                              | 24 |
| 4.6. Resultaten cross-case analyse.....                                     | 28 |
| 5. Discussie, conclusies en aanbevelingen .....                             | 31 |
| 5.1. Discussie – reflectie .....  | 31 |
| 5.2. Conclusies .....   | 34 |
| 5.3. Aanbevelingen voor de praktijk .....                                   | 36 |
| 5.4. Aanbevelingen voor verder onderzoek .....                              | 36 |
| Literatuurlijst .....   | 37 |

|   |    |
|---|----|
| Bijlage 1: Artikelen gevonden in het literatuuronderzoek.....     | 39 |
| Bijlage 2: Synthese van kritieke succesfactoren .....             | 43 |
| Bijlage 3: Interview protocol .....                               | 46 |
| Bijlage 4: Lijst codes .....                                      | 49 |
| Bijlage 5: Groepering van de codes per organisatie .....          | 51 |
| Bijlage 6: Groepering van de codes per functie.....               | 52 |
| Bijlage 7: Matrix met combinaties van subcodes met aantallen..... | 53 |

# 1. Inleiding

## 1.1. Aanleiding

In de digitale wereld van tegenwoordig, waarin met de dag meer (big) data verzameld wordt door bedrijven, is het de kunst om informatie uit deze stroom van data te halen. Dit kan onder andere met behulp van data science (DS). Zo kan DS als tool gebruikt worden voor een tal van zaken, bijvoorbeeld om kosten te verlagen, om nieuwe markten te vinden, om patronen te herkennen, om aanbevelingen te doen, voor het detecteren van afwijkingen, het optimaliseren van supply chain activiteiten, het voorspellen van bepaalde gebeurtenissen en dit is nog maar een tipje van de sluier. Voorbeelden van concrete toepassingen zijn foto herkenning, fraudevoorspelling, klantprofilering, identificatie van tumoren, zelfrijdende auto's, om maar enkelen te noemen, het internet staat er vol mee (Analytics Vidhya, 2015; Priyadarshiny, 2019). Een DS project kan dan ook beschreven worden als een project dat statistische en machine-learning technieken (algoritmes) gebruikt om grote hoeveelheden ongestructureerde en / of gestructureerde gegevens te analyseren, met als doel correlaties en causale relaties te identificeren, gebeurtenissen te classificeren en te voorspellen, patronen en anomalieën te identificeren en waarschijnlijkheden, interesses en sentiment af te leiden (Das, Cui, Campbell, Agrawal, & Ramnath, 2015). Uit onderzoek is gebleken dat bedrijven die aan DS doen, 5 tot 6 % productiever en winstgevender zijn (McAfee, Brynjolfsson, Davenport, Patil, & Barton, 2012). Ondanks dit gegeven blijkt echter ook dat 85% van de DS projecten falen (Walker, 2017). Hoewel de termen big data projecten en DS projecten regelmatig door elkaar gebruikt worden, wordt big data vaak gezien als een subset van DS, waar de hoeveelheid gegevens het gebruik van speciale hulpmiddelen en algoritmen vereist (Saltz, J., Shamshurin, I., & Connors, C. (2017). Verder als er gesproken wordt over big data, worden vaak de volgende dimensies (of een subset daarvan) aangehaald, specifieke kenmerken van big data waarmee omgegaan moet kunnen worden (vaak uitdagingen genoemd), bijvoorbeeld de 7 V's, namelijk volume, verscheidenheid ('variety'), waarheidsgetrouwheid / kwaliteit ('veracity'), snelheid ('velocity'), veranderlijkheid ('variability'), visualisatie ('visualization') en waarde ('value') (Sivarajah, Kamal, Irani, & Weerakkody, 2017).

## 1.2. Probleemstelling

Terwijl er vele verschillende oorzaken kunnen zijn voor het feit dat DS projecten falen, valt het op, dat er binnen de DS community veel nadruk gelegd is op technische oorzaken, zoals het verbeteren van beschikbare technieken om data te analyseren. Er is echter tot nu toe relatief weinig aandacht geweest is voor bijvoorbeeld de samenwerking in een DS project en voor de uitdagingen die een team heeft, die gezamenlijk een DS project uitvoert (Saltz, Crowston, & Shamshurin, 2017).

In een ander recent onderzoek gebaseerd op geclusterd commentaar van 19 experts uit de DS branche (Becker, 2017) kwam ook al naar voren dat de meeste redenen dat een DS project faalde meestal niet technisch van aard waren, maar vaak juist projectmatig en organisatorisch van aard waren. Hierdoor lijdt de industrie onnodig aan lage succespercentages op het gebied van DS, dit terwijl er volgens de jaarlijkse surveys, uitgevoerd door New Vantage Partners (2016), steeds meer door bedrijven geïnvesteerd wordt in DS projecten. Omdat er zo onnodig veel DS projecten falen, wordt niet het maximale uit deze investeringen gehaald.

Er blijkt inderdaad ook nog geen optimale procesmethodologie te bestaan voor de uitvoering van een DS project (Saltz et al., 2017). Veel DS projecten worden op ad hoc wijze uitgevoerd, of op basis van project management, dat niet bevredigend genoeg aansluit bij een DS project (Mariscal, Marbán, & Fernández, 2010). Dit terwijl bekend is dat het toepassen van een verbeterde procesmethodologie, de resultaten van DS teams zou kunnen verbeteren (Saltz, Hotz, Wild, & Stirling, 2018). Het succes van welk project dan ook, is afhankelijk van de effectieve uitvoering van een project. Uit (Saltz, 2015) blijkt immers dat nuttige resultaten een uitdaging zijn, zonder goede managementmethodologieën en -instrumenten. Opvallend vaak falen data DS projecten en hier zijn verschillende redenen voor, niet alleen technische redenen, maar vaak organisatorische redenen.

Probleemstelling: Lage succespercentages bij data science projecten, doordat er bij het managen van data science projecten te weinig aandacht is voor de projectmatige en organisatorische aspecten

### 1.3. Onderzoeksvragen

Het doel van dit onderzoek is om inzicht te verkrijgen in de verschillende zaken waar organisaties tegen aanlopen bij het uitvoeren van DS projecten. Bepaalde zaken gaan goed, maar sommige zaken hebben wellicht meer aandacht nodig.

Tevens zal er een categorisatie gemaakt worden van de verschillende kritieke succesfactoren (KSF) bij DS projecten. Deze inzichten kunnen vervolgens gebruikt worden om de lage succespercentages bij DS projecten te verhogen.

Om de doelstelling, die hierboven beschreven staat te realiseren, moet bepaalde kennis vergaard worden. Daarom is een centrale onderzoeksvraag opgesteld die in dit onderzoek centraal staat:

*Hoe kan een DS project uitgevoerd worden, met meer aandacht voor projectmatige en organisatorische aspecten, zodat zij succesvol is?*

Een uitgevoerd DS project wordt overigens als succesvol gezien, als het resultaat van het DS project, waarde heeft toegevoegd heeft aan de organisatie. Waarde hoeft hierbij niet per se te slaan op monetaire inkomsten.

Om antwoord te kunnen geven op deze hoofdvraag zijn 3 deelvragen opgesteld:

1. *Wat zijn de kritieke succesfactoren van data science projecten?*
2. *Welke verschillende categorieën succesfactoren kunnen onderscheiden worden bij data science projecten?*
3. *Hoe kan de kennis van kritieke succesfactoren ingezet worden, zodat data science projecten beter gemanaged worden?*

### 1.4. Relevantie

#### **Wetenschappelijke relevantie**

Aangezien er in de wetenschappelijke literatuur relatief weinig geschreven is over methodologieën, tools en frameworks waarmee teams effectiever en efficiënter DS projecten kunnen uitvoeren (Saltz et al., 2017), draagt dit onderzoek bij aan project management voor DS. Dat hier behoefte aan is, blijkt wel uit het feit dat er al enkele DS projectmanagementbenaderingen opgekomen zijn, die Agile project management concepten toepassen op CRISP-DM, zoals bijvoorbeeld het Microsoft's Team DS Process en de Domino DS Lifecycle (Emerging Approaches – DS Project Management, 2020). Agile



project management is een framework dat een projectteam beter in staat stelt om snel projecten aan te passen aan snel evoluerende omgevingen waarin systemen worden gebruikt, dit door middel van een iteratief proces van continue evaluatie en korte ontwerptermijnen (Cervone, 2012). Door te analyseren welke elementen uit eventueel toegepaste project management methodes werken of minder goed werken, kunnen deze bevindingen gebruikt worden voor de optimalisatie van project management methodes of leiden tot het gebruik van geheel andere project management methodes.

### **Maatschappelijke relevantie**

Verkregen inzichten kunnen gebruikt worden door organisaties om hogere succespercentages te behalen bij DS projecten. Deze projecten kunnen de gehele maatschappij helpen, zowel personen als organisaties. Zo kunnen organisaties waarde creëren, dit kan voor de organisatie zelf zijn, maar bijvoorbeeld ook voor personen, denk bijvoorbeeld aan een DS project in het kader van het detecteren van een ziekte, of een DS project in kader van het milieu.

## 1.5. Onderzoeksaanpak

In eerste instantie zullen er op basis van de probleemstelling door middel van een verkennend literatuuronderzoek onderzoeksvragen opgesteld worden. Vervolgens zal er kwalitatief onderzoek gedaan worden naar het managen van DS projecten. Dit onderzoek sluit aan bij de verkenning van een onderwerp en vanwege het feit dat er op zoek gegaan wordt naar rijke beschrijvingen van hoe DS projecten gemanaged worden bij organisaties. Een manier om kwalitatief onderzoek te doen is door middel van casestudy. Voor de casestudy zullen 3 verschillende organisaties die DS projecten bedrijven, elk binnen hun eigen context, uitvoerig bestudeerd worden. Vervolgens zullen de 3 organisaties vergeleken worden met elkaar en worden ook bevindingen uit het literatuuronderzoek (conform Okoli & Schabram (2010)) rondom bijvoorbeeld KSF hierin meegenomen. Informatie zal dus door middel van literatuuronderzoek maar ook via semigestructureerde interviews bij organisaties vergaard worden. Deze interviews zullen vervolgens uitgeschreven worden. Er zullen zogenaamde transcripten van gemaakt worden, dit kan eventueel ook door software of partijen die dit als dienst verlenen. Vervolgens kan er structuur in de verzamelde data gemaakt worden door ofwel handmatig te coderen of wel dit met behulp van een software pakket te doen. Een software pakket geeft gelijk de mogelijkheid om deze data gelijk te analyseren, hiermee zullen dan de onderzoeksvragen beantwoorden. Tot slot zal het onderzoek grondig geëvalueerd worden.

## 2. Theoretisch kader

### 2.1. Onderzoeksaanpak

Het doel van dit hoofdstuk is om het theoretische kader te schetsen. Het theoretische kader dient als theoretische basis voor het verdere onderzoek. In dit theoretische kader zal door synthese van informatie een uiteenzetting gegeven worden van de belangrijkste begrippen en wetenschappelijke bevindingen rondom het succesvol managen van een DS project, dit door middel van een literatuuronderzoek. Verder zullen met de bevindingen uit het literatuuronderzoek zoveel mogelijk de onderzoeksvragen beantwoord worden voor zover mogelijk en zal bepaald worden welke zaken er nog extra of uitvoeriger onderzocht moeten worden, bijvoorbeeld vanwege hiaten in de wetenschappelijke literatuur. Het literatuuronderzoek dat gedaan is, bestond uit een groot aantal stappen die ook bekend zijn bij een systematisch literatuuronderzoek. Een systematisch literatuuronderzoek is volgens Okoli & Schabram (2010): “explicit, and reproducible method for identifying, evaluating, and synthesizing the existing body of completed and recorded work produced by researchers”. Hoewel hier dus niet een systematisch literatuuronderzoek uitgevoerd wordt, kan veel van dezelfde methodologie gehanteerd worden bij dit literatuuronderzoek, dat als

theoretische basis dient voor deze scriptie. Nu volgt een toelichting op de uitgevoerde stappen van dit literatuuronderzoek op basis van Okoli & Schabram (2010).

### **Stappen methodologie literatuuronderzoek**

1. Zoeken van relevante literatuur:  
Gelet op de onderzoeksvragen zijn er zoektermen (zie tabel voor specificatie van de zoektermen) bedacht waarmee relevante wetenschappelijke literatuur gevonden kan worden. De zoektermen hebben te maken met het succesvol managen van DS projecten en verschillen enigszins afhankelijk van welke bron gebruikt is. De bronnen die gebruikt zijn, zijn de universiteitsbibliotheek van de Open Universiteit, Google Scholar en Google. Google voor bijvoorbeeld eventuele interessante websites of blogs.
2. Practical screen (uitsluiten van literatuurbronnen):
  1. Van elk artikel is de "abstract" en samenvatting gelezen om de relevantie te toetsen, alvorens het artikel op te nemen in de literatuurlijst;
  2. Er is alleen geput uit bronnen die opgenomen zijn binnen journals en conferences, die ook "peer-reviewed" zijn;
    - Universiteitsbibliotheek : Om het aantal artikelen beperkt te houden, zijn alleen peer-reviews opgenomen, tevens zijn krantenartikelen/book-reviews/proefschriften uitgesloten. Er resteren zo alleen artikelen die geschreven zijn door experts en beoordeeld zijn door verschillende andere experts in het veld, voordat het artikel wordt gepubliceerd (om de kwaliteit van het artikel te waarborgen).
    - Google Scholar: Patenten zijn uitgesloten, omdat die niet relevant zijn. Citaten zijn verder ook uitgesloten, opdat er alleen artikelen gevonden werden.
    - Google: Er zijn slechts 30 relevante resultaten weergegeven, omdat ervoor gekozen is om de duplicate content (verschillende websites met dezelfde inhoud), weg te laten, dit om de meest relevante resultaten te tonen en het aantal resultaten beperkt te houden.
3. Quality appraisal:  
In de kwaliteitsbeoordeling zijn documenten nauwkeurig bestudeerd, om te bepalen of er antwoorden op onderzoeksvragen in de artikelen gevonden kan worden. Wanneer dit het geval was, is het artikel meegenomen. Om de kwaliteit verder te borgen van de artikelen, is onderzocht of de onderzoeken in de artikelen ook verantwoord zijn uitgevoerd. Er is hierbij vooral naar de methodologie sectie van het paper gekeken, of het onderzoek dus methodologisch wetenschappelijk verantwoord is geweest.
4. Data extractie:  
Van elk gekozen artikel is een kleine samenvatting gemaakt en op basis van alle samenvattingen is er 1 grote mindmap gemaakt, dat als hulpmiddel dient bij de analyse van de bevindingen.
5. Analyse van bevindingen:  
Op basis van de mindmap in stap 4, is de informatie uit de gebruikte artikelen bekeken in relatie tot de probleemstelling en onderzoeksvragen. De mindmap heeft vervolgens geholpen om te komen tot een samensmelting (synthese) van de informatie uit alle artikelen.
6. Schrijven van review:  
Op basis van de analyse in stap 5 is beschreven wat er bekend is in de literatuur met betrekking tot de probleemstelling. Tot slot is bepaald in hoeverre de literatuur ondersteunend was voor het beantwoorden van de onderzoeksvragen en zijn de

onderzoeksvragen beantwoord voor zover mogelijk op basis van de gevonden relevante literatuur.

## 2.2. Uitvoering

Het literatuuronderzoek is gedaan op basis van 3 bronnen, namelijk op basis van de digitale universiteitsbibliotheek van de Open Universiteit, Google Scholar en Google. Google is aanvullend gebruikt om eventuele interessante websites of blogs te vinden rondom proces management van DS / big data projecten, daar dit een relatief nieuw onderwerp is, waarover nog niet zoveel geschreven is. Ondanks dat dit dus geen wetenschappelijke literatuur hoeft te zijn, kunnen er mogelijk toch interessante websites gevonden worden, waar het onderwerp proces management van DS / big data projecten, door deskundigen in het werkveld besproken wordt. Overigens zijn de gebruikte zoektermen in de 'queries' door de zoekmethodes van Google Scholar en de universiteitsbibliotheek niet helemaal hetzelfde, omdat de zoekmethoden niet helemaal hetzelfde werken en ook niet tot dezelfde orde van resultaten leiden. Het aantal artikelen die gevonden zijn op basis van specifieke zoektermen kan hieronder bekeken worden in tabel 2.1. De 'practical screen', besproken in het vorige hoofdstuk, is gedaan op basis van de getoonde filters.

| Bron   | Zoektermen  | Filter  |
|--|---|---|
| Universiteitsbibliotheek (Open Universiteit) | (TitleCombined:(“DS” OR “DS team” OR “DS projects” OR “data scientists” OR “big data projects” OR “big data” OR “CRISP?DM”)) AND (TitleCombined:(manag* OR “project management” OR “process management” OR method* OR “process model” OR project* OR process* OR guide OR framework OR tool* OR framework OR holistic)) AND (TitleCombined:(success* OR fail* OR challeng* OR key?factors OR effectiveness OR performance)) | Alleen peer-reviews zijn opgenomen, tevens zijn krantenartikelen/book-reviews/proefschriften uitgesloten. |
| Google Scholar                               | allintitle: "DS project" OR "big data project" OR "DS factors" OR "big data factors" OR "DS succes*" OR "big data succes*"  | Patenten en citaten zijn uitgesloten.   |
| Google                                       | "DS project management" "big data project"  | Gebruik gemaakt van de optie in Google om alleen de meest relevante resultaten weer te geven.             |

Tabel 2.1: zoektermen literatuuronderzoek

Voor elk van de in tabel 2.1 genoemde bronnen kan in bijlage 1: Artikelen gevonden in het literatuuronderzoek worden opgezocht hoeveel resultaten er waren na het draaien van de 'queries', hoeveel artikelen er relevant genoeg bleken te zijn en hoeveel artikelen er uiteindelijk gebruikt zijn na een finale kwaliteitscheck. Tot slot zijn ook de artikelen zelf die gevonden zijn, opgenomen in bijlage 1: Artikelen gevonden in het literatuuronderzoek.

## 2.3. Resultaten en conclusies

Om een DS project te laten slagen is het van belang inzicht te hebben in de KSF van een DS project. Het gebruik van KSF is een gebruikelijke methode om de prestaties van een organisatie te meten (Rockart, 1979; Bullen & Rockart, 1981; Rockart & Crescenzi, 1984). Rockart (1979) definieerde KSF als: "*the limited number of areas in which results, if they are satisfactory, will ensure successful competitive performance for the organization. They are the few key areas where 'things must go right' for the business to flourish*".

In de literatuur worden een tal van KSF benoemd. Vaak worden deze succesfactoren onderverdeeld in: mensen, processen en technologie (Koronios et al., 2014; Mikalef, Boura, Lekakos, & Krogstie, 2019; Miller, 2018; Saltz & Shamshurin, 2016).

Er zijn echter ook nog gedetailleerdere onderverdelingen te vinden zoals de onderverdeling in:

- Data, processen, technologie, organisatie, mensen en context (Mikalef et al., 2019);
- Data, processen, tools, governance, team en meetbare doelen (Saltz & Shamshurin, 2016).

Deze onderverdelingen lijken eigenlijk erg op elkaar en zijn daarom, op de context na, ook wel uitwisselbaar. Zo bevatten beide onderverdelingen data en processen. Daarnaast zullen tools vaak onder technologie vallen en slaat het team op de mensen. Verder zijn 'governance' en meetbare doelen onderdeel van processen. De reden dat deze apart genoemd worden door Saltz & Shamshurin (2016), geeft wel aan dat zij vinden dat deze zaken bijzondere aandacht vereisen. Tot slot geeft de context aan, dat naast KSF, ook nog de context van een organisatie er toe doet, zoals bijvoorbeeld de organisatiegrootte, branche, ofwel de specifieke situatie waarin een organisatie zich bevindt.

Naast de categorisatie van KSF, worden KSF door Koronios et al. (2014) ook gekoppeld aan de verschillende fases die een rol spelen in een DS project, zoals de de 'business'-fase, de data-fase, de analyse-fase, de implementatie-fase, de meet-fase en de leer-fase. Bepaalde KSF lijken dus in een bepaalde specifieke fase meer of minder een rol te spelen.

Een aantal vaak voorkomende voorbeelden van KSF, aangehaald door meerdere onderzoekers (Gao, Koronios, & Selle, 2015; Koronios et al., 2014; LaValle, Lesser, Shockley, Hopkins, & Kruschwitz, 2011; Mikalef, Boura, Lekakos, & Krogstie, 2019; Miller, 2018; Saltz & Shamshurin, 2016; Watson, 2014, 2019) zijn op het laagste niveau (op het niveau van mensen, processen en technologie):

- Mensen: analytische vaardigheden, technologische kennis, multidisciplinair team, coördinatie stakeholders, een op feiten gebaseerde besluitvormingscultuur;
- Processen: waarde creatie, (big) data strategie, datakwaliteit, meetbare project doelen, projectplanning, sponsoring van hoger management, 'privacy by design', goed gedefinieerde rollen en verantwoordelijkheden, proces management als vaardigheid, goede afstemming tussen bedrijfs- en IT-strategieën;
- Technologie: opslag van data, prestaties van systemen, een flexibele (schaalbare) IT-architectuur, investeringen in nieuwe technologieën, toegang tot data, uitdagingen bij gebruik van big data (Sivarajah et al., 2017; Vanauer, Böhle, & Hellingrath, 2015)

Zie voor een volledige lijst van gevonden KSF bijlage 2: Synthese van kritieke succesfactoren. Hier zijn alle gevonden KSF gevonden in het literatuuronderzoek gecategoriseerd onder data, processen, technologie, organisatie, mensen (Mikalef et al., 2019). Er had ook gekozen kunnen worden voor de

onderverdeling in data, processen, tools, governance, team en meetbare doelen (Saltz & Shamshurin, 2016), maar zoals eerder aangegeven komen deze ongeveer neer en is voor de categorisatie van Mikalef et al. (2019) gekozen.

Ondanks dat er behoorlijk wat artikelen in de literatuur te vinden zijn over KSF, falen nog veel DS projecten. Er is duidelijk meer nodig dan alleen kennis over KSF. Er is behoefte aan een proces, dat houvast biedt bij het managen van een DS project (Miller, 2018; Saltz et al., 2018; Saltz & Shamshurin, 2016). Daarnaast is er ook een projectplanning en een duidelijke project scope en doel nodig (Miller, 2018). In de praktijk blijkt dat DS projecten vaak ad hoc uitgevoerd worden, of een proces gebruiken die niet toereikend is (Saltz, 2015; Saltz et al., 2018; Saltz & Shamshurin, 2016). Niet altijd wordt er een proces management methodologie (PMM) toegepast die past bij de context, maar wordt bijvoorbeeld een methodologie gebruikt die men al kent. De context is belangrijk, zo is bijvoorbeeld bekend dat organisaties die bijvoorbeeld verschillen met betrekking tot infrastructuur, organisatiegrootte en branche, anders gemanaged moeten worden (Mikalef et al., 2019; Saltz, Shamshurin, & Connors, 2017). Dit geldt ook voor in hoeverre een datavraag helder of vaag is, dat beantwoord moet worden met een DS project.

Een PMM wordt ook niet zomaar geaccepteerd door een team. Zo zijn er sleutel acceptatiefactoren zoals de geschiktheid, complexiteit en relatieve voordelen (bijvoorbeeld effectiviteit, communicatie) van een methodologie, die een rol spelen bij het in gebruik nemen van een PMM (Saltz, 2018). Er worden verschillende PMM-en gezien in de praktijk, zoals bijvoorbeeld SCRUM en Kanban (Saltz et al., 2018). Daarnaast valt het op dat er hybride proces methodologieën beginnen te ontstaan, die zogenaamde 'Knowledge Discovery' procesmodellen voor data mining (KDDM: Knowledge Discovery and Data Mining) als basis hebben (Reinartz, 2002; Cios & Kurgan, 2005) of een combinatie zijn van bestaande software engineering methodologieën met enkele aanvullende elementen uit KDDM of andere raamwerken of methodologieën (Saltz et al., 2018). De term KDDM is voorgesteld als de meest geschikte naam voor het totale proces van 'Knowledge Discovery' (Reinartz, 2002; Cios & Kurgan, 2005). Voorbeelden van KDDM zijn bijvoorbeeld CRISP-DM, ASUM-DM en SEMMA (Angée, Lozano-Argel, Montoya-Munera, Ospina-Arango, & Tabares-Betancur, 2018).

Feit blijft dat ondanks de eventuele kennis van KSF en de eventuele kennis van enkele PMM-en en de acceptatiefactoren die daarbij een rol spelen, het toch moeilijk is om in te schatten voor een organisatie, hoe dan een optimale PMM er uit zou moeten zien en waar deze dan aan zou moeten voldoen. Het moet in ieder geval een flexibele methodologie zijn, al was het alleen maar omdat de context voor elke organisatie anders kan zijn, maar ook vanwege iteratieve en verkennende taken in DS projecten. Wat hierbij kan helpen, is een methodologie te kiezen, aan te passen of te bedenken waarbij er rekening gehouden wordt met het karakter van DS, zoals bijvoorbeeld de systematiek van Das, Cui, Campbell, Agrawal, & Ramnath (2015), hoewel hierbij alleen de gegevens centraal staan in elk van de iteratieve analytische stappen bij de data gedreven planning en uitvoering van de onderzoek pijplijn.

Bij het holistische raamwerk van Dutta & Bose (2015) staan niet alleen de gegevens centraal, vandaar dat dit raamwerk kan helpen bij het aanpassen of ontwerpen van een geschikte PMM voor een DS project. Zij presenteren een gemakkelijk te begrijpen maar gedetailleerde 'roadmap', die kan worden gevolgd door organisaties die van plan zijn DS-analyseprojecten te implementeren. Essentieel voor het succes van een DS project hierin zijn een duidelijk begrip van het bedrijfsprobleem, een gedetailleerd en goed geplande stap-voor-stap projectkaart, een cross-functioneel (multidisciplinair) projectteam, acceptatie van innovatieve visualisatie technieken, beschermheerschap en actieve betrokkenheid van het topmanagement en een cultuur van data gedreven beslissingen maken (Dutta & Bose, 2015). Dit raamwerk zal een goede basis bieden voor

vervolgonderzoek, aangezien het ten opzichte van andere data mining methodologieën zoals KDD, CRISP-DM, SEMMA en ASUM-DM een stuk gedetailleerder is. Zie hieronder het raamwerk (tabel 2.3) van Dutta & Bose (2015) ten opzichte van KDD, CRISP-DM, SEMMA en ASUM-DM. Door gebruik te maken van het meest uitgebreide raamwerk van Dutta & Bose (2015) in combinatie met de kennis over de KSF kan er nu ingezoomd worden op delen van het DS proces. Zo kunnen inzichten verkregen worden over wanneer welke kritieke succesfactor nu een rol speelt en hoe een organisatie dus ook presteert op specifieke delen in het DS proces.

| Framework for implementation of Big Data projects in firms | Dutta & Bose                    | KDD   | ASUM-DM             | CRISP-DM               | SEMMA              |        |
|--|---------------------------------|---|---------------------|------------------------|--------------------|--------|
| No. Of Steps   | 10                              | 9   | 6                   | 6                      | 5                  |        |
| Name of Steps  | Business problem                | Developing and Understanding of the Application   | Analyze             | Business Understanding |                    |        |
|  | Research                        |   | Design              |                        |                    |        |
|  | Cross Functional Team Formation |   |                     |                        |                    |        |
|  | Project Roadmap                 |   |                     |                        |                    |        |
|  | Data collection and examination | Creating a Target Data Set                        |                     |                        | Data Understanding | Sample |
|  |                                 | Data Cleaning and Pre-processing                  |                     |                        | Explore            |        |
|  |                                 | Data Transformation                               |                     |                        | Data Preparation   | Modify |
|  | Data analysis and modelling     | Choosing the suitable Data Mining Task            | Configure and build |                        | Modeling           | Model  |
|  |                                 | Choosing the suitable Data Model Mining Algorithm |                     |                        |                    |        |
|  |                                 | Employing Data Mining Algorithm                   |                     |                        |                    |        |
| Data visualization   | Interpreting Mined Patterns     |   |                     | Evaluation             | Assessment         |        |

|  |                             |                            |                      |            |  |
|--|-----------------------------|----------------------------|----------------------|------------|--|
|  | Insight generation          | Using Discovered Knowledge | Deploy               | Deployment |  |
|  | Integration with IT systems |                            | Operate and Optimize |            |  |
|  | Training people             |                            |                      |            |  |

Tabel 2.3: Raamwerk Dutta & Bose (2015) versus KDD, ASUM-DM, CRISP-DM en SEMMA

## 2.4. Doel van het vervolgonderzoek

Er is dus genoeg onderzoek gedaan naar wat belangrijk is voor een DS project. Wat echter nog ontbreekt, is om met deze kennis te kunnen bepalen hoe DS projecten nu het beste uitgevoerd kunnen worden en dus ook te bepalen wat nu een optimale PMM is, gegeven de context van een organisatie. Bestaande genoemde raamwerken die rekening houden met het karakter van DS projecten en zoveel mogelijke relevante KSF kunnen integreren in de methodologie (voor zover niet gedaan), zouden hierbij kunnen helpen. Het is daarom ook interessant om te onderzoeken wat nu de optimale combinatie is van KSF, de fases in een project en de specifieke context van een organisatie en PMM. Dit kan bijdragen aan een raamwerk, waarmee organisaties die DS projecten bedrijven, hun voordeel kunnen doen. Dit met als doel om DS projecten uit te voeren, zodat zij succesvol kunnen zijn.

## 3. Methodologie

Hier volgt een verantwoording van het uitgevoerde empirische onderzoek.

### 3.1. Conceptueel ontwerp: keuze van onderzoeksmethode(n)

Om te bepalen hoe DS projecten op succesvolle wijze uitgevoerd kunnen worden, waarbij ook meer aandacht is voor de projectmatige en organisatorische aspecten, is het enerzijds van belang om inzicht te verkrijgen in de KSF van een DS project op basis van literatuuronderzoek en anderzijds is het belangrijk om inzicht te verkrijgen in welke uitdagingen organisaties in de praktijk nu hebben bij hun DS projecten en welke zaken juist hebben bijgedragen, zodat zij succesvol konden zijn met DS.

Er moet dus gezocht worden naar rijke beschrijvingen van hoe DS projecten nu uitgevoerd worden. Wat verstaat een organisatie überhaupt onder DS en hoe is DS georganiseerd? Wellicht zijn er ook 'lessons learned' gedocumenteerd. Het is interessant om te weten hoe er gestart wordt met een DS project en hoeveel ervaring de organisatie heeft met DS projecten en of zij ook al succesvol zijn geweest met DS projecten en wat de uitdagingen momenteel zijn. Aangezien er data verzameld zal moeten worden met betrekking tot ervaringen die organisaties gehad hebben met hun DS projecten, zal er kwalitatief onderzoek gedaan worden. Dit onderzoek probeert namelijk onderzoeksvragen te beantwoorden op basis van onder andere interpretaties, ervaringen en betekenis. De casestudy is zo een vorm van kwalitatief onderzoek en leent zich dus uitermate voor het verkrijgen van de context van een onderzoek en de processen die worden doorlopen (Morris & Wood, 1991). Dit kan bijvoorbeeld door middel van observatie, met behulp van interviews en vragenlijsten. Gegeven de beperkte tijd voor dit onderzoek, is observeren geen optie en geven vragenlijsten geen mogelijkheid tot wedervraag en is er dus een voorkeur voor 'face-to-face' semigestructureerde interviews bij een drietal organisaties.

De voordelen van 'face tot face' zijn dat de persoonlijke sfeer ervoor kan zorgen dat de respondenten zich sneller openstellen en meer vertellen en respondenten geobserveerd kunnen worden, waardoor de onderzoeker kan ingaan op gezichtsuitdrukkingen, zoals verwarring of ongemak.

De casestudy zal dus een 'multiple' casestudy zijn, dit biedt namelijk de mogelijkheid tot het vergelijken van organisaties. Vaak zal dit zijn, bij een speciale eenheid van data en analytics, de zogenaamde 'Analytics Competency Centers' (Schüritz, Brand, Satzger, & Bischhoffshausen, 2017), of bij een andere afdeling waar data analisten of 'data scientists' werkzaam zijn of bij een afdeling dat zich bezig houdt met productontwikkeling.

De bestudering van meerdere cases zal op deductieve wijze plaatsvinden, deductief onderzoek begint altijd met een theorie. Zo zal bijvoorbeeld de theorie van de KSF en van PMM-en gebruikt worden. Verder is zoals ook vermeld in hoofdstuk 2.4 gebleken dat de context waarin een organisatie zich bevindt invloed kan hebben op de wijze waarop DS projecten uitgevoerd worden, vandaar dat er een holistische benadering bij de casestudy wordt genomen, zodat er dus ook flink aandacht besteed wordt aan de context van de organisaties, waarin zij zich bevinden.

### 3.2. Technisch ontwerp: uitwerking van de methode

Bij de selectie van de organisaties zal gebruik gemaakt worden van het netwerk van de onderzoeker. Idealiter zal er iemand geïnterviewd worden, die de kennis beschikt over de wijze van het managen van DS producten, wellicht een project manager, maar het zou ook een lead data scientist / ingenieur kunnen zijn of een datamanager of hoofd van een 'Analytics Competency Center'.

Voor het onderzoek zal de aanpak van Yin (2014) gevolgd worden, waarbij het formuleren en onderzoeken van proposities centraal staat. Het is een aanpak die veel lijkt op het toetsen van hypothesen toetsen, zoals je dat bij kwantitatief onderzoek op statistische wijze doet. De theorie van KSF bij DS projecten geeft een theoretische verwachting. Deze verwachtingen kunnen vervolgens vergeleken worden met eigen gevonden empirische bevindingen door middel van de te nemen semigestructureerde interviews, die moeten leiden tot zo rijk mogelijke beschrijvingen van het managen van DS projecten.

De interviews van maximaal 2 keer een uur per organisatie, zullen bij voorkeur ingepland worden bij de organisatie zelf, een omgeving waar de geïnterviewde zich vertrouwd zal voelen. Verder zullen de interviews met permissie van de geïnterviewden opgenomen worden met een IOS app (Memorecorder voor Iphone), hiermee kunnen hoogwaardige, langdurige geluidsopnames gemaakt worden. Ook zal er metadata verzameld en opgeslagen worden van de respondent en de organisatie, zoals de functie van de respondent, grootte van de organisatie etc.

Ter behoeve van de interviews is er een interview protocol opgesteld, zodat er aan de ene kant geen zaken vergeten worden voorafgaand en tijdens de interviews en zodat er ook in grote lijnen dezelfde vragen worden gesteld aan de verschillende respondenten, zie voor details bijlage 3: Interview protocol. Dit maakt de vergelijking van de verschillende organisaties makkelijker. De vragen in het interview behelzen in ieder geval de volgende onderwerpen:

- Algemeen
- Ervaringen met DS projecten
- Cultuur & organisatie
- Het DS proces



Los van bovenstaande onderwerpen, hebben de geïnterviewden ook de mogelijkheid gekregen om onderwerpen te benoemen die zij zelf relevant vonden in kader van het uitvoeren van DS projecten, maar niet aan bod waren gekomen.

Tot slot zullen de opnames getranscribeerd worden, uitgeschreven worden in 'transcripten', waarbij spraak naar tekst omgezet wordt. Tegenwoordig bestaan er ook allerlei software pakketten op de markt die het uittypen van transcripten automatiseren of vergemakkelijken, zoals 'Express Scribe', waarvan ook een gratis versie bestaat en dus wellicht gebruikt kan worden. Een eerste versie zal ook voorgelegd worden aan de respondent, zodat deze kan controleren of de tekst overeenkomt met hetgeen hij/zij ook bedoelde te zeggen, zie ook hoofdstuk 3.4 (validiteit). Tot slot zullen alle interviewdata opgenomen worden in een geheime bijlage, dit in verband met de vertrouwelijkheid van de data.

### 3.3. Gegevensanalyse

Het houden van interviews kan tot grote hoeveelheden (geschreven) teksten leiden, daarom is het belangrijk om daar overzicht over te behouden. Dan is het ook makkelijker om patronen te herkennen en zelfs cases te vergelijken. Voor het analyseren van kwalitatieve data kan een standaard aanpak van coderen en categoriseren toegepast worden. Een uitgeschreven interview (het transcript) wordt in detail geanalyseerd door telkens codes (begrippen, onderwerpen) aan te geven. Vervolgens probeert de onderzoeker de codes in categorieën in te delen, bij voorkeur is er voor iedere code een categorie (Gibbs, 2007).

Categorieën kunnen vooraf zijn vastgesteld op basis van literatuuronderzoek, maar ook naar voren komen uit de verzamelde data. Het eerste noemt men 'concept-driven', het tweede 'data-driven'. Uiteindelijk probeert de onderzoeker ook de categorieën met elkaar in verband te brengen. Zo kan bijvoorbeeld een kritieke succesfactor gecodeerd worden met een label, die vervolgens de waarde 0 of 1 kan ('dummy coding') aannemen, naar gelang de succesfactor aanwezig is bij het managen van DS projecten bij de betreffende organisatie (Gibbs, 2007).

Voor het coderen en categoriseren zijn er tegenwoordig ook goede software pakketten op de markt, zoals 'NVivo', 'ATLAS.ti' en 'Kwalitan', dit zijn programma's die de onderzoeker ondersteuning biedt bij kwalitatieve, interpretatieve analyses. Derhalve zal hier ook gebruik van gemaakt worden. Alle drie de pakketten zorgen voor een efficiënte opslag van de gegevens en bieden diverse hulpmiddelen om het kwalitatieve materiaal te analyseren, zoals coderen, het selecteren en terugzoeken van gegevens, het ordenen van codes, overzichten van codes of van woorden in de tekst, 'key words in context' en het schrijven van memo's. Het inschakelen van een computerprogramma ter ondersteuning van de analyse moet vooral gezien worden in het bevorderen van het systematisch werken en in het controleerbaar maken van de analyses. Nadeel kan zijn dat er tijd geïnvesteerd moet worden om het computerprogramma te leren kennen.

### 3.4. Reflectie t.a.v. validiteit, betrouwbaarheid en ethische aspecten

#### **Interne validiteit**

De interne validiteit gaat over hoe goed het ontwerp van het onderzoek is en of er vanuit dat ontwerp dus de juiste conclusies getrokken kunnen worden (Saunders et al., 2016). Om de onderzoeksvragen te beantwoorden is er literatuuronderzoek gedaan en is er data verzameld met behulp van semigestructureerde interviews, die vervolgens geanalyseerd is. De volgende zaken zijn gedaan om de interne kwaliteit te borgen. Om de kwaliteit van het literatuuronderzoek zo hoog

mogelijk te houden is het literatuuronderzoek gedaan op basis van een groot aantal stappen conform het systematische literatuuronderzoek van Okoli & Schabram (2010). Om de kwaliteit van het semigestructureerde interviews te borgen is er een interview protocol opgesteld, hetgeen ook gereviseerd is door de begeleider van dit onderzoek. Respondenten zijn verder geselecteerd op basis van kennis en expertise over het onderwerp. Bovendien zijn er ook meerdere organisaties en meerdere respondenten per organisatie geïnterviewd, dit maakt het ook makkelijker om patronen of tegenstellingen te vinden. Door ook de bevindingen van het literatuuronderzoek te vergelijken met bevindingen vanuit de interviews kon ook vaak getrianguleerd worden en daar waar dit niet kon is hier expliciet op in gegaan en is er een toelichting gegeven. Ook zijn verzamelde gegevens teruggestuurd naar de respondenten voor eventuele correcties in kader van eventuele interpretatiefouten. Tot slot is bij de analyse van gegevens gebruik gemaakt van coderen conform Gibbs (2007) en zijn coderingen zoveel mogelijk overzichtelijk gepresenteerd met behulp van tabellen, op basis waarvan conclusies zijn getrokken en beargumenteerd.

### **Construct validiteit**

De construct validiteit gaat over de mate waarin daadwerkelijk gemeten wordt, wat beoogd wordt te meten (Saunders et al., 2016). Om een zo hoog mogelijke construct validiteit te bewerkstelligen worden eventuele ruimtes voor interpretaties geminimaliseerd, doordat in het theoretische kader de kernbegrippen van het onderzoek geoperationaliseerd worden. Zo wordt in het theoretische kader uiteengezet dat het succes van een DS project gemeten kan worden op basis van KSF, omdat dit een gebruikelijke methode is in de wetenschap, om de prestaties van een organisatie te meten (Rockart, 1979). Daarnaast is gecontroleerd of de formulering van de interviewvragen precies genoeg is, door het testen van vragen op een aantal proefrespondenten, om te achterhalen of de vragen inderdaad meten wat ik wil meten en of ze goed begrepen worden.

### **Externe validiteit**

De externe validiteit gaat om de mate van generaliseerbaarheid van de onderzoeksresultaten (Saunders et al., 2016). Het gaat bij de casestudy dan wel om de theoretische generaliseerbaarheid (Yin, 2004). Kan dus dat wat waargenomen wordt bij de organisaties, vertaald worden naar het theoretisch kader in de vorm van (een set van) proposities? In het bijzonder gaat het er in dit onderzoek dus om of eigen gevonden empirische bevindingen door middel van genomen semigestructureerde interviews te generaliseren zijn. Vanwege de beperkte tijd zijn er 3 organisaties (2 interviews per organisatie) bestudeerd, vanwege dit beperkte aantal kan dit in theorie een risico vormen voor de externe validiteit. Echter omdat bij dit beperkte aantal organisaties, heel veel KFS uit de literatuur bevestigd zijn en er ook veel overeenkomsten tussen de organisaties zijn gevonden, is de overtuiging groot dat de externe validiteit toch behoorlijk geborgd is. Daarnaast, zijn mensen wanneer zij in een vertrouwde omgeving worden geïnterviewd, mogelijk veel eerlijker in hun antwoorden, hetgeen uiteindelijk ook weer de generalisatie (ecologische generalisatie) ten goede komt. Daarom zijn interviews via Skype gedaan en de camera hoefde niet per se gebruikt te worden als de respondent dat prettiger vond. Hierdoor hebben alle interviews plaatsgevonden onder vertrouwde omstandigheden voor de respondenten.

### **Betrouwbaarheid**

Met betrekking tot de betrouwbaarheid van het onderzoek gaat het erom aannemelijk te maken dat een herhaling van het onderzoek tot dezelfde resultaten zal leiden (Saunders et al., 2016). De procedures die gevolgd zijn, zijn zodanig beschreven, zodat het proces voor een buitenstaander navolgbaar, en dus in principe herhaalbaar is. Op die manier kan iedereen nagaan hoe te werk is gegaan, hoe geïnterpreteerd is en hoe ik tot conclusies ben gekomen. Het proces van het onderzoek is dus uitvoerig gerapporteerd, met in ieder geval informatie over:

- selectie van cases, metadata van cases
- opname van interviews

- transcripten
- coderen / analyseren van transcripten / gebruikte kwalitatieve onderzoek software

### **Ethische aspecten**

Met betrekking tot ethische aspecten rondom data verzameling, is er per email om toestemming gevraagd of interviews afgenomen mogen worden en opgenomen mogen worden. Daarbij zijn de volgende zaken in acht genomen:

- Vrijwillige deelname: controleren of mensen die deelnemen aan mijn dataverzameling begrijpen dat ze niet hoeven mee te werken en dat ze op elk moment hun deelname kunnen staken;
- 'Informed consent': controleren of mensen die deelnemen aan de dataverzameling snappen waar ze bij betrokken worden;
- Breng geen schade toe: eventuele gevoelige onderwerpen op integere manier behandelen;
- Beschermd identiteit: persoonsgegevens en bedrijven zullen geanonimiseerd worden;
- Neutraliteit: objectief blijven;
- Minimalisme: controleren of alleen wat ik moet weten verzameld wordt.

## 4. Resultaten

### 4.1. Beschrijving uitvoering onderzoek

Het doel van dit onderzoek is om te bepalen hoe DS projecten nu het beste gemanaged kunnen worden, daarom zijn er voor deze casestudy organisaties geselecteerd, die al minimaal 3 jaar DS projecten uitvoeren of al minimaal 10 DS projecten uitgevoerd hebben. Zo worden er dus organisaties verkregen, die voldoende ervaring hebben met DS projecten. Aangezien organisaties ook vergeleken zullen worden, vereist dit ook dat het ongeveer even grote organisaties zijn, zodat zij ook ongeveer dezelfde middelen hebben om bijvoorbeeld in DS te kunnen investeren en daardoor ook redelijk vergelijkbaar zijn.

Om dit te bewerkstelligen is een selectie gemaakt uit dienstverlenende beursgenoteerde bedrijven, die ook tot de top van de 50 grootste Nederlandse bedrijven behoren. Verder, omdat het ook interessant kan zijn om juist verschillen tussen organisaties te ontdekken, is er niet per se gekozen voor organisaties die in dezelfde branche zitten. Tot slot, moest elke organisatie bereid zijn om 2 persoonlijke interviews van een uur bij 2 verschillende personen te laten opnemen, waarbij bovendien elk te interviewen persoon een belangrijke rol moest hebben bij DS projecten, zoals bijvoorbeeld een manager DS of een data scientist. In verband met COVID-19 is er uiteindelijk afgestemd met de organisaties om de interviews niet 'face-to-face', maar via een Skype-verbinding af te nemen en de interviews te mogen opnemen met de opname mogelijkheden van de Skype applicatie.

Op basis van de Skype opnames zijn vervolgens transcripten gemaakt van de opnames. Op basis van de transcripten zijn met behulp van het software pakket Atlas.ti de transcripten gecodeerd op een 'data-driven' manier, dat wil zeggen dat codes (categorieën) niet vooraf zijn vastgesteld, maar op basis van de verzamelde data zijn vastgesteld. Codes kunnen in veel gevallen wel lijken op KSF, omdat veel, maar zeker niet alle interviewvragen betrekking hebben op bekende KSF in de literatuur. De verschillende vragen met antwoorden uit de interviews zijn dus gelabeld, wanneer dit relevant leek.

Dit heeft geresulteerd in een lijst met codes voor:

- Respondenten

- Organisaties
- Hoofdlabels: ‘Algemeen’, ‘Ervaringen met DS projecten’, ‘Cultuur & organisatie’, ‘Het DS proces’, ‘Niet besproken’
- Sublabels: Dit is een detaillering van de verschillende hoofdlabels

Zie voor de volledige lijst codes bijlage 4: Lijst codes. Overigens kan een sublabel onder verschillende labels vallen. Zo kan een bepaalde sublabel onder meerdere respondenten vallen, maar ook onder meerdere organisaties en ook onder meerdere hoofdcodes. Nadat alle relevante stukken tekst gecodeerd waren, zijn met Atlas.ti de volgende rapporten zijn gemaakt, om gecodeerde quotes te analyseren:

1. Groepering van de codes per organisatie (zie bijlage 5)
  - Ter behoeve van het bepalen van de verschillen tussen de organisatie
  - Ter behoeve van het bepalen van de overeenkomsten tussen organisaties
  - Ter behoeve van het bepalen van welke codes bij alle organisaties voorkomen
  - Ter behoeve van het bepalen van welke codes niet bij een organisatie voorkomen
  - Ter behoeve van het bepalen van welke codes frequent bij een organisatie voorkomen
2. Groepering van de codes per functies (zie bijlage 6)
  - Ter behoeve van het bepalen van de verschillen tussen de functies
  - Ter behoeve van het bepalen van de overeenkomsten tussen de functies
  - Ter behoeve van het bepalen van welke codes bij de verschillende functies voorkomen
  - Ter behoeve van het bepalen van welke codes niet bij de verschillende functies voorkomen
  - Ter behoeve van het bepalen van welke codes frequent bij de verschillende functies voorkomen
3. Matrix met combinaties van subcodes met aantallen (zie bijlage 7)
  - Ter behoeve van het bepalen van relaties tussen de verschillende subcodes

## 4.2. Algemene resultaten

Op basis van het rapport in bijlage 5 blijkt dat de 3 meest voorkomende codes die ‘overall’ (over alle organisaties heen) voorkomen zijn:

- Bedrijfscultuur
- Organisatie DS
- Proces management methodologie

De thema’s die betrekking hebben op deze codes spelen dus een belangrijke rol bij de case study. De reden dat bedrijfscultuur zo vaak ter sprake komt, zal nader onderzocht moeten worden, maar dat ‘Organisatie DS’ vaak voorkomt is niet gek, omdat dit onderwerp duidelijk op alle organisaties van toepassing is, aangezien zij DS projecten uitvoeren. Verder lijkt het er op dat ‘proces management methodologie’ ook een belangrijk thema is, hier zal dus zeker op het niveau van de organisaties of functies nog scherper gekeken naar worden.

Verder blijkt op basis van het rapport in bijlage 7 dat bepaalde combinaties van codes vaker voorkomen. Hoewel dit een logische reden kan hebben, omdat bepaalde codes nu heel veel te maken hebben met elkaar. Toch zal hier naar worden gekeken, omdat er wellicht ook niet voor de hand liggende relaties naar voren kunnen komen. De paren die meer voorkomen dan anderen zijn:

- Bedrijfscultuur & Data toegankelijkheid
- Bedrijfscultuur & Management support
- Bedrijfscultuur & Belangrijkheid van DS projecten
- Bedrijfscultuur & Data governance
- Bedrijfscultuur & DS expertise op bestuursniveau
- Aanleiding voor DS projecten & Belangrijkheid data
- Aanleiding voor DS projecten & Management support
- Belangrijkheid van DS projecten & Bezig met DS projecten
- Organisatie DS & Bezig met DS projecten
- Organisatie DS & Type DS project
- Data governance & Data toegankelijkheid
- Data governance & DS expertise op bestuursniveau
- Management support & DS expertise op bestuursniveau
- Data governance & Management support

Zoals verwacht zien we een aantal relaties die niet vreemd zijn, zoals de “Aanleiding voor DS projecten’ wanneer data heel belangrijk is (‘Belangrijkheid data’) voor een organisatie. Door de quotes die betrekking hebben op bovenstaande zaken nog eens onder de loep te nemen, zou de vraag gesteld kunnen worden of data toegankelijkheid beter geregeld is, zodra data governance hoger op de agenda staat en of data governance beter geregeld is, zodra er meer expertise is op bestuursniveau. Uit de analyse van de quotes volgt dat respondenten de indruk hebben, dat zodra de expertise op bestuursniveau toeneemt, dat impact heeft op een betere data governance, betere data toegankelijkheid en verbeterde management support en dus samengevat een verbeterde bedrijfscultuur. Verder kan geconcludeerd worden dat de reden dat bedrijfscultuur ‘overall’ zo vaak ter sprake is gekomen, is omdat het met nogal wat codes samenhangt. Tot slot lijken de codes die ‘overall’ weinig voorkomen, het resultaat te zijn van specifieke codes die meestal maar door 1 van de organisaties benoemd zijn, ook daar zal meer ingezoomd worden bij de behandeling van de verschillende organisaties. In de volgende hoofdstukken zullen achtereenvolgens de profielen van de verschillende organisaties en de bijbehorende resultaten per respondent worden beschreven. Aan het eind van elke paragraaf zullen de resultaten geaggregeerd worden over de respondenten en dus per organisatie samengevat worden in tabelvorm op basis van het framework van Dutta & Bose (2015), zie daarvoor eventueel ook tabel 2.3 in hoofdstuk 2.3.

### 4.3. Profiel en resultaten organisatie 1

Organisatie 1 is al 7 jaar geleden begonnen met DS projecten en voert nu jaarlijks zo een 15 tot 20 projecten uit op jaarbasis. De duur van de projecten is verschillend, maar de langere DS projecten lopen gemiddeld 3 tot 6 maanden, waarbij dus de helft van de tijd zuiver nodig is om te industrialiseren, het dus feitelijk naar productie brengen van een DS oplossing. De organisatie heeft een centrale DS afdeling met onder andere data scientisten en engineers, die in projecten meedraaien met de business. Inmiddels zitten er ook data scientisten vast in een business team. De centrale DS afdeling bestaat voornamelijk uit ongeveer 20 jonge professionals. Verder wordt er binnen de organisatie gebruik gemaakt van een Scrum, een Agile procesmanagement methodologie (Scrum vs Agile, wat is het verschil eigenlijk?, 2020). De codes waarop de meeste quotes betrekking hebben voor organisatie 1:

1. Bedrijfscultuur

2. Data toegankelijkheid
3. Organisatie DS
4. Aanleiding voor DS projecten
5. Urgentie voor datakwaliteit

Met name de codes hierboven bij 1, 2 en 5 zijn interessant om extra te onderzoeken, hoewel code 1 wel te verwachten is op basis van de algemene geaggregeerde resultaten, verder zijn de overige 2 codes wel te verwachten, vanwege de thema's.

Tot slot hebben een aantal quotes betrekking op de volgende codes, die alleen bij deze organisatie ter sprake zijn gekomen, dat zijn:

- Betrouwbaarheid DS oplossing
- Nieuwe IT technologie
- Best practice process for DS
- Softe of randvoorwaardelijke elementen

Bovenstaande genoemde relevante codes zullen extra aandacht krijgen en besproken worden in de resultaten op basis van de 2 afgenomen interviews bij deze organisatie. Elk interview wordt afzonderlijk besproken, omdat is opgevallen dat er interessante verschillen zijn in het perspectief van bijvoorbeeld een data scientist en een manager DS, maar ook bij de andere vertegenwoordigers van een zelfde organisatie. Door de resultaten per geïnterviewde te bespreken, kan worden ingegaan op eventuele opmerkelijke verschillen tussen de geïnterviewden bij dezelfde organisatie.

### **Resultaten op basis van het interview met de Groupmanager Operations Research & Big Data bij organisatie 1 (GMO)**

De GMO is de manager van een centrale DS club, die als het ware als een interne consultancy club functioneert en mensen uitleent aan de verschillende business teams. Hij geeft aan dat hij het idee heeft dat ze met betrekking tot het toepassen van DS technieken aardig voorop lopen ten opzichte van andere organisaties. Verder is hij van mening dat zeker in het begin van de DS journey, externe expertise op senior niveau enorm kan helpen om kennis hiaten op te vullen. Hiervan hebben zij zelf ook veel profijt van gehad, ondanks dat de externe mensen nu helaas weg bezuinigd zijn. Zo gaf de GMO aan:

*“Uhm nou kijk wat ik wel belangrijk vind, wat helpt in ieder geval, zeker in de beginfase, is dat je externen hebt, die senior, die een brede ervaring hebben. Zeker de meeste teams die ik zie bij bedrijven, zijn vaak juniors of mensen die zijn doorgestroomd, ook junior zijn daardoor, dan helpt het enorm als je in zo een team ook een ervaren iemand hebt, dat er experts zijn die een team kunnen helpen” (GMO).*

De manier van werken bij de organisatie is vooral door 2 data scientisten (ook engineers) mee te laten draaien in meerdere multidisciplinaire business teams. Opvallend is dat er enerzijds zo kennis wordt geborgd door in paartjes van data scientisten te werken, dat er zo ook veel kennis wordt opgedaan door in meerdere teams te werken en anderzijds is opvallend dat er zo ook direct een goede betrokkenheid van de business aanwezig is, aangezien de business het team ook leidt. In de business teams wordt gebruik gemaakt van Scrum, een Agile proces management methodiek. Dit werkt goed, doordat er aan de ene kant door stapsgewijs en iteratief te werken er constant feedback verkregen wordt van de business, waarbij de business ook kritisch mag zijn en projecten ook tijdig gestopt kunnen worden, als dat nodig is. Door middel van proof of concepts (concept waarmee

gedemonstreerd kan worden dat een bepaalde DS oplossing geschikt is voor een bepaald bedrijfsdoel) laten de data scientists een eventuele toegevoegde waarde zien van een DS oplossing. Op die manier proberen ze dan ook de business mee te krijgen en te overtuigen om de oplossing op te schalen. Bij het opschalen van een DS oplossing speelt het naar productie brengen een rol en dat is ook gelijk het belangrijkste aandachtspunt volgens de GMO. Zo zegt hij:

*“Ja, het is altijd van heb je wel de juiste mensen, de juiste data en de juiste techniek beschikbaar. Dus de mensen dat lukt aardig, data kwaliteit dat varieert nogal, om het op een gegeven moment naar productie te brengen, ja daar ligt wel een uitdaging, nieuwe technieken die nog onbekend zijn binnen de IT hoek, dat vraagt nog wel om extra tijd en moeite” (GMO).*

Een punt waarover hij tevreden is, is de DS community die ze hebben, die er voor zorgt dat zowel de ‘vaste’ data scientists in de organisatie als ook de data scientists uit de centrale afdeling, ervaringen en kennis kunnen uitwisselen door periodiek bij elkaar te komen en trainingen en demonstraties te krijgen. Tot slot heeft hij aangestipt dat het voor het vertrouwen van de business als ook voor het hoger management belangrijk is, dat oplossingen die opgeleverd worden zodanig betrouwbaar zijn, dat ze ook de belofde toegevoegde waarde opleveren. Zo gaf hij aan:

*“Uiteindelijk voordat je de gehele organisatie ergens mee laat werken, dan moet je ook zeker weten dat het voldoende betrouwbaar is, dat is 1 maar dat het ook voldoende oplevert” (GMO).*

### **Resultaten op basis van het interview met de product owner bij organisatie 1 (PO)**

De PO van één van de business teams en eigenaar van allerlei uitgaven data, probeert inzichten te verkrijgen uit deze data om tot kostenbesparingen en verbetervoorstellen te komen. In het kader daarvan staat zij heel erg open voor DS oplossingen, maar zij ziet dat zij met name beperkt wordt door softe of rand voorwaardelijke zaken, waar meer aandacht voor zou mogen komen, zoals bijvoorbeeld de toegang tot data waar zij geen eigenaar van is. Zo gaf zij aan:

*“...maar zoals ik al zei als alleen al de data krijgen, allerlei politieke discussies met zich meebrengt en mensen die het niet willen geven en noem het maar op dan ben je zo al 2 tot 3 maanden verder en dan heb je een heel mooi, logisch proces, maar de werkelijkheid is vaak iets complexer, ik denk dat daar ook de grootste uitdaging zit, niet zo zeer in techniek, maar het ligt voornamelijk als het ware in rand voorwaardelijke of softe elementen” (PO).*

Het verkrijgen van data lijkt regelmatig een lastig tijdrovend en politiek proces te zijn. Een ander item betreft de data kwaliteit, zo benoemt zij vervuilingen en inconsistentie in data, die meer de aandacht mogen krijgen van het hoger management.

De PO vraagt zich ook af of er iets bestaat als een best practice proces voor data science, zelf is ze namelijk wel te spreken over het werken met de Scrum PMM, waarbinnen er ook gewoon ruimte is om te experimenteren en te verkennen door de data scientists. Een ander aandachtspunt die ze wel ziet, is de behoefte aan flexibele IT support, aangezien de IT afdeling heel erg projectmatig werkt. Tot slot gaf ze aan dat er nog te weinig DS oplossingen opgeschaald worden, omdat het hoger management in haar ogen nog te huiverig is om te experimenteren met DS oplossingen en te investeren in DS projecten, zonder dat die direct een ‘return on investment’ opleveren. Zo gaf ze aan:

“Ja, want dat is wel altijd toch de eerste vraag die gesteld wordt als je een project wil opstarten. Wat is je return on investment?” (PO).

Het zou wat haar betreft dan ook kunnen helpen als de data expertise op bestuursniveau omhoog gaat. Door de codes tussen de GMO en PO te vergelijken (zie bijlage 6), valt het op dat het aantal quotes op bedrijfscultuur bij de PO aanmerkelijk hoger is. Door hierop in te zoomen valt duidelijk op dat de PO de bedrijfscultuur het belangrijkste aandachtspunt vindt. Zo is haar belangrijkste boodschap dat er meer management support mag komen, om data kwaliteit, data toegang en IT support meer aandacht te geven en de belangrijkste boodschap van de GMO is dat ze nog niet zo snel DS oplossingen naar productie kunnen brengen vanwege nieuwe technologieën en de kennis daarover.

Zie hieronder de genoemde resultaten samengevat voor organisatie 1 in tabelvorm. De genoemde zaken die goed of zeer goed gaan (aangegeven met een ‘+’ en ‘++’) en de zaken die beter kunnen of slecht gaan (aangegeven met een ‘-’ en ‘--’), geplot over de verschillende fases in een DS project. Er zit geen model achter de beoordeling en het dient dus meer ter indicatie om de organisaties onderling te kunnen vergelijken.

| Hoofdfases                 | Subfases   | Organisatie 1 | Toelichting  |
|----------------------------|--|---------------|--|
| Strategische voorbereiding | Business problem (helder begrip vraagstuk)   | +             | Data scientisten zitten in een multidisciplinair business team, waarin alle begrip gevormd kan worden  |
|                            | Research (identificatie mogelijke oplossingen en bepalen of bestaande infrastructuur voldoet)  | -             | Naar productie brengen van oplossingen gaat nog niet zo snel, dus hier wordt nog niet genoeg aandacht aan gegeven in ieder geval aan het begin van het traject |
|                            | Cross Functional Team Formation (instellen van een multidisciplinair team met alle belanghebbenden ondersteund door IT)                      | +             | Het team is er en deels kan het team ook zelf IT zaken oppakken  |
|                            | Project Roadmap (flexibele roadmap met identificatie van belangrijkste activiteiten, samen met tijdlijnen, aangewezen personen en mijlpalen) | ++            | Er wordt Scrum gewerkt, dus iets dergelijk is ook aanwezig   |
| Data analytics             | Data collection and examination (structured/unstructured data, analyse data kwaliteit en data schoning)                                      | --            | Data toegang, data kwaliteit en data governance verdienen meer urgentie  |
|                            | Data analysis and modelling / Data visualization (trends ontdekken, verklaren en voorspellen met behulp van machine learning)                | ++            | Genoeg data scientisten aanwezig, genoeg verschillende technieken die gebruikt worden, ook dankzij externe kennis  |
|                            | Insight generation (transformatie van analyse van gegevens naar  | ++            | Er worden regelmatig proof of concepts opgeleverd  |



|               |   |    |  |
|---------------|---|----|--|
|               | bruikbare inzichten voor de organisatie)  |    |  |
| Implementatie | Integration with IT systems (makkelijke integratie door een robuuste IT architectuur)                                       | -- | Dit gaat nog stroef en kan veel beter.     |
|               | Training people (oplossing leren gebruiken, men is meestal niet gelijk comfortabel met nieuwe tooling en manier van werken) | ++ | Er is een hele goede link met de business. |

Tabel 4.1: Samenvatting resultaten organisatie 1 in relatie tot fases in een DS project

## 4.4. Profiel en resultaten organisatie 2

De organisatie is ongeveer 5 jaar geleden begonnen met DS projecten en voert nu jaarlijks zo een 15 tot 20 projecten uit op jaarbasis. De duur van de projecten is verschillend, maar de langere DS projecten lopen gemiddeld 3 tot 6 maanden, waarbij dus de helft van de tijd zuiver nodig is om te industrialiseren. De manager DS geeft leiding aan een centrale DS afdeling, daarnaast geeft de manager DS ook leiding aan 3 andere teams, namelijk 2 data engineering teams en 1 data integratie team. Ze zitten ook allemaal bij elkaar op dezelfde afdeling. De data scientisten werken op project basis binnen de business teams, maar er zijn ook units binnen de organisatie waar data scientisten vast werkzaam zijn. De centrale DS bestaat uit een tiental data scientisten, met in ieder geval 2 data scientisten die senior zijn, er zijn verder nauwelijks junioren op de afdeling. Verder wordt er binnen de organisatie voornamelijk gebruik gemaakt van een Scrum, een Agile procesmanagement methodologie, maar ook Kanban wordt gebruikt. De codes waarop de meeste quotes betrekking hebben, zijn voor organisatie 2:

1. Bedrijfscultuur
2. Organisatie DS
3. Management support
4. Proces management methodologie
5. DS expertise op bestuursniveau
6. Start DS project

De code bij 2 is wel weer te verwachten, maar de overige zijn zeker interessant om extra te onderzoeken. Verder hebben een aantal quotes betrekking op de volgende codes, die alleen bij deze organisatie ter sprake zijn gekomen, dat zijn:

- Leidend worden
- Standaardisatie in technologie
- Uniformiteit in manier van werken
- Vrijheid in tooling
- Hybride DS organisatie

Bovenstaande genoemde relevante codes zullen extra aandacht krijgen en besproken worden in de resultaten op basis van de 2 afgenomen interviews bij deze organisatie.

### Resultaten op basis van het interview met de Manager DS bij organisatie 2 (MDS)

De MDS is de supplier van de DS en engineering skills, dus eigenlijk vergelijkbaar met de GMO van organisatie 1. Hij is redelijk tevreden over hoe de DS projecten nu bij hun lopen, dit is zeker niet altijd zo geweest. Daar waar in het verleden, proof of concepts vaak niet opgeschaald werden vanwege andere KPI's of plannings van business teams, is zijn afdeling tegenwoordig in ieder geval

bij aanvang standaard bij elk project betrokken. Dit heeft te maken met hoe ze nu georganiseerd zijn, namelijk als een centrale afdeling en voorheen als een innovatie afdeling met een zwakkere link naar de verschillende business teams toe. In deze organisatie worden met name 2 proces management methodologieën gebruikt, namelijk Scrum en nummer 2 is Kanban. De MDS gaf aan dat dit over het algemeen wel redelijk werkt, maar dat men wat hem betreft flexibel zou moeten zijn in te gebruiken proces management methodologieën en heel goed zou moeten kijken wat er wel werkt en wat niet. Zo antwoordde hij op de vraag of ze binnen hun team gebruik maken van een PMM:

*“Nou alleen voor DS zou ik het geen Scrum willen noemen, eerder Kanban, daar komt het meer op neer. Ik heb heel veel gescrumd, ik was ook Scrum master in 1 van mijn vorige rollen, nog voordat het data avontuur begon en ik zie het niet als zo geschikt voor DS. Ik vind het te veel afleiden eigenlijk als je constant aan het plannen, pokeren en andere rituelen aan het doen bent, je moet het namelijk vooral gewoon gaan doen en in ieder geval gewoon aan de slag gaan” (MD).*

Als positief punt noemde hij de goede communicatie bij Scrum, als negatief punt noemde hij de vele rituelen bij een Scrum werkwijze die erg kunnen afleiden, afhankelijk van het type analyse dat gedaan moet worden door de data scientists. Verder is hij in ieder geval erg tevreden over hun data governance board, dat bestaat uit een stuurgroep, waarbinnen beleid en verantwoordelijkheden en eigenaarschap rondom data is vastgesteld, zoals de aanstelling van belangrijke stakeholders in de organisatie, data eigenaren en data stewards. Deze board draagt zo enorm bij aan de snelheid van knopen die gehakt moeten worden rondom bijvoorbeeld het gebruik van bepaalde data of bij het gebruiken van DS oplossingen. Er is ook genoeg ruimte tot verbetering, zo is de MDS momenteel erg bezig uit te zoeken hoe hij met zijn organisatie nog volwassener kan worden met behulp van zijn centrale DS afdeling en hoe hij met zijn afdeling nog meer de stempel zou kunnen drukken in een leidende rol in de organisatie. Hij ziet hier met name ook taken weggelegd voor het hoger management. Aan de ene kant verwacht hij nog meer top-down sturing, zodat hij bijvoorbeeld kan afdwingen dat infrastructuur zo veel mogelijk gestandaardiseerd mag worden, zodat ze efficiënter kunnen worden bij het in productie brengen van DS oplossingen. Maar daarvoor is het nodig dat de data ‘literacy’ op bestuursniveau omhoog gaat, dus dat het bestuur ook beter weet wat nu wel of niet kan met DS en dat ze ook meer het belang gaan zien van investeren in DS, maar ook in data kwaliteit en slimme opslag van bijvoorbeeld ongestructureerde data. Zo gaf hij aan:

*“... dat de hoge leiding ook echt goed snap wat wel kan en wat niet kan met artificial intelligence en DS en ook hoe ze daar het beste op kunnen gaan sturen. Op een bepaald niveau wil je toch een soort van bullshit detector zijn en dan is het echt de vraag of de hoge leiding op dit moment daartoe in staat is en dat is even de lijn die ik nu voor ogen heb. Dus dat is het level waar ik graag nog heen zou willen gaan. Of terwijl heel kort samengevat, ik vind dat de data literacy van ons bestuur echt fors omhoog moet, zodat ze echt gaan snappen en gaan kunnen handlen naar onze stukken en ik kan me alleen al niet anders voorstellen dat daarom ook betekent dat ons team nog verder wordt uitgebreid of groter gaat worden” (MD).*

Hij is ook een grote voorstander van een uniforme manier van werken, dus van standaardisatie en volgens een vaste methodiek werken. Zo zei hij:

*“... maar je hebt ook gewoon het feit dat het je gewoon niet gaat helpen in je efficiency als je allerlei verschillende tooling gaat gebruiken. Als de 1 met Microsoft Teams aan het samenwerken is, heeft het geen zin als de ander het allemaal in Gitlab gaat doen. Dat soort dingen, het lijken een beetje*

*simpele dingen, dat zie ik nog wel als de volgende stap voor onze organisatie in onze volwassenwording. Dan gaat het ook allemaal makkelijker naar productie toe” (MD).*

Tot slot vindt hij dat het optimaliseren en onderhouden van bestaande DS oplossingen soms onderbelicht wordt, omdat het accent wel heel erg ligt op nieuwe DS oplossingen bij de organisatie. Daar zei hij over:

*“Het is constant ontwikkelen, ontwikkelen, ontwikkelen ontwikkelen van nieuwe taken en er gaat bijna geen aandacht uit naar het optimaliseren van die bestaande taken. En dat is iets waarvan ik vind dat je nog niet helemaal volwassen bent in je data organisatie. En als dat nou 1 van de dingen is, als onze operations officer daar nou nog wat meer tijd en aandacht aan gaat geven, dan gaan wij daar gewoon meer voordeel op pakken” (MD).*

### **Resultaten op basis van het interview met de Senior Data Scientist (SD)**

De SD is verantwoordelijk voor het toepassen van DS technieken om zo waarde te genereren voor de organisatie. Naast deze verantwoordelijkheid wonen de seniors ook technische overleggen bij, waarbij er bijvoorbeeld geschakeld wordt met data architecten. Tot slot organiseren de seniors ook data events, om de rest van de organisatie (belangrijke stakeholders) mee te nemen door middel van opleiding en communicatie over nieuwe gebruikte technologieën, door de data scientisten. Dit wordt ook wel de ‘data academy’ genoemd en draagt ook zeker bij het worden van een data gedreven organisatie. Opvallend is direct het verschil in invalshoek van aan de ene kant de MDS en aan de andere kant de SD. De SD spreekt meer over het DS proces zelf, de tooling die hij gebruikt, de problemen die hij ondervindt in het DS proces. Zo krijgt hij geregeld te maken met een situatie, waarbij het voor hem duidelijk is dat het business probleem nog niet duidelijk is en dat de oplossing niet per se ook een DS oplossing hoeft te zijn. Hij geeft aan dat de business vaak wil ‘rennen’, terwijl het juist belangrijk is om zaken goed door te spreken en te ‘scopen’ en dat er diverse aannames gedaan zijn, die eerst uitgedaagd moeten worden. Hij geeft verder aan dat de door hun gebruikte proces methodologie zoals Scrum handig is bij het samenwerken met de business, maar dat het aan de andere kant heel veel tijd vreet vanwege diverse rituelen en talloze overleggen. Vaak kennen de mensen in een interdisciplinair of multidisciplinair team elkaar nog niet zo goed, wat dus ook veel tijd kost. Bovendien geeft hij aan dat bijvoorbeeld een exploratieve data analyse zich moeilijk laat managen in een sprint en wellicht beter niet gedaan kan worden in een business team (eerder met een collega data scientist) en zeker niet op een Scrum wijze. Daar zei hij over:

*“Veel gebeurt binnen mijn organisatie op de ‘Scrum’, ‘Agile’ manier, dat is niet mijn favoriete methodiek om DS te doen. Zoals ik al zei, voor een exploratieve analyse is dat heel moeilijk te plannen. Omdat je niet weet wat je allemaal gaat tegenkomen, als je überhaupt nooit naar die data hebt gekeken. Hoe lang je erover gaat doen om die data goed schoon te krijgen, op te schonen om het uiteindelijk in een model te stoppen. Daar gaat echt wel tachtig procent van je tijd in zitten” (SD).*

*“Volgens mij werkt Scrum en Agile goed, als je precies weet wat je precies gaat doen of moet gaan doen en Scrum komt oorspronkelijk uit software development waarbij je alles in kleine taakjes kan opschalen en van hier schrijf deze vijf regels code en jij schrijft die vijf regels code en aan het einde van de week hebben we samen honderd regels code geschreven en dat werkt voor een deel voor als je DS gaat productionizen, want dan ga je veel meer op een software development achtige manier te werk en ga je veel minder beetje dingen in elkaar hacken en modellen uitproberen van alles en nog wat doen om het beste eruit te pompen. Hoewel dat ook wel moet gebeuren tijdens productionizen,*

*maar dan doe je dat, dan ga je dat meer standaardiseren, daarvoor is het denk ik wel een prima oplossing” (SD).*

Iets dat opvalt is de tevredenheid in de vrijheid in te gebruiken technieken en tooling, dit wordt ook gewaardeerd door de data scientist, maar dit staat wel haaks op de standaardisatie die de MDS graag ziet. Iets anders dat de SD noemde en wel bevestigd wordt door de MDS, is dat het nog een beetje ontbreekt aan visie op bestuursniveau, over waar te staan bijvoorbeeld over 3 jaar wat DS, IT en business afdelingen betreft. Iets anders, dat ook bevestigd wordt door de SD en waar de SD content mee is, is dat er steeds meer commitment in de vorm van meer vrijgemaakte tijd en energie door de business is gekomen, terwijl voorheen de business vaak nog schermde met hun eigen prioriteiten of KPI's. De SD geeft aan dat dit komt door verbeterde sponsorship van het hoger management. De SD is ook tevreden over het feit dat er altijd minimaal 2 data scientists werken aan 1 project, waardoor je met elkaar kunt sparren. Daar zei hij over:

*“Maar dat is meer omdat we vanuit ons, vanuit centrale DS eigenlijk altijd minstens met twee data scientists aan een project werken. Je moet iemand hebben die dezelfde taal met je spreekt en waar je niet allemaal dingen nog moet uitleggen voordat je de diepte in kan” (SD).*

Wat verder nog opvalt is dat hij aangeeft dat zo min mogelijk afhankelijkheden in een multidisciplinair team zich uitbetalen in meer efficiency bij het industrialiseren van DS oplossingen, door middel van end-to-end teams. Hier ziet hij duidelijk een positieve impact op het naar productie brengen van DS oplossingen, wat hij in die vorm (end-to-end, alle benodigde skills) onlangs ook persoonlijk ervaren heeft. Zo is er bij organisatie 1 ook een multidisciplinair team, maar is dit in feite niet een end-to-end team, omdat er bijvoorbeeld afhankelijkheden zijn met de IT afdeling. Tot slot noemt hij een interessant potentieel verbeterpunt voor wat betreft de organisatie van DS. Dit is iets waar hij de laatste tijd vaak over denkt maar intern nog niet besproken heeft en dat is dat hij denkt dat organisaties met zowel een centrale DS organisatie als ook data scientists in de business elkaar kunnen versterken door vanuit centraal perspectief een strategie op te zetten. Daarover zei hij:

*“Maar juist de hybride manier van organiseren is denk ik wel iets waarvan wij denken, daar krijg je the best of both worlds. Een centrale team die het overzicht kan houden van wat er allemaal gebeurt en verschillende teams elkaar kunnen koppelen en met elkaar dat kunnen versterken. En vanuit centraal perspectief ook een soort van strategie kunnen opzetten. Teams die meer in een bepaalde business unit zitten, die kijken vaak ook niet verder dan hun business units. Zij zijn heel erg goed in waar hun business unit voor bedoeld is, maar krijgen ook niet de mogelijkheid om zich veel te ontwikkelen. Om bijvoorbeeld in even 2 weken een deep learning model te gaan trainen. Omdat dat niet binnen hun Scrum valt en niet binnen de KPI's valt die de managers hanteren. Die ze om de zoveel tijd goed moeten afstemmen. Dat soort echte DS werk zou ik willen zeggen, dat komt er vaak niet van in die teams” (SD).*

Zie hieronder de genoemde resultaten samengevat voor organisatie 2 in tabelvorm. Zaken die goed gaan en beter kunnen in relatie tot de fases in een DS project.

| Hoofdfases                 | Subfases                                   | Organisatie 2 | Toelichting   |
|----------------------------|--|---------------|---|
| Strategische voorbereiding | Business problem (helder begrip vraagstuk) | -             | Dit vereist behoorlijk wat energie, vaak zijn er door de business allerlei aannames |

|                |  |    |   |
|----------------|--|----|---|
|                |  |    | gedaan, die uitgedaagd moeten worden  |
|                | Research (identificatie mogelijke oplossingen en bepalen of bestaande infrastructuur voldoet)  | ++ | Er wordt goed gekeken welke oplossing nodig is en of er überhaupt sprake is van een DS oplossing  |
|                | Cross Functional Team Formation (instellen van een multidisciplinair team met alle belanghebbenden ondersteund door IT)                      | ++ | Er is vaak zelfs sprake van end-to-end teams  |
|                | Project Roadmap (flexibele roadmap met identificatie van belangrijkste activiteiten, samen met tijdlijnen, aangewezen personen en mijlpalen) | ++ | Dit is aanwezig, bovendien wordt er ook vaak Agile en Scrum, Kanban gewerkt,  |
| Data analytics | Data collection and examination (structured/unstructured data, analyse data kwaliteit en data schoning)                                      | -  | Er is uitgesproken dat er meer geïnvesteerd mag worden in data kwaliteit en slimme data opslag.   |
|                | Data analysis and modelling / Data visualization (trends ontdekken, verklaren en voorspellen met behulp van machine learning)                | ++ | Er zijn senior data scientisten aanwezig, verder ook genoeg verschillende technieken die gebruikt worden                                    |
|                | Insight generation (transformatie van analyse van gegevens naar bruikbare inzichten voor de organisatie)                                     | ++ | Er worden regelmatig proof of concepts opgeleverd   |
| Implementatie  | Integration with IT systems (makkelijke integratie door een robuuste IT architectuur)  | -  | Ondanks dat er weinig klachten gehoord worden, is de doorlooptijd hetzelfde als die van organisaties die aangeven hier moeite mee te hebben |
|                | Training people (oplossing leren gebruiken, men is meestal niet gelijk comfortabel met nieuwe tooling en manier van werken)                  | +  | Er is een hele goede link met de business.  |

Tabel 4.2: Samenvatting resultaten organisatie 2 in relatie tot fases in een DS project

Door de codes tussen de MDS en SD te vergelijken (zie bijlage 6), wordt weer de invalshoek tussen die van de MDS en SD bevestigd. Terwijl aan de ene kant de MDS meer strategisch bezig is om zijn afdeling en organisatie op een hoger niveau te tillen, ligt de focus van de SD met name op het DS proces en komen de volgende codes vaker voor bij de SD dan bij de MDS:

1. Proces management methodologie
2. Start DS project
3. Relatie met de business

Door in te zoomen op de quotes die betrekking hebben op bovenstaande komt duidelijk naar voren dat de SD vindt dat de huidige gebruikte proces management methodologie verre van optimaal is

voor het uitvoeren van DS projecten. Ondanks dat hij ook de voordelen ziet van de door hun gebruikte Scrum methodiek, met name in de relatie met de business, geeft hij aan dat doordat er met interdisciplinaire teams gewerkt wordt en men veelal elkaar niet goed kent, het veel tijd kost om elkaar te kennen en dezelfde taal te spreken. Zo gaf hij aan:

*“Het hangt er vanaf met wat voor een business team je te maken hebt. Als zij heel erg gewend zijn allemaal om Scrum en Agile te werken, dan vinden ze het heel spannend om dat niet te doen. En op zekere hoogte ben ik er wel voor, dat je een geregeld moment hebt om het er met elkaar over te hebben. Maar als ik nu in een team zit, waarom ik het nu dus zo ook druk heb, we zijn net begonnen met het project, is een geheel nieuw team die vanuit drie verschillende teams samengesteld wordt. Dan moet je heel veel overleggen, in het begin, om een beetje op hetzelfde niveau te komen. Wat weet die persoon, wat weet die persoon, praten we dezelfde taal? Wat bedoelen we hiermee en daarmee? En daar gaat heel veel tijd in zitten. Daarnaast krijg je ook nog eens de normale soort van Agile en Scrum dingen, van we moeten een sprint planning doen, we moeten een sprint review doen aan het einde van de eerste week. Hoe hebben we het gedaan? Dat moeten we presenteren aan onze stakeholders. Zijn die stakeholders echt geïnteresseerd in wat je na één week voor elkaar hebt gekregen? Moet je echt na een week al een retrospective inplannen van anderhalf uur? Als je daar met tien mensen in zit, dan brandt je zo door een week werkuren heen.” (SD).*

Bovendien vindt hij dat bepaalde analyses zoals een exploratieve analyse zich helemaal niet lenen voor de door hun gebruikte Scrum methodiek, vanwege de enorme onbekendheden in een dergelijke analyse. Opvallend is dat Scrum als de norm wordt gezien om te werken in de organisatie en dat er daarom niet gedacht wordt aan het feit dat de Scrum methodiek zich zeker ook voor leent om aanpassingen te doen, waar de data scientisten gebaat mee kunnen zijn.

#### 4.5. Profiel en resultaten organisatie 3

De organisatie is ongeveer 3 jaar geleden begonnen met DS projecten en voert nu jaarlijks zo een 12 tot 15 projecten uit op jaarbasis. De duur van de projecten is verschillend en is ook afhankelijk aan wie het gevraagd wordt. Eén van de geïnterviewden geeft aan dat de projecten gemiddeld 12 tot 16 weken lopen, de ander geeft aan dat de projecten 3 tot 6 maanden kunnen duren voor de in ieder geval wat langere DS projecten, waarbij dus de helft van de tijd zuiver nodig is om te industrialiseren. De organisatie heeft een centrale DS afdeling, dat bestaat uit 15 data scientisten. Deze afdeling wordt aangestuurd door de manager DS. De data scientisten werken in 4 teams onder leiding van 4 principal data scientisten op hun eigen afdeling op project basis. De principal data scientist is een ervaren scientist, die ook de rol van projectleider heeft en dus verantwoordelijk is voor stakeholders management en project management. De centrale DS afdeling werkt dus vaak voor de business teams, maar niet binnen een business team. Voor een project moet de business ook 2 dagen per week beschikbaar zijn voor vragen en dergelijke. Er is ook een engineering team, dit team wordt aangestuurd door een platform owner. Verder zitten beide teams op dezelfde afdeling. De platform owner en de manager DS vallen wel beiden onder dezelfde director insights. Naast de centrale DS afdeling zijn er ook business units binnen de organisatie waar data scientisten vast werkzaam zijn. Hoewel er binnen de organisatie door de business teams gebruik gemaakt wordt van Scrum, is dit dus niet van toepassing voor de centrale DS afdeling. De codes waarop de meeste quotes betrekking hebben, zijn voor organisatie 2:

1. Bedrijfscultuur
2. Proces management methodologie
3. Management support

#### 4. Participatie van de business

De code bij 1 is wel weer te verwachten, maar de overige zijn zeker interessant om extra te onderzoeken. Opvallend is verder door de verschillende organisaties te vergelijken dat er bij deze organisatie niet wordt gesproken over de belangrijkheid van senioriteit van data scientisten en het zelfstandig in productie brengen van DS oplossingen, wat wel besproken wordt bij de andere 2 organisaties. Bovenstaande genoemde zaken zullen extra de aandacht krijgen en besproken worden in de resultaten op basis van de 2 afgenomen interviews bij deze organisatie.

#### **Resultaten op basis van het interview met de Manager Analytics & Decision Support (MADS)**

De MADS is verantwoordelijk voor het binnenhalen van projecten, het prioriteren van projecten en de 'staffing' van projecten. Tevens monitort ze de projecten en bewaakt daarmee ook de voortgang van projecten en ziet er op toe dat er aan het einde er een goede presentatie is met een goed duidelijk advies. Opvallend is dat ze ook om de zoveel tijd zelf steekproefsgewijs projecten oppakt. Hiermee is ze ook goed op de hoogte van uitdagingen en succesfactoren van DS projecten die binnen haar afdeling spelen. Wat gelijk opvalt aan deze organisatie is dat zij eigenlijk zelden binnen een business team opereren, dit in tegenstelling tot organisatie 1 en organisatie 2. Dit betekent ook dat ze niet echt samenwerken met de engineers, hoewel ze daar wel veel contact mee hebben. Bovendien hebben ze als harde eis bij de business, dat wanneer ze een project uitvoeren voor de business, de business dan ook tenminste 2 dagen per week beschikbaar is. Er wordt ook enorm doorgevraagd bij de business, om te achterhalen wat nu de business vraag is. Beschikbaar zijn, betekent overigens niet dat de business naast de data scientisten zitten, maar wel tijd hebben voor een stand-up en als er bijvoorbeeld vragen zijn en ze dan ook gedurende die 2 dagen snel een antwoord sturen op vragen, daarnaast wordt ook commitment gevraagd dat ze ook iets met het advies doen van de data scientisten. In de praktijk zie je echter dat het toch lastig plannen is met de business. Daarnaast zou je denken dat dit tot veel minder verzoeken voor DS oplossingen van de business zou kunnen leiden, maar er komen juist veel projecten tot stand op aanvraag van de business. Dit betekent dat de afdeling inmiddels goed bekend is in de organisatie en dat business afdelingen binnen de organisatie een enorme bereidheid hebben om te werken met DS oplossingen. De reden voor dit succes ligt in het snel iets laten zien aan de business en zo ook snel feedback te krijgen om juist met behulp die feedback mensen mee te krijgen. Opvallend is dat ze verder als reden voor succes een goede samenwerking tussen alle stakeholders, vergelijkbaar met die van een multidisciplinair team noemt, ondanks dat ze zelf niet volledig meedraaien in een multidisciplinair team. Zo zei ze:

*“Ja, wat vooral bijdraagt, is een multidisciplinair team en dan heb ik het over data scientisten en echt de business en eventueel ook de operatie, dat je echt alle mensen die straks iets te maken hebben met de oplossing, dat je die meteen aan boord hebt, trouwens ook de data engineers, als het dan gaat om de oplossing te implementeren. Die heb je keihard nodig, die samenwerking en het is nodig in mijn ogen om echt snel iets te laten zien, dus gewoon demo's te hebben en ook om na een aantal sprints te laten zien, dit is wat we neergezet hebben, om zo steeds de mensen mee te krijgen en zo ook feedback te krijgen voor je oplossing en dan is data kwaliteit ook heel belangrijk, omdat je daar als bedrijf ook van leert” (MADS).*

Iets dat volgens de MADS beter kan en ook nodig is voor succes, is de beschikbaarheid van goede data, dus urgentie voor datakwaliteit. Zij geeft aan dat het niet alleen belangrijk zou zijn om meer data KPI's te introduceren voor datakwaliteit, maar dat bedrijfsprocessen hierop ook aangestuurd zouden moeten worden. Dit houdt in dat iedereen in de organisatie weet hoe belangrijk data kwaliteit is en welke zaken dit negatief kunnen beïnvloeden. In kader hiervan, om dit te

bewerkstelligen (meer aandacht voor datakwaliteit) zou meer data expertise op senior management niveau heel welkom zijn. Hierover zei ze:

*“Ik denk dat een training op senior management niveau heel belangrijk zou zijn. Ik denk ook dat het belangrijk zou zijn om meer data KPI's te introduceren als een bedrijf, bijvoorbeeld over de datakwaliteit maar ook over de kennis, en dat ook op de vloer” (MADS).*

Tot slot geeft ze aan dat ze verder zelf ook veel aandacht besteed aan een bepaalde DS oplossing of aanpak, die ze reeds ontwikkeld hebben, ook ergens anders kunnen toepassen, vooral omdat DS projecten best wat tijd kunnen kosten.

### **Resultaten op basis van het interview met de Director Insights (DI)**

De DI is verantwoordelijk voor alle data capabilities, zoals data analytics, master data management en business intelligence, binnen de organisatie. Tevens heeft hij ook het laatste woord samen met de MADS over welke DS projecten nu uitgevoerd worden, daarnaast reviewt hij ook de einddocumenten. De DI is zeer tevreden over hoe de DS projecten meestal lopen. Dit is niet altijd zo geweest. Zo geeft hij aan dat het de eerste paar jaren echt moeilijk was, om in beeld te komen bij de business teams. Het was dus moeilijk bij het grootste deel van de organisatie om de connectie te leggen tussen een business probleem en DS. Er is daarom op een bepaald moment ook een trainingsprogramma ontwikkeld om de organisatie meer data gedreven te krijgen en dit heeft de organisatie geholpen om DS in de kern van de organisatie te brengen.

Er wordt bovendien ook maandelijks door de DS afdeling een data event georganiseerd voor iedereen die belangstelling heeft in de organisatie. In dit data event kunnen technieken besproken worden maar ook interessante sprekers in kader van DS uitgenodigd worden. Iets dat volgens de DI bijdraagt aan de adoptie van de business, is dat wanneer een traject afgerond is, men ervoor zorgt dat de business ook de oplossing begrijpt en goed kan toepassen binnen hun processen. Zo zei hij:

*“maar wat lastiger is bij DS projecten is, op het moment dat je dus een traject hebt afgerond, dat de business ook begrijpt wat er nu eigenlijk gebeurd is, omdat als je niet uitkijkt het toch best heel specifiek is zeg maar” (DI).*

De DI geeft aan dat ze verder niet per se aan financiële benefit tracking doen, omdat het ook risico's met zich meebrengt, aangezien men dan meer tijd kwijt kan zijn aan het bewijzen dat iets toegevoegde waarde heeft, terwijl de toegevoegde waarde gewoon beredeneerd kan worden. Sponsorship van hoger management is zeer goed te noemen, aangezien volgens de DI zo een beetje in elk jaarplan, gemaakt vanuit verschillende directieleden, concreet staat dat ze DS nodig hebben. Een aandachtspunt die hij benoemt, is de participatie van de business in DS projecten, ondanks de afspraken met de business daaromtrent. Daar zei hij over:

*“Nou wat ik in ieder geval belangrijk vind en ik weet niet of dat dan beter is, omdat we dat denk ik al best wel goed doen, maar wat belangrijk is dat je de participatie van de business hebt, dus mensen in de business die gewoon domein kennis hebben, dus verstand hebben van logistiek van HR of van Finance, gewoon omdat ze daarin zitten. Dat is essentieel” (DI).*

Een ander aandachtspunt die hij benoemt is de capaciteit op engineering vlak. Opmerkelijk is dat hij capaciteit voorstelt en bijvoorbeeld niet een andere manier van werken op de afdeling, zoals 'continuous integration' en 'continuous development'. De engineering afdeling is namelijk een best



grote afdeling, die ook flink gegroeid is. Zie hieronder de genoemde resultaten samengevat voor organisatie 3 in tabelvorm. De zaken die goed gaan en beter kunnen in relatie tot de fases in een DS project.

| Hoofdfases                 | Subfases  | Organisatie 3 | Toelichting   |
|----------------------------|---|---------------|---|
| Strategische voorbereiding | Business problem (helder begrip vraagstuk)  | +             | De business wordt scherp bevraagd, maar de participatie van de business is een aandachtspunt  |
|                            | Research (identificatie mogelijke oplossingen en bepalen of bestaande infrastructuur voldoet)   | +             | Er is veel contact met de engineering afdeling, waar zaken afgestemd worden   |
|                            | Cross Functional Team Formation (instellen van een multidisciplinair team met alle belanghebbenden ondersteund door IT)                       | --            | Er wordt zelden multidisciplinair gewerkt, terwijl men daar enorm de voordelen van ziet   |
|                            | Project Roadmap (flexibele roadmap met identificatie van belangrijkste activiteiten, samen met tijdslijnen, aangewezen personen en mijlpalen) | +             | De organisatie werkt voornamelijk op Scrum wijze, echter de data scientisten zelf niet, wel worden ze aangestuurd door een principal data scientist, die ook projectleider is |
| Data analytics             | Data collection and examination (structured/unstructured data, analyse data kwaliteit en data schoning)                                       | -             | Er is een datalake voor toegang tot veel data, wel mag er organisatiebreed meer aandacht komen voor datakwaliteit   |
|                            | Data analysis and modelling / Data visualization (trends ontdekken, verklaren en voorspellen met behulp van machine learning)                 | ++            | Genoeg ervaring (principals) aanwezig, goed opleidingsprogramma in samenwerking met een universiteit aanwezig en genoeg verschillende technieken die gebruikt worden          |
|                            | Insight generation (transformatie van analyse van gegevens naar bruikbare inzichten voor de organisatie)                                      | ++            | Er worden regelmatig proof of concepts opgeleverd   |
| Implementatie              | Integration with IT systems (makkelijke integratie door een robuuste IT architectuur)   | -             | Engineering capaciteit wordt het issue genoemd, maar het zou net zo goed aan een niet robuuste architectuur of aan de nieuwe technologie kunnen liggen                        |
|                            | Training people (oplossing leren gebruiken, men is meestal niet gelijk comfortabel met nieuwe tooling en manier van werken)                   | ++            | Er wordt hier bewust tijd aan besteed en heeft enorm de aandacht  |

Tabel 4.3: Samenvatting resultaten organisatie 3 in relatie tot fases in een DS project

Door de codes tussen de MADS en DI te vergelijken (zie bijlage 6), vallen de verschillen tussen de MADS en DI voor de volgende codes op:

1. Proces management methodologie
2. Organisatie DS
3. Urgentie data kwaliteit
4. Relatie met de business

Met name de codes onder 1, 3 en 4 verdienen weer extra aandacht. Door deze quotes te bestuderen, wordt duidelijk dat de MADS het belangrijk vindt om een goede relatie te hebben met alle stakeholders in een project, met name met de engineers en de business en hier draagt een multidisciplinair team dus aan bij. Maar tegelijkertijd draaien ze zelf niet volledig mee in de multidisciplinaire teams. Ook op de centrale DS afdeling zelf wordt niet met Scrum gewerkt, hoewel ze aangeeft dat ze daar in een transitie te zitten en voornemens zijn om dit wel steeds meer te doen, met als doel om toch sneller zaken te kunnen opleveren. Verder valt op dat urgentie voor data kwaliteit niet benoemd door de DI, maar wel door de MADS. Tot slot benadrukt de MADS, dat een goede relatie met de business belangrijk is. Dit wordt dan wel weer bevestigd door de DI, die aangeeft dat ook wanneer een traject DS traject is afgerond, men zich er wel van moet vergewissen dat de business ook goed begrijpt, hoe de data DS oplossing nu precies werkt en hoe het nu precies gebruikt moet worden. Wat ook opviel, was dat er bij deze organisatie verder niet gesproken is over de relevantie van de senioriteit van data scientists, wat op zich niet zo gek is omdat er genoeg senioriteit aanwezig is op de afdeling in de vorm van 4 principals. Dat het zelfstandig in productie brengen van DS oplossingen niet aan bod is gekomen heeft te maken met het feit dat men niet volledig meedraait in een multidisciplinair team en dat de implementatie fase tot de verantwoordelijkheid van de engineers behoort.

## 4.6. Resultaten cross-case analyse

Als we nu de organisaties vergelijken voor zaken die goed gaan bij het uitvoeren van DS projecten en zaken die beter kunnen over de fases heen, dan kan geconcludeerd worden, dat de uitdagingen bij de organisaties liggen in:

1. het rand voorwaardelijk zijn voor data analytics (bijvoorbeeld goede datakwaliteit, slimme opslag van data);
2. de implementatie (hebben van een robuuste architectuur en kennis daarover) Verder kan geconcludeerd worden dat zaken die vooral goed gaan bij de verschillende organisaties, te maken hebben met.

| Hoofd fases    | Subfases  | Organisatie 1 | Organisatie 2 | Organisatie 3 |
|----------------|---|---------------|---------------|---------------|
| Data analytics | Data collection and examination (structured/unstructured data, analyse data kwaliteit en data schoning) | --            | -             | -             |
| Implementatie  | Integration with IT systems (makkelijke integratie door een robuuste IT architectuur)                   | --            | -             | -             |

Tabel 4.4: Resultaten vergelijking van organisaties voor wat betreft de uitdagingen bij organisaties

Zaken die behoorlijk goed gaan bij alle organisaties:

1. Data analytics fase: het concreet toepassen van DS technieken en daar waarde uithalen;
2. Strategische voorbereiding: gebruik maken van een project roadmap;
3. Implementatie: trainen van personeel die met de DS oplossing moet werken.

| Hoofdfases                 | Subfases   | Organisatie 1 | Organisatie 2 | Organisatie 3 |
|----------------------------|--|---------------|---------------|---------------|
| Data analytics             | Data analysis and modelling / Data visualization (trends ontdekken, verklaren en voorspellen met behulp van machine learning)                | ++            | ++            | ++            |
|                            | Insight generation (transformatie van analyse van gegevens naar bruikbare inzichten voor de organisatie)                                     | ++            | ++            | ++            |
| Strategische voorbereiding | Project Roadmap (flexibele roadmap met identificatie van belangrijkste activiteiten, samen met tijdlijnen, aangewezen personen en mijlpalen) | ++            | ++            | +             |
| Implementatie              | Training people (oplossing leren gebruiken, men is meestal niet gelijk comfortabel met nieuwe tooling en manier van werken)                  | ++            | +             | ++            |

Tabel 4.5: Resultaten vergelijking van organisaties voor wat betreft zaken die goed gaan bij organisaties

Verder zijn er nog resultaten die een wisselend beeld geven bij de organisaties.

- Bij organisatie 1 speelt ten opzichte van de andere organisaties dat er te weinig aandacht gegeven wordt bij het begin van het DS traject, aan de technische implicaties van een eventueel DS project;
- Bij organisatie 2 speelt ten opzichte van de andere organisaties, dat er vaak nog goed uitgezocht moet worden, of er überhaupt spraken is van een DS probleem en moeten verschillende aannames, gedaan door de business, nog goed getoetst worden.
- Bij organisatie 3 speelt ten opzichte van de andere organisaties, dat men wel veel waarde hecht aan de samenwerking zoals in een multidisciplinair team, maar zitten de data scientisten, vooralsnog in ieder geval niet of nauwelijks, in dergelijke multidisciplinaire business teams en is het dus veel plannen met de verschillende stakeholders.

| Hoofdfases                 | Subfases  | Organisatie 1 | Organisatie 2 | Organisatie 3 |
|----------------------------|---|---------------|---------------|---------------|
| Strategische voorbereiding | Business problem (helder begrip vraagstuk)  | +             | -             | +             |
|                            | Research (identificatie mogelijke oplossingen en bepalen of bestaande infrastructuur voldoet) | -             | ++            | +             |
|                            | Cross Functional Team Formation (instellen van een multidisciplinair team met alle            | +             | ++            | --            |

|  |                                      |  |  |  |
|--|--------------------------------------|--|--|--|
|  | belanghebbenden ondersteund door IT) |  |  |  |
|--|--------------------------------------|--|--|--|

Tabel 4.6: Resultaten vergelijking van organisaties voor wat betreft zaken die een wisselend beeld geven bij de organisaties

Op basis van de ervaringen van de respondenten van de verschillende organisaties zijn er KSF gevonden, die niet direct toegewezen kunnen worden aan één van de KSF die uit literatuuronderzoek zijn gebleken (bijlage 2: Synthese van kritieke succesfactoren). Dit betekent dat er toch meer KSF zijn, dan die bekend zijn in de literatuur en dat KSF ook nog in ontwikkeling zijn. De KSF die alleen op basis van deze casestudy naar voren zijn gekomen, zijn:

1. Senioriteit in data scientisten en engineers is belangrijk, voor de beginperiode om een vliegende start te maken;
2. Data scientisten in paartjes en in meerdere business teams laten werken, voor kennisborging en kennisontwikkeling;
3. Opleveren van betrouwbare producten is promotie voor nieuwe DS projecten en draagt dus zo ook bij aan het succes van DS projecten;
4. Importantie van meer data expertise op bestuursniveau, maar met name de relatie die het ook heeft met verschillende andere KSF, want meer data expertise heeft namelijk in potentie een effect op heel veel zaken, zoals verbeterde management support, willen investeren in data kwaliteit en slimme data opslag, maar ook het willen experimenteren met DS. Dus eigenlijk een effect op de complete bedrijfscultuur en dit zorgt ook voor visie van bestuurders, waardoor DS in de organisatie volwassener kan worden;
5. Open staan voor veranderingen in de gebruikte proces management methodologie en het meenemen van ervaringen van data scientisten hierin;
6. Standaardisatie van architectuur en tooling, om in potentie DS oplossingen eerder naar productie te brengen, weeg dit wel goed af tegen de tooling die data scientisten graag willen gebruiken. Bij alle 3 de organisaties duurt het gemiddeld 3 extra maanden om een DS oplossing naar productie te brengen, wellicht kan het lonen om process management methodologieën voor software engineering zoals 'continuous integration' en 'continuous delivery' (methodiek om kort cyclisch en vaak software naar productie te brengen) te adopteren, die het mogelijk maken om sneller en frequenter DS oplossingen naar productie te brengen;
7. Hybride organisatie van DS: als een organisatie zowel een centrale DS afdeling als vaste data scientisten in business teams heeft, biedt een centrale afdeling de mogelijkheid om de vaste data scientisten in aanraking te laten komen met technieken waar ze normaal gesproken van uit hun eigen business niet mee in aanraking zouden komen, tegelijkertijd kan er enorm geprofiteerd worden van de domeinkennis van een data scientist die vast werkt in een bepaalde business unit;
8. Genoeg aandacht besteden aan het onderhouden van bestaande DS oplossingen, om DS oplossingen betrouwbaar te houden.

Er zijn verder veel KSF uit het literatuuronderzoek, die overigens geen betrekking hadden op Nederlandse bedrijven, die direct of indirect bevestigd zijn door de ervaringen van de verschillende respondenten bij de verschillende organisaties (vastgesteld door de interviews). Op basis van het literatuuronderzoek is een synthese uitgevoerd voor alle gevonden KSF. Veel KSF in deze synthese zijn tijdens de interviews aan de orde gekomen, maar enkelen ook niet. Het valt op dat een aantal van de KSF die niet tijdens de interviews aan de orde zijn gekomen, een impact kunnen hebben op

de implementatie fase van een DS project, terwijl de implementatie nou juist ook een zorgpunt is bij alle organisaties. De KSF die niet genoemd zijn:

- Een robuuste flexibele IT architectuur
- Beschikbaarheid van documentatie ter behoeve van techniek
- Continue bewaking van prestaties en regelmatige bijstelling van het systeem om aan operationele doelstellingen te voldoen
- Systeem veiligheid
- Bewustwording bij IT-personeel dat het creëren en onderhouden van de documentatie van nieuw geïmplementeerde systemen van vitaal belang is
- Vroegtijdig kunnen reageren wanneer gebruikte systemen moeten worden uitgebreid

Verder is er nog een kritieke succesfactor, die niet genoemd is en dat is:

- Analyse van vaardigheden die ontbreken bij een DS project

Over het algemeen werd aangegeven dat alle kennis wel aanwezig is om DS trajecten uit te voeren, maar veel kennis relatief nieuw is en nog niet zoveel affiniteit is met bepaalde zaken, zeker als het gaat om kennis rondom de implementatie fase van een DS project.

De KSF die niet aan de orde zijn gekomen tijdens de interviews, zijn in het rood in de synthese aangegeven in bijlage 2: Synthese van KSF. De KSF zijn ook te zien in onderstaande tabel 4.7 (KSF op basis van literatuuronderzoek, die niet genoemd zijn tijdens de interviews).

| Technology   | People  | Process  |
|--|---|--|
| document collection is a significant problem, only with the right documents provided the needed answers can be found within the data | IT staff needs to be aware that creating and maintaining the documentation of the new implemented systems is vital, especially because these are innovative systems | skill gap analysis   |
| adapt architectural principles   |   | the growth of analysis data has to be forecasted to be able to react early when used systems need to be expanded |
| performance has to be monitored continuously and the system needs to be tuned on a regular basis to meet operational goals           |   |  |
| system security  |   |  |

Tabel 4.7: KSF op basis van literatuuronderzoek, die niet genoemd zijn tijdens de interviews

## 5. Discussie, conclusies en aanbevelingen

### 5.1. Discussie – reflectie

Eerder onderzoek heeft aangetoond dat de meeste redenen dat een DS project faalde, meestal niet technisch van aard waren, maar vaak juist projectmatig en organisatorisch van aard waren (Becker,

2017). Dit onderzoek kan dat grotendeels bevestigen, maar niet helemaal. Enkele belangrijke aandachtsgebieden bij de organisaties vallen inderdaad onder niet technische redenen, zoals de importantie van een goede relatie met alle belanghebbenden in een DS project. Daarnaast is er behoefte aan meer data expertise op bestuursniveau, met het idee dat dit resulteert in meer management support en meer aandacht voor data kwaliteit, data governance, slimme opslag van data en een visie voor DS. Bovendien is het zo dat het thema waar het meest over gesproken is door de organisaties, ook de bedrijfscultuur van de organisatie is, dus in hoeverre de organisatie al data gedreven is. Echter een belangrijk aandachtspunt, die ook duidelijk naar voren is gekomen bij alle 3 de organisaties, betreft de implementatie fase van DS projecten. In het bijzonder de lange doorlooptijd van het in productie brengen van DS oplossingen. De redenen daarvoor liggen wel degelijk ook in de technische sfeer. Hoewel bij geen van de organisaties is aangegeven dat de lange doorlooptijd heeft geleid tot het falen van DS projecten, is wel bij elke organisatie merkbaar dat dit een knelpunt betreft. Alle 3 de organisaties die onder de loep zijn genomen, gebruiken DS oplossingen voor de organisatie zelf, zoals de optimalisatie van allerlei interne processen. Het is daarom achteraf gezien ook interessant om organisaties te interviewen, die DS oplossingen maken voor andere partijen. Immers wordt dan duidelijk of deze organisaties andere doorlooptijden kennen, voor het naar productie brengen van DS oplossingen. Aangezien we tegenwoordig in een tijd leven waarin de behoefte van organisaties snel veranderen, kan ik mij voorstellen dat de doorlooptijden bij deze organisaties korter zijn en dan wordt ook duidelijk hoe deze organisaties dergelijke kortere doorlooptijden bewerkstelligen.

Andere onderzoeken hebben aangetoond dat er een behoefte is aan een proces, dat houvast biedt bij het managen van een DS project (Miller, 2018; Saltz et al., 2018; Saltz & Shamshurin, 2016), daarnaast dat er een projectplanning en een duidelijke project scope en doel nodig is (Miller, 2018). Bij dit onderzoek wordt in ieder geval bevestigd dat er vaak een proces management methodologie gebruikt wordt, die gehanteerd wordt door business teams en dus bekend is bij organisatie, dit is vaak een Agile methodiek. Er was echter 1 organisatie, namelijk organisatie 3, die zelden meedraaide met business teams. In het verleden werkten ook nog niet alles business teams met de Scrum methodiek, waardoor dit wellicht ook niet nodig was. Een echt bevredigend antwoord heb ik helaas niet gekregen, waarom ze nog steeds zelden met de business meedraaien, terwijl inmiddels bijna alle afdelingen binnen de organisatie met Scrum werken en ze nota bene meer behoefte hebben aan schakelmomenten met de business. In deze organisatie werken data scientisten overigens onder leiding van een principal data scientist, die tevens projectleider is. Het zou interessant zijn om te weten welke methodologie hierbij gehanteerd wordt. In kader hiervan zou het handig geweest zijn om ook een principal data scientist geïnterviewd te hebben, zodat op de gehanteerde methodiek doorgevraagd wordt, helaas liet de relatief beperkte tijd waarin dit onderzoek afgerond moet worden, het niet toe om een derde persoon bij deze organisatie te interviewen. Vanwege de zojuist genoemde beperkte tijd in het huidige onderzoek zijn er voor deze casestudy 3 organisaties geselecteerd en bij elke organisatie zijn er 2 interviews afgenomen en dit brengt limitaties met zich mee. Het beperkte aantal organisaties kan er in theorie voor zorgen dat de resultaten niet goed generaliseerbaar zijn, maar omdat er toch behoorlijk wat overeenkomsten tussen de organisaties gevonden zijn, kan wel verwacht worden dat zelfs al zouden de resultaten niet helemaal generaliseerbaar zijn, de bevindingen toch voor veel andere organisaties met een vergelijkbaar profiel moeten gelden.

Een ander onderzoek gaf aan, dat er tot nu toe relatief weinig aandacht geweest is voor de samenwerking in een DS project en voor de uitdagingen die een team heeft, die gezamenlijk een DS project uitvoert (Saltz, Crowston, & Shamshurin, 2017). Dat wordt ook bevestigd door dit onderzoek. Alle 3 de organisaties geven aan dat een goede samenwerking met alle belanghebbenden in een DS project noodzakelijk en belangrijk is. Het gaat hierbij dus met name om business experts

(bijvoorbeeld product owner, business analyst, data eigenaar en andere relevante kennishouders bij de business), data scientisten en engineers (data integratie experts en ontwikkelaars).

Door 2 van de organisaties wordt Scrum, een Agile methodiek gebruikt (Scrum vs Agile, wat is het verschil eigenlijk?, 2020) en bij organisatie 3 wordt met name veel ad hoc afgestemd met de verschillende belanghebbenden. In ieder geval wordt er door organisatie 2, met name door de daar werkende data scientist (SD), maar ook door de manager (MD) wordt aangegeven, dat de Scrum methodiek niet altijd plezierig werkt. Bovendien ervaart organisatie 3 heel veel moeite bij het managen en plannen met de verschillende belanghebbenden. Het is mij echter niet duidelijk geworden of er genoeg aandacht was voor de huidige knelpunten bij de samenwerking, bijvoorbeeld in een 'retrospective' (reflectiemiddel) bij de Scrum methodiek en in hoeverre knelpunten bij samenwerking dus een aandachtsgebied is bij het uitvoeren van DS projecten. Uit onderzoek (Saltz, 2018) is gebleken dat een proces management methodologie (PMM) niet zomaar geaccepteerd wordt door een team en dat hierbij sleutel acceptatiefactoren een rol spelen zoals de geschiktheid, complexiteit en relatieve voordelen (bijvoorbeeld effectiviteit, communicatie) van een methodologie. Ik had graag aanbevelingen willen doen, hoe dan een optimale PMM er uit zou moeten zien en waar deze dan aan zou moeten voldoen, maar dan had ik achteraf gezien meer data scientisten moeten interviewen en wellicht had het perspectief van de data scientist in een DS proces dan meer aan bod mogen gekomen. Want met name de SD is ingegaan op knelpunten bij de Scrum methodologie. Dat neemt verder niet weg dat ook op basis van dit éne interview met de SD het wel duidelijk werd, dat een gehanteerde PMM, die bijvoorbeeld de norm is in een organisatie, flexibeler toegepast kan worden, zodat ook data scientisten zich comfortabeler voelen bij een dergelijke PMM. Niet voor niets ontstaan er steeds meer hybride PMM-en, die een combinatie zijn van bestaande software engineering methodologieën met enkele aanvullende elementen uit KDDM of andere raamwerken of methodologieën (Saltz et al., 2018).

Heel veel zaken die benoemd zijn door de organisaties, in relatie tot de ervaringen die deze organisaties hebben gehad bij het uitvoeren van DS projecten, bevestigen het uitgevoerde literatuuronderzoek rondom KSF. Bovendien laat dit onderzoek zien dat de gevonden KSF van uit literatuuronderzoek, die eigenlijk alleen betrekking hadden op onderzoek bij niet Nederlandse bedrijven, ook van toepassing zijn voor Nederlandse organisaties. Zo zijn een aantal zaken die bijvoorbeeld sterk naar voren zijn gekomen als kritieke succesfactor, die ook vanuit literatuuronderzoek bekend zijn, bijvoorbeeld:

- Bedrijfscultuur (bijvoorbeeld data expertise op bestuursniveau, DS community)
- Urgentie voor data kwaliteit
- DS visie
- Management support
- Multidisciplinair team
- Data governance
- Data toegankelijkheid
- Proces management methodologie
- (Investeren in) technische middelen / know-how

Verder zijn bedrijven niet alleen bekeken en vergeleken op zuiver uitdagingen of zaken die goed zijn gegaan, maar door ook echt in te zoomen op de fase waarin de uitdagingen of de positieve zaken betrekking hadden. Een belangrijke bevinding hierbij was dat de implementatie fase van een DS project genoeg ruimte heeft om verbeterd te worden en dat dit bovendien een kritieke succesfactor is, die voornamelijk onder techniek valt. Dit betekent dat dus ondanks er bij DS projecten veel verbeterd kan worden op procesmatig en op rand voorwaardelijk vlak, dat ook de technische KSF nog steeds de nodige aandacht vereisen.

Er zijn ook echt nieuwe bevindingen gedaan op basis van dit onderzoek, zoals de belangrijke samenhang van een aantal KSF, namelijk de relatie tussen data expertise op bestuursniveau, management support, investeren in data kwaliteit en slimme data opslag, het willen experimenteren met DS en een visie voor DS (zie eventueel de lijst van gevonden KSF in hoofdstuk 4.3). Maar bijvoorbeeld ook het voorstel voor een hybride organisatie van DS en te zoeken naar oplossingen voor het sneller in productie brengen van DS oplossingen, bijvoorbeeld door standaardisatie van architectuur, tooling en platform. Tot slot is bijvoorbeeld ook naar voren gekomen dat het opleveren van betrouwbare DS oplossingen heel belangrijk is en dus tevens als promotie dient voor andere toekomstige DS projecten. Het verdient daarom dan ook de aanbeveling om genoeg aandacht te besteden aan het onderhouden van bestaande DS oplossingen en niet slechts te focussen op het ontwikkelen van nieuwe DS oplossingen.

## 5.2. Conclusies

In dit onderzoek is gezocht naar een antwoord op de vraag: ‘Hoe moet een DS project uitgevoerd worden, zodat zij succesvol is?’. Hiervoor zijn 3 deelvragen geformuleerd. De eerste deelvraag luidde als volgt: ‘Wat zijn de KSF van DS projecten?’

Zowel op basis van het literatuuronderzoek naar KSF als op basis van de meervoudige casestudy bij Nederlandse organisaties zijn er KSF gevonden. Aan de éne kant zijn er met behulp van de casestudy KSF gevonden, die niet gevonden zijn bij het literatuuronderzoek en aan de andere kant worden heel veel KSF, die gevonden zijn met behulp van het literatuuronderzoek bevestigd door de casestudy (zie ook tabel 5.1 onderaan deze paragraaf). Hierdoor kan ook de conclusie getrokken worden dat de KSF die vanuit literatuuronderzoek gevonden zijn en van toepassingen waren op niet Nederlandse bedrijven, ook heel vaak van toepassing zijn op Nederlandse bedrijven, vergelijkbaar met die uit deze casestudy.

De tweede deelvraag luidt als volgt: ‘Welke verschillende categorieën succesfactoren kunnen onderscheiden worden bij DS projecten?’

Op basis van het literatuuronderzoek zijn KSF gevonden die gecategoriseerd kunnen worden op basis van verschillende thema’s en op basis van verschillende fases in een DS project. De thema’s betreffen data, processen, technologie, organisatie, mensen en context. De fases betreffen de ‘business’-fase, de data-fase, de analyse-fase, de implementatie-fase, de meet-fase en de leer-fase.

Tot slot luidt de derde deelvraag als volgt: ‘Hoe kan de kennis van KSF ingezet worden zodat DS projecten beter gemanaged kunnen worden?’

Doordat er nu aan de éne kant een uitgebreide lijst van gecategoriseerde KSF (naar thema) bekend is, op basis van zowel literatuuronderzoek als ook bevestigd door deze casestudy en aan de andere kant ook de verschillende fases in een DS project bekend zijn, draagt dit bij aan kennis over waar precies in een DS project uitdagingen kunnen voorkomen. Met deze kennis kunnen de verschillende fases in een DS project geanalyseerd en beoordeeld worden aan de hand van KSF, waarbij extra aandacht gegeven moet worden aan zaken waar organisaties uit deze casestudy de nodige uitdagingen mee hebben:

- Bedrijfscultuur (veel genoemd, bestaat uit veel belangrijke KSF zoals management support, data expertise van bestuur, data governance, data toegankelijkheid)
- Proces management methodologie moet flexibel zijn, houdt te weinig rekening met data scientisten en karakter (verkennend, onzekerheden) van data science
- Implementatie fase (nieuwe technologie, flexibele robuuste architectuur)



Nu de deelvragen beantwoord zijn, wordt de hoofdvraag beantwoord, die als volgt luidt: ‘Hoe moet een DS project uitgevoerd worden, zodat zij succesvol is’. Hoe meer KSF meegenomen kunnen worden in een DS project, des te groter is de kans dat een DS project slaagt. Echter, dan moet men wel kennis hebben van deze KSF. Omdat de KSF niet relevant hoefden te zijn voor Nederlandse organisaties en er misschien hele andere KSF een rol spelen voor Nederlandse organisaties, is er een casestudy gedaan. Met behulp van deze casestudy wordt geconcludeerd, dat ondanks er nog enkele nieuwe KSF zijn gevonden, de meeste KSF van het literatuuronderzoek ook gelden voor Nederlandse organisaties. Dit betekent dat de kennis over de KSF en de kennis over de verschillende fases in een DS project gebruikt kunnen worden om te achterhalen wat er nu goed gaat in een DS project en wat nu verbeterd kan worden, bijvoorbeeld met behulp van een te ontwikkelen audit. Er moet hier in ieder geval extra aandacht gegeven worden aan zaken als de bedrijfscultuur, de gebruikte proces management methodologie en de implementatie fase van een data science project. De conclusie kan getrokken worden dat de kennis over KSF door middel van literatuuronderzoek en casestudy en de kennis van de uitdagingen bij de organisaties in deze casestudy, ingezet kan worden zodat DS projecten aangepast en beter gemanaged kunnen worden en daarmee succesvoller zijn. Hiermee is de hoofdvraag van dit onderzoek beantwoord.

| Toelichting   | Process  | Technology   | People  | Organization  | Data                                    |
|---|--|--|---|---|---|
| Nieuw gevonden KSF  | Opleveren van betrouwbare producten is promotie voor nieuwe data science projecten en draagt dus zo ook bij aan het succes van data science projecten; | •Standaardisatie van architectuur en tooling   | Senioriteit in data scientisten en engineers is belangrijk, voor de beginperiode om een vliegende start te maken;   | Data scientisten in paartjes en in meerdere business teams laten werken, voor kennisborging en kennisontwikkeling;  | data value                              |
| Niet besproken KSF tijdens interviews                           | Genoeg aandacht besteden aan het onderhouden van bestaande data science oplossingen, om data science oplossingen betrouwbaar te houden.                | document collection is a significant problem, only with the right documents provided the needed answers can be found within the data             | Data science / big data expertise op bestuursniveau   | Hybride organisatie van data science: als een organisatie zowel een centrale data science afdeling als vaste data scientisten in business teams heeft, biedt een centrale                   | data variability                        |
| KSF uit de literatuur, ook aan de orde gekomen in de interviews | Open staan voor veranderingen in de gebruikte proces management methodologie en het meenemen van ervaringen van data scientisten hierin;               | performance has to be monitored continuously and the system needs to be tuned on a regular basis to meet operational goals                       | IT staff needs to be aware that creating and maintaining the documentation of the new implemented systems is vital, especially because these are innovative systems | management priority / sponsorship / support   | data visualization                      |
|   | skill gap analysis   | adapt architectural principles   | data analysis and visualisation / interpretation of analytical results skills   | big data strategy alignment (with organization's vision)  | data quality management                 |
|   | the growth of analysis data has to be forecasted to be able to react early when used systems need to be expanded                                       | system security  | mathematical modeling / statistical skills  | well defined organizational structure   | data ownership                          |
|   | specified business case with demonstrable ROI  | technology for data availability   | development of skills / training  | performance management  | data integration                        |
|   | the size and complexity of the project has to be determined carefully  | data architecture  | people skills & ability to self-organize when needed  | data protection and privacy by design   | representativeness of data              |
|   | focus on small projects and known questions  | investment in flexible IT-infrastructure, technology & tools / adequate hardware   | data science technology (for example text analysis methodologies)   | culture of being data-driven / culture of fact based decision making  | document collection / access to sources |
|   | close collaboration between IT and business  | awareness of recent developments in the Big Data analysis area, including evaluation, selection and implementation of new software methodologies | multidisciplinary team (i.e., across different departments)   | information strategy for big data   | high data quality                       |
|   | communication about the data and initiatives   | evaluation of the used hardware to make sure it meets the requirement of the growing data amount also in future                                  | stakeholder coordination / shared understanding   | big data as strategic instrument  | data variety (unstructured/structured)  |
|   | flexibility and agility with freedom for experimentation   | developing, integration and application of logical data models for the new analysis platform   | analytical skillset (including data analysis and visualisation)   | business strategy / vision  | data security                           |
|   | focus on change management   | investment in data sources & data storage  | technical skillset (programming skills / skills for big data projects / technologies)   | management should provide project direction   | data privacy                            |
|   | project difficulty explored and communicated   | reporting- and visualization technology / discovery technology   | business & management skills  | client consultation (involves engaging the internal and external stakeholders to give them the opportunity to air their views, influence the project plans, and know what has been decided) | data volume                             |
|   | clarity of project deliverables (clear or ambiguous) / clear project goal with deadline  | identification and access to needed data sources / combine different data sets   | data scientist in the team  | Creation of a clear vision and implementation strategy for the software and hardware components of the new analysis platform  | data veracity                           |
|   | identifiable business value for the business is delivered from the project, the project must support the business needs                                | cloud-based solutions  | communication and storytelling  |   | data velocity                           |
|   | creation and disseminating of documentation needs to be a fixed part of the project  | integration of new solutions   | domain knowledge  |   |   |
|   | processes need to be established in order to ensure high data quality of the documents used for the analysis   | analytic tools / innovative analysis tools   |   |   |   |
|   | a measurable project outcome has to be defined and in fact measured to evaluate the success of the project   |  |   |   |   |
|   | a clear goal of the project has to be defined, and measurements have to be established, the project needs to have a clear start and end                |  |   |   |   |
|   | feasibility study  |  |   |   |   |
|   | well defined and manageable project scope (understood by team)   |  |   |   |   |
|   | fast delivering of results   |  |   |   |   |
|   | iterative process model  |  |   |   |   |
|   | project management competence  |  |   |   |   |
|   | include business users in the project  |  |   |   |   |
|   | acceptance   |  |   |   |   |
|   | project management process defined   |  |   |   |   |
|   | process management for data science / big data   |  |   |   |   |

### 5.3. Aanbevelingen voor de praktijk

Omdat de conclusie getrokken is dat de meeste van de gevonden KSF op basis van literatuuronderzoek ook voor Nederlandse organisaties geldt, verdient het de aanbeveling voor iedereen die een rol speelt bij het managen van DS projecten, de bestaande literatuur rondom KSF te bestuderen. Deze literatuur is te vinden in de literatuurlijst.

Omdat de informatie rondom de uitdagingen en succesfactoren bij de geanalyseerde organisaties, de kennis van KSF en de kennis van de fases in een DS project de mogelijkheid geeft om DS projecten te beoordelen, verdient het ook de aanbeveling om op basis hiervan zijn of haar DS projecten te analyseren en te beoordelen.

Tot slot is gebleken dat een gebruikte PMM de ruimte heeft om verbeterd te worden. Daarom verdient het ook de aanbeveling om met name de mening van data scientisten mee te nemen rondom een gebruikte PMM. Dit kan wellicht helpen om een PMM te verbeteren of te kiezen voor een andere PMM.

### 5.4. Aanbevelingen voor verder onderzoek

Omdat er interessante opmerkingen door de SD geplaatst zijn met betrekking tot de PMM en er in de praktijk ook steeds meer hybride PMM-en aan het ontstaan zijn (Saltz et al., 2018), verdient het de aanbeveling om vervolgonderzoek te doen, door meer data scientisten te interviewen over zijn of haar ervaringen met de gebruikte PMM door de organisatie. In het bijzonder verdient het de aanbeveling om data scientisten te bevragen bij organisaties met verschillende PMM.

Daarnaast is het opgevallen dat enkele KSF die bekend zijn uit het literatuuronderzoek, niet aan de orde zijn gekomen in de interviews, zoals bijvoorbeeld documentatie rondom de interpretatie van data, systeem performance, systeem veiligheid en de belangrijkheid van documentatie van nieuw geïmplementeerde systemen (voor de volledigheid zie tabel 4.6, hoofdstuk 4.3). Nu zijn er ook geen directe vragen gesteld tijdens de interviews over deze onderwerpen, maar dan nog hadden ze aan de orde kunnen komen aan het einde van het interview, waarbij de geïnterviewde de mogelijkheid had om zaken aan te stippen, die nog niet besproken waren. Om erachter te komen of deze zaken nu wel of geen rol spelen bij Nederlandse organisaties zou in een vervolgonderzoek extra vragen gesteld kunnen worden over deze onderwerpen. Vooral nog lijkt het er op dat deze KSF in ieder geval niet tot de belangrijkste KSF horen bij Nederlandse organisaties met een vergelijkbaar profiel als in deze casestudy.

Tot slot een laatste aanbeveling voor vervolgonderzoek, die eigenlijk uit 2 componenten bestaat, namelijk:

1. In een vervolgonderzoek ook engineers interviewen, die vaak de productie van DS oplossingen voor hun rekening nemen. Dit is interessant, omdat gebleken is uit de resultaten, dat de doorlooptijd van de implementatie fase van DS oplossingen een aandachtsgebied is;
2. Data scientisten en engineers bij organisaties te interviewen die DS oplossingen projecten uitvoeren voor derden, omdat de verwachting is dat bij deze organisaties de doorlooptijd voor de implementatie fase van DS oplossingen korter zal zijn en dat ze wellicht bepaalde methodologieën, die interessant zijn, gebruiken om dit te bereiken.

## Literatuurlijst

- Analytics Vidhya. (2015). 13 Amazing Applications / Uses of DS Today. Geraadpleegd van <https://www.analyticsvidhya.com/blog/2015/09/applications-data-science/>
- Angée, S., Lozano-Argel, S. I., Montoya-Munera, E. N., Ospina-Arango, J.-D., & Tabares-Betancur, M. S. (2018). *Towards an improved ASUM-DM process methodology for cross-disciplinary multi-organization big data & analytics projects*. Paper gepresenteerd op: International Conference on Knowledge Management in Organizations.
- Becker, D. K. (2017). *Predicting outcomes for big data projects: Big Data Project Dynamics (BDPD): Research in progress*. Paper gepresenteerd op: 2017 IEEE International Conference on Big Data (Big Data).
- Bullen, C. V., & Rockart, J. F. (1981). A primer on critical success factors.
- Cervone, H. F. (2011). Understanding agile project management methods using Scrum. OCLC Systems & Services: International digital library perspectives.
- Cios, K. J., & Kurgan, L. A. (2005). Trends in data mining and knowledge discovery. In *Advanced techniques in knowledge discovery and data mining* (pp. 1-26): Springer.
- Das, M., Cui, R., Campbell, D. R., Agrawal, G., & Ramnath, R. (2015). *Towards methods for systematic research on big data*. Paper gepresenteerd op: 2015 IEEE International Conference on Big Data (Big Data).
- Dutta, D., & Bose, I. (2015). Managing a Big Data project: The case of Ramco Cements Limited. *International Journal of Production Economics*, 165, 293-306. doi:10.1016/j.ijpe.2014.12.032
- Emerging Approaches – DS Project Management. (2020, 30 juni). DS Project Management. <http://www.datascience-pm.com/emerging-approaches/>
- Gao, J., Koronios, A., & Selle, S. (2015). Towards a process view on critical success factors in big data analytics projects.
- Gibbs, G. R. (2007). Thematic coding and categorizing. In Gibbs, G. R. *Qualitative Research kit: Analyzing qualitative data* (pp. 38-55). London, England: SAGE Publications, Ltd doi: 10.4135/9781849208574
- Kaisler, S., Armour, F., Espinosa, J. A., & Money, W. (2013). *Big data: Issues and challenges moving forward*. Paper gepresenteerd op: 2013 46th Hawaii International Conference on System Sciences.
- Koronios, A., Gao, J., & Selle, S. (2014). *Big data project success-A meta analysis*. Paper gepresenteerd op: Pacis.
- LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big data, analytics and the path from insights to value. *MIT sloan management review*, 52(2), 21-32.
- Mariscal, G., Marbán, O., & Fernández, C. (2010). A survey of data mining and knowledge discovery process models and methodologies. *Knowledge Eng. Review*, 25, 137-166. doi:10.1017/S0269888910000032
- McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D., & Barton, D. (2012). Big data: the management revolution. *Harvard business review*, 90(10), 60-68.
- Mikalef, P., Boura, M., Lekakos, G., & Krogstie, J. (2019). Big data analytics and firm performance: Findings from a mixed-method approach. *Journal of Business Research*, 98, 261-276. doi:10.1016/j.jbusres.2019.01.044
- Miller, G. J. (2018). Quantitative Comparison of Big Data Analytics and Business Intelligence Project Success Factors. In *Information Technology for Management: Emerging Research and Applications* (pp. 53-72): Springer.
- Morris, T., & Wood, S. (1991). Testing the survey method: continuity and change in British industrial relations. *Work, Employment and Society*, 5(2), 259-282.
- Okoli, C., & Schabram, K. (2010). A guide to conducting a systematic literature review of information systems research.
- Project Management Institute. (2000). *A guide to the project management body of knowledge*

- (*PMBOK guide*). Newtown Square, PA: Project Management Institute
- Priyadarshiny, U. (2019). Top 10 DS Applications. Geraadpleegd van <https://www.edureka.co/blog/data-science-applications/>
- Reinartz, T. (2002). *Stages of the discovery process*. Paper gepresenteerd op: Handbook of data mining and knowledge discovery.
- Rockart, J. F. (1979). Chief executives define their own data needs. *Harvard business review*, 57(2), 81-93.
- Rockart, J. F., & Crescenzi, A. D. (1984). Engaging top management in information planning and development: a case study.
- Saltz, J. S. (2015). *The need for new processes, methodologies and tools to support big data teams and improve big data project effectiveness*. Paper gepresenteerd op: 2015 IEEE International Conference on Big Data (Big Data).
- Saltz, J. S. (2018). *Identifying the Key Drivers for Teams to Use a DS Process Methodology*. Paper gepresenteerd op: ECIS.
- Saltz, J. S., Crowston, K., & Shamshurin, I. (2017). *Comparing DS project management methodologies via a controlled experiment*. Paper gepresenteerd op: Proceedings of the 50th Hawaii International Conference on System Sciences.
- Saltz, J. S., Hotz, N., Wild, D., & Stirling, K. (2018). Exploring Project Management Methodologies Used Within DS Teams.
- Saltz, J. S., & Shamshurin, I. (2016). *Big data team process methodologies: A literature review and the identification of key factors for a project's success*. Paper gepresenteerd op: 2016 IEEE International Conference on Big Data (Big Data).
- Saltz, J. S., Shamshurin, I., & Connors, C. (2017). Predicting DS sociotechnical execution challenges by categorizing DS projects. *Journal of the Association for Information Science and Technology*, 68(12), 2720-2728. doi:10.1002/asi.23873
- Saunders et al. (2016). Research methods for business students: Pearson education.
- Schüritz, R., Brand, E., Satzger, G., & Bischhoffshausen, J. (2017). How to cultivate analytics capabilities within an organization?—design and types of analytics competency centers.
- Scrum vs Agile, wat is het verschil eigenlijk? (2020, 6 maart). Agile Scrum Group. <https://agilescrumgroup.nl/scrum-vs-agile/>
- Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, 70, 263-286. doi:10.1016/j.jbusres.2016.08.001
- Vanauer, M., Böhle, C., & Hellingrath, B. (2015). *Guiding the introduction of big data in organizations: A methodology with business-and data-driven ideation and enterprise architecture management-based implementation*. Paper gepresenteerd op: 2015 48th Hawaii International Conference on System Sciences.
- Walker, J. (2017). Big data strategies disappoint with 85 percent failure rate. *Digital Journal*, 23rd November. Geraadpleegd van: <http://www.digitaljournal.com/tech-and-science/technology/big-datastrategies-disappoint-with-85-percent-failure-rate/article/508325> (bekeken op 28 oktober 2019).
- Watson, H. J. (2014). Tutorial: Big data analytics: Concepts, technologies, and applications. *Communications of the Association for Information Systems*, 34(1), 65.
- Watson, H. J. (2019). Update Tutorial: Big Data Analytics: Concepts, Technology, and Applications. *Communications of the Association for Information Systems*, 44(1), 21.
- Yin., R.K. (2014). Case Study Research: Design and Methods. 5th edition. SAGE Publications.

## Bijlage 1: Artikelen gevonden in het literatuuronderzoek

| Bron                     | Aantal resultaten | Na practical screen | Na quality appraisal | Opmerking   |
|--------------------------|-------------------|---------------------|----------------------|---|
| Google Scholar           | 77                | 13                  | 5                    | Alle 77 artikelen kort langsgelopen en gescreend op relevante inhoud, dit heeft tot 13 bronnen met relevante inhoud geleid, door vervolgens naar de kwaliteit en methodologie te kijken zijn er 5 bronnen overgebleven.   |
| Universiteitsbibliotheek | 107               | 14                  | 3                    | Alle 107 artikelen kort langsgelopen en gescreend op relevante inhoud, dit heeft tot 14 bronnen met relevante inhoud geleid, door vervolgens naar de kwaliteit en methodologie te kijken zijn er 3 bronnen overgebleven.  |
| Google                   | 4.100             | 30                  | 3                    | Ongeveer 4.100 resultaten, maar door alleen de meest relevante resultaten (geen duplicate content) weer te geven, blijven er 30 resultaten over. Er bleven uiteindelijk 5 relevante weblinks over, op basis waarvan er 3 wetenschappelijke artikelen gevonden zijn, die ook van goede kwaliteit zijn. |
| Snowballing              | 5                 | 2                   | 2                    | Door te speuren in de referenties van bovengenoemde artikelen (snowballing) naar andere relevante artikelen. Op basis van de practical screen zijn er 2 artikelen overgebleven.   |

Tabel 1.1: Resultaten literatuuronderzoek

| Literatuur   | Bron                     | Soort bron  | T.b.w. practical screen  | T.b.v. quality appraisal   |
|--|--------------------------|---|--|--|
| Das, M., Cui, R., Campbell, D. R., Agrawal, G., & Ramnath, R. (2015). Towards methods for systematic research on big data. Paper gepresenteerd op: 2015 IEEE International Conference on Big Data (Big Data).  | Google                   | 2015 IEEE International Conference on Big Data (Big Data)           | gebruiken: relevant, gaat over datagedreven planning en uitvoering van de onderzoekslijn   | gebruiken: H-index=16, redelijk goede conferentie  |
| Saltz, J. S. (2015). The need for new processes, methodologies and tools to support big data teams and improve big data project effectiveness. Paper gepresenteerd op: 2015 IEEE International Conference on Big Data (Big Data).  | Google                   | 2015 IEEE International Conference on Big Data (Big Data)           | gebruiken: relevant paper bespreekt de belangrijkste onderzoeksvragen met betrekking tot methodologieën, tools en frameworks om de effectiviteit van big data teams te verbeteren  | gebruiken: Dr. Jeffrey Saltz (B.S. in computer science from Cornell University, an M.B.A. from The Wharton School at the University of Pennsylvania and a Ph.D. in Information Systems from the New Jersey Institute of Technology.) H-index =16   |
| Saltz, J. S., Hotz, N., Wild, D., & Stirling, K. (2018). Exploring Project Management Methodologies Used Within Data Science Teams.  | Google                   | International Conference on Information Systems                     | gebruiken: relevant vanwege de project management methodologieën die besproken worden  | gebruiken: Dr. Jeffrey Saltz (B.S. in computer science from Cornell University, an M.B.A. from The Wharton School at the University of Pennsylvania and a Ph.D. in Information Systems from the New Jersey Institute of Technology.)   |
| Dutta, D., & Bose, I. (2015). Managing a Big Data project: The case of Ramco Cements Limited. International Journal of Production Economics, 165, 293-306. doi:10.1016/j.ijpe.2014.12.032  | Google Scholar           | International Journal of Production Economics                       | gebruiken: relevant, presenteert een framework om big data projecten te implementeren, van business problem tot aan het trainen van personeel. Welke uitdagingen en succesfactoren horen hierbij?  | gebruiken: 152 keer geciteerd door anderen, impact factor = 4,998 Journal Metrics CiteScore: 7,13, Impact Factor: 4,998, 5-Year Impact Factor: 5,631, Source Normalized Impact per Paper (SNIP): 2,486, SCImago Journal Rank (SJR): 2,475  |
| Koronios, A., Gao, J., & Selle, S. (2014). Big data project success-A meta analysis. Paper gepresenteerd op: Pacis.  | Google Scholar           | Pacific Asia Conference on Information Systems (PACIS)              | gebruiken: relevant, noemt een waslijst met succesfactoren. Door succesfactoren af te zetten tegen bepaalde fases in een DS project, zijn er 6 kritieke succesfactoren bepaald.  | gebruiken: 19 keer geciteerd door anderen. H-index = 9   |
| Miller, G. J. (2018). Quantitative Comparison of Big Data Analytics and Business Intelligence Project Success Factors. In Information Technology for Management: Emerging Research and Applications (pp. 53-72): Springer.   | Google Scholar           | Advanced Information Technologies for Management (conference) Paper | gebruiken, relevant hoewel het gaat over zowel BI en BD projecten, kunnen er succesfactoren uitgehaald worden  | gebruiken: is 2 keer geciteerd door anderen. H-index=31  |
| Saltz, J. S. (2018). Identifying the Key Drivers for Teams to Use a Data Science Process Methodology. Paper gepresenteerd op: ECIS.  | Google Scholar           |   | gebruiken: relevant, gaat over de drivers om proces management methodologie te gebruiken   | gebruiken: Dr. Jeffrey Saltz (B.S. in computer science from Cornell University, an M.B.A. from The Wharton School at the University of Pennsylvania and a Ph.D. in Information Systems from the New Jersey Institute of Technology.) Impact Factor: 2,738 H-index =124   |
| Saltz, J. S., & Shamshurin, I. (2016). Big data team process methodologies: A literature review and the identification of key factors for a project's success. Paper gepresenteerd op: 2016 IEEE International Conference on Big Data (Big Data).  | Google Scholar           | 2016 IEEE International Conference on Big Data (Big Data)           | gebruiken: relevant, kunnen succesfactoren uitgehaald worden   | gebruiken: Dr. Jeffrey Saltz (B.S. in computer science from Cornell University, an M.B.A. from The Wharton School at the University of Pennsylvania and a Ph.D. in Information Systems from the New Jersey Institute of Technology.) Impact Factor: 2,738 H-index =124   |
| Kaisler, S., Armour, F., Espinosa, J. A., & Money, W. (2013). Big data: Issues and challenges moving forward. Paper gepresenteerd op: 2013 46th Hawaii International Conference on System Sciences.  | Snowballing              | Hawaii International Conference on System Sciences                  | gebruiken: relevant om big data te beschrijven.  | 796 keer door anderen geciteerd, HICSS ranked second in citation ranking among 18 Information Systems (IS) conferences [1], ranked third in value to the MIS field among 13 Management Information Systems (MIS) conferences [2], and ranked second in conference rating among 11 IS conferences [3] ( <a href="https://aisel.aisnet.org/hicss/">https://aisel.aisnet.org/hicss/</a> ) H-index =72 |
| Vanauer, M., Böhle, C., & Hellingrath, B. (2015). Guiding the introduction of big data in organizations: A methodology with business-and data-driven ideation and enterprise architecture management-based implementation. Paper gepresenteerd op: 2015 48th Hawaii International Conference on System Sciences. | Snowballing              | 2015 48th Hawaii International Conference on System Sciences.       | gebruiken: relevant, bespreekt een methodologie voor big data projecten  | gebruiken: H-index =72   |
| Mikalef, P., Boura, M., Lekakos, G., & Krogstie, J. (2019). Big data analytics and firm performance: Findings from a mixed-method approach. Journal of Business Research, 98, 261-276. doi:10.1016/j.jbusres.2019.01.044   | Universiteitsbibliotheek | Journal of Business Research  | gebruiken: relevant, while the hype around big data analytics is continuously growing, the conditions under which such investments lead to business value remain largely unexplored in empirical research. that what is important is not the technologies surrounding big data analytics, but rather, the organizational diffusion of such technologies towards the generation of insight that can help attain business goals. | gebruiken: 24 keer geciteerd door anderen, impact factor = 4,028 (2018) Journal Metrics, CiteScore: 5,32, Impact Factor: 4,028, 5-Year Impact Factor: 4,747, Source Normalized Impact per Paper (SNIP): 1,920, SCImago Journal Rank (SJR): 1,684   |
| Saltz, J. S., Shamshurin, I., & Connors, C. (2017). Predicting data science sociotechnical execution challenges by categorizing data science projects. Journal of the Association for Information Science and Technology, 68(12), 2720-2728. doi:10.1002/asi.23873   | Universiteitsbibliotheek | Journal of the Association for Information Science and Technology   | gebruiken, relevant vanwege data science aspect  | gebruiken: Dr. Jeffrey Saltz (B.S. in computer science from Cornell University, an M.B.A. from The Wharton School at the University of Pennsylvania and a Ph.D. in Information Systems from the New Jersey Institute of Technology.) Impact Factor: 2,738 H-index =124   |
| Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. Journal of Business Research, 70, 263-286. doi:10.1016/j.jbusres.2016.08.001   | Universiteitsbibliotheek | Journal of Business Research  | gebruiken: relevant vanwege de verschillende uitdagingen bij een compleet project van A to Z.  | gebruiken: 503 keer geciteerd door anderen, impact factor = 4,028 (2018)   |

Tabel 1.2: Toelichting op geselecteerde bronnen n.a.v. practical screen en quality appraisal

| Literatuur  | Bron                     | Soort bron  | T.b.v. practical screen  | T.b.v. quality appraisal  |
|---|--------------------------|---|--|---|
| Ali, I. M., Jusoh, Y. Y., Abdullah, R., & Nor, R. N. H. (2017). A Conceptual Framework For Measuring The Performance of Big Data Analytics Process. <i>Acta Informatica Malaysia (AIM)</i> , 1(2), 13-14.   | Google Scholar           | Acta Informatica Malaysia (AIM)                                     | valt af: performance measurement framework, zoomt meer in op de technische aspecten  | n.v.t.  |
| Ali, I. M., Jusoh, Y. Y., Abdullah, R., Nor, R. N. H., & Affendey, L. S. (2019). Measuring the performance of big data analytics process. <i>Journal of Theoretical and Applied Information Technology</i> , 97(14).  | Google Scholar           | Journal of Theoretical and Applied Information Technology           | valt af: evaluation for BDA process performance, zoomt meer in op de technische aspecten   | n.v.t.  |
| Angé, S., Lozano-Argeles, S. I., Montoya-Munera, E. N., Ospina-Arango, J.-D., & Tabares-Betancur, M. S. (2018). Towards an improved ASUM-DM process methodology for cross-disciplinary multi-organization big data & analytics projects. Paper gepresenteerd op: International Conference on Knowledge Management in Organizations. | Google Scholar           | International Conference on Knowledge Management in Organizations   | gebruiken: interessant vanwege relevantie proces methodologie  | Onder andere gefinancierd door Colombiaanse overheid en instituten, geen bedrijven. Geciteerd door 8 anderen. |
| Miller, G. J. (2019). The influence of big data competencies, team structures, and data scientists on project success. Paper gepresenteerd op: 2019 IEEE Technology & Engineering Management Conference (TEMSCON).  | Google Scholar           | 2019 IEEE Technology & Engineering Management Conference (TEMSCON)  | valt af: lijkt niet extra veel toe te voegen op artikel van Miller (2018)  | n.v.t.  |
| Papadaki, D., Bakas, D. N., Karamitsos, D., & Kirkham, D. (2019). Big data from social media and scientific literature databases reveals relationships among risk management, project management and project success. <i>Project Management and Project Success</i> (September 26, 2019).   | Google Scholar           | PM World Journal  | valt af: niet relevant genoeg, gaat meer over de relaties tussen risico management, project management en project successen.               | n.v.t.  |
| Sachdeva, N., Singh, O., & Kapur, P. (2015). Modeling critical success factors for adoption of big data analytics project: an ISM-MICMAC based analysis. <i>Communications in Dependability and Quality Management</i> , 18(4), 93-110.   | Google Scholar           | Communications in Dependability and Quality Management              | valt af: gaat alleen over de adoptie van big data projecten  | n.v.t.  |
| Santos, E. F. d. (2017). Critical success factors in the implementation of a Big Data project: a case study. <i>Instituto Superior de Economia e Gestão</i> .   | Google Scholar           | Instituto Superior de Economia e Gestão                             | valt af: Portugees en geen full text   | n.v.t.  |
| Shaheen, S. (2018). Moderating role of decision making on relationship Between big data analytics and project success. <i>Bahria University Islamabad Campus</i> .  | Google Scholar           | Bahria University Islamabad Campus                                  | valt af: de full tekst van de studie niet beschikbaar is via de universiteitsbibliotheek, tevens is het slechts een gedeeltelijke scriptie | n.v.t.  |
| Chang, W., et al. Big Data: Challenges, practices and technologies: NIST Big Data Public Working Group workshop at IEEE Big Data. Big Data (Big Data), IEEE International Conference on.  | Snowballing              | 2014 IEEE International Conference on Big Data (Big Data)           | valt af: zoomt meer in op de technische aspecten.  | n.v.t.  |
| Jagadish, H. V., et al. (2014). "Big data and its technical challenges." <i>Communications of the ACM</i> 57(7): 86-94.   | Snowballing              | Journal of the Association for Computing Machinery (ACM).           | valt af: zoomt meer in op de technische aspecten.  | n.v.t.  |
| Katal, A., et al. (2013). Big data: issues, challenges, tools and good practices. 2013 Sixth international conference on contemporary computing (IC3), IEEE.  | Snowballing              | 2013 Sixth International Conference on Contemporary Computing (IC3) | valt af: zoomt meer in op de technische aspecten.  | n.v.t.  |
| Almeida, F. L. (2017). Benefits, challenges and tools of big data management. <i>Journal of Systems Integration</i> , 8(4), 12-20.  | Universiteitsbibliotheek | Journal of Systems Integration                                      | valt af: zoomt meer in op de technische aspecten.  | n.v.t.  |
| Al-Taie, M. Z., Salim, N., & Obasa, A. I. (2017). Successful Data Science Projects: Lessons Learned from Kaggle Competition. <i>Kurdistan Journal of Applied Research</i> , 2(3), 40-49. doi:10.24017/science.2017.3.18   | Universiteitsbibliotheek | Kurdistan Journal of Applied Research                               | relevant: succesfactoren obv Kaggle  | lage kwaliteit journal -> h5-index:4; h5-mediaan:5  |
| Amankwah-Amoah, J., & Adomako, S. (2019). Big data analytics and business failures in data-Rich environments: An organizing framework. <i>Computers in Industry</i> , 105, 204-212. doi:10.1016/j.comind.2018.12.015  | Universiteitsbibliotheek | Computers in Industry   | valt af: zoomt meer in op de technische aspecten.  | n.v.t.  |
| Birney, E. (2012). Lessons for big-data projects. <i>Nature</i> , 489(7414), 49-51. doi:10.1038/489049a   | Universiteitsbibliotheek | Nature  | valt af: niet de meeste relevante journal  | n.v.t.  |
| Ferraris, A., Mazzoleni, A., Devalle, A., & Couturier, J. (2019). Big data analytics capabilities and knowledge management: impact on firm performance. <i>Management Decision</i> , 57(8), 1923-1936. doi:10.1108/MD-07-2018-0825  | Universiteitsbibliotheek | Management Decision   | valt af: zoomt meer in op de technische aspecten.  | n.v.t.  |
| Gil, D., & Song, I.-Y. (2016). Modeling and Management of Big Data: Challenges and opportunities. <i>Future Generation Computer Systems</i> , 63, 96-99. doi:10.1016/j.future.2015.07.019   | Universiteitsbibliotheek | Future Generation Computer Systems                                  | valt af: zoomt meer in op de technische aspecten.  | n.v.t.  |
| Pawar, A. M. (2016). Big Data mining: challenges, technologies, tools and applications. <i>Database Systems Journal</i> , 7(2), 28-33.  | Universiteitsbibliotheek | Database Systems Journal  | valt af: zoomt meer in op de technische aspecten.  | n.v.t.  |
| Pugna, I. B., Dutescu, A., & Stanila, O. G. (2019). Corporate Attitudes towards Big Data and Its Impact on Performance Management: A Qualitative Study. <i>SUSTAINABILITY</i> , 11(3), 684. doi:10.3390/su11030684  | Universiteitsbibliotheek | SUSTAINABILITY  | valt af: performance measurement framework, zoomt meer in op de technische aspecten  | n.v.t.  |
| Rodríguez-Mazahua, L., Rodríguez-Enriquez, C.-A., Sánchez-Cervantes, J. L., Cervantes, J., García-Alcaraz, J. L., & Alor-Hernández, G. (2016). A general perspective of Big Data: applications, tools, challenges and trends. <i>The Journal of Supercomputing</i> , 72(8), 3073-3113. doi:10.1007/s11227-015-1501-1                | Universiteitsbibliotheek | The Journal of Supercomputing                                       | valt af: zoomt meer in op de technische aspecten.  | n.v.t.  |
| Sena, V., Bhaumik, S., Sengupta, A., & Demirbag, M. (2019). Big Data and Performance: What Can Management Research Tell us? <i>British Journal of Management</i> , 30(2), 219-228. doi:10.1111/1467-8551.12362  | Universiteitsbibliotheek | British Journal of Management                                       | valt af: slaat meer op performance management  | n.v.t.  |
| Vargas-Solar, G., Zechinelli-Martini, J. L., & Espinosa-Oviedo, J. A. (2017). Big Data Management: What to Keep   | Universiteitsbibliotheek | Data Science and Engineering  | valt af: zoomt meer in op de technische aspecten.  | n.v.t.  |

Tabel 1.3: Toelichting op niet geselecteerde bronnen n.a.v. practical screen en quality appraisal

| Artikel  | Bron                     |
|--|--------------------------|
| Das, M., Cui, R., Campbell, D. R., Agrawal, G., & Ramnath, R. (2015). Towards methods for systematic research on big data. Paper gepresenteerd op: 2015 IEEE International Conference on Big Data (Big Data).  | Google                   |
| Saltz, J. S. (2015). The need for new processes, methodologies and tools to support big data teams and improve big data project effectiveness. Paper gepresenteerd op: 2015 IEEE International Conference on Big Data (Big Data).  | Google                   |
| Saltz, J. S., Hotz, N., Wild, D., & Stirling, K. (2018). Exploring Project Management Methodologies Used Within DS Teams.  | Google                   |
| Dutta, D., & Bose, I. (2015). Managing a Big Data project: The case of Ramco Cements Limited. International Journal of Production Economics, 165, 293-306. doi:10.1016/j.ijpe.2014.12.032  | Google Scholar           |
| Koronios, A., Gao, J., & Selle, S. (2014). Big data project success-A meta analysis. Paper gepresenteerd op: Pacis.  | Google Scholar           |
| Miller, G. J. (2018). Quantitative Comparison of Big Data Analytics and Business Intelligence Project Success Factors. In Information Technology for Management: Emerging Research and Applications (pp. 53-72): Springer.   | Google Scholar           |
| Saltz, J. S. (2018). Identifying the Key Drivers for Teams to Use a DS Process Methodology. Paper gepresenteerd op: ECIS.  | Google Scholar           |
| Saltz, J. S., & Shamshurin, I. (2016). Big data team process methodologies: A literature review and the identification of key factors for a project's success. Paper gepresenteerd op: 2016 IEEE International Conference on Big Data (Big Data).  | Google Scholar           |
| Kaisler, S., Armour, F., Espinosa, J. A., & Money, W. (2013). Big data: Issues and challenges moving forward. Paper gepresenteerd op: 2013 46th Hawaii International Conference on System Sciences. International Conference on, 2013, pp. 995-1004: IEEE.   | Snowballing              |
| Vanauer, M., Böhle, C., & Hellingrath, B. (2015). Guiding the introduction of big data in organizations: A methodology with business-and data-driven ideation and enterprise architecture management-based implementation. Paper gepresenteerd op: 2015 48th Hawaii International Conference on System Sciences. | Snowballing              |
| Mikalef, P., Boura, M., Lekakos, G., & Krogstie, J. (2019). Big data analytics and firm performance: Findings from a mixed-method approach. Journal of Business Research, 98, 261-276. doi:10.1016/j.jbusres.2019.01.044   | Universiteitsbibliotheek |
| Saltz, J. S., Shamshurin, I., & Connors, C. (2017). Predicting DS sociotechnical execution challenges by categorizing DS projects. Journal of the Association for Information Science and Technology, 68(12), 2720-2728. doi:10.1002/asi.23873   | Universiteitsbibliotheek |
| Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. Journal of Business Research, 70, 263-286. doi:10.1016/j.jbusres.2016.08.001   | Universiteitsbibliotheek |

Tabel 1.4: Resultaten literatuuronderzoek



## Bijlage 2: Synthese van kritieke succesfactoren

| Data                                    | Organization   | Process  | People  | Technology   |
|---|--|--|---|--|
| data quality management                 | management priority / sponsorship / support              | close collaboration between IT and business              | development of skills / training                            | awareness of recent developments in the Big Data analysis area, including evaluation, selection and implementation of new software methodologies |
| data ownership                          | big data strategy alignment (with organization's vision) | communication about the data and initiatives             | people skills & ability to self-organize when needed        | document collection is a significant problem, only with the right documents provided the needed answers can be found within the data             |
| data integration                        | well defined organizational structure                    | flexibility and agility with freedom for experimentation | DS technology (for example text analysis methodologies)     | evaluation of the used hardware to make sure it meets the requirement of the growing data amount also in future                                  |
| representativeness of data              | performance management                                   | focus on change management                               | multidisciplinary team (i.e., across different departments) | developing, integration and application of logical data models for the new analysis platform   |
| document collection / access to sources | data protection and privacy by design                    | project difficulty explored and communicated             | stakeholder coordination / shared understanding             | performance has to be monitored continuously and the system needs to be tuned on a regular basis to meet operational goals                       |

|  |  |   |   |  |
|--|--|---|---|--|
| high data quality                      | culture of being data-driven / culture of fact based decision making                           | clarity of project deliverables (clear or ambiguous) / clear project goal with deadline                                 | IT staff needs to be aware that creating and maintaining the documentation of the new implemented systems is vital, especially because these are innovative systems | investment in data sources & data storage                                      |
| data variety (unstructured/structured) | information strategy for big data  | identifiable business value for the business is delivered from the project, the project must support the business needs | analytical skillset (including data analysis and visualisation)   | reporting- and visualization technology / discovery technology                 |
| data security                          | big data as strategic instrument   | creation and disseminating of documentation needs to be a fixed part of the project                                     | technical skillset (programming skills / skills for big data projects / technologies)   | identification and access to needed data sources / combine different data sets |
| data privacy                           | business strategy / vision   | the growth of analysis data has to be forecasted to be able to react early when used systems need to be expanded        | business & management skills  | cloud-based solutions  |
| data volume                            | management should provide project direction  | processes need to be established in order to ensure high data quality of the documents used for the analysis            | data scientist in the team  | adapt architectural principles   |
| data veracity                          | client consultation (involves engaging the internal and external stakeholders to give them the | a measureable project outcome has to be defined and in fact measured to evaluate the success of the project             | communication and storytelling  | integration of new solutions   |

|                    |  |   |   |  |
|--------------------|--|---|---|--|
|                    | opportunity to air their views, influence the project plans, and know what has been decided)                                 |   |   |  |
| data velocity      | Creation of a clear vision and implementation strategy for the software and hardware components of the new analysis platform | a clear goal of the project has to be defined, and measurements have to be established, the project needs to have a clear start and end | domain knowledge  | analytic tools / innovative analysis tools                                       |
| data variability   |  | the size and complexity of the project has to be determined carefully   | data analysis and visualisation / interpretation of analytical results skills | data architecture  |
| data visualization |  | focus on small projects and known questions   | mathematical modelling / statistical skills                                   | investment in flexible IT-infrastructure, technology & tools / adequate hardware |
| data value         |  | specified business case with demonstrable ROI   |   | technology for data availability   |
|                    |  | feasibility study   |   | system security  |
|                    |  | skill gap analysis  |   |  |
|                    |  | well defined and manageable project scope (understood by team)  |   |  |
|                    |  | fast delivering of results  |   |  |
|                    |  | iterative process model   |   |  |
|                    |  | project management competence   |   |  |

|  |  |  |  |  |
|--|--|--|--|--|
|  |  | Include business users in the project acceptance |  |  |
|  |  | project management process defined               |  |  |
|  |  | process management for DS / big data             |  |  |

## Bijlage 3: Interview protocol

### Script voorafgaand aan interview:

#### Introductie

Voorstelling van mezelf/mijn organisatie: refereren naar telefoongesprek dat ook al gevoerd is.

#### Dankwoord

Allereerst dank voor je bereidheid om deel te nemen aan het interviewaspect van mijn studie.

#### Doel van het interview

Zoals ik al eerder heb aangegeven, is het doel van mijn onderzoek te begrijpen hoe DS projecten het beste uitgevoerd kunnen worden, zodat zij succesvol zijn. Ik doe dit aan de hand van interviews en heb hiervoor mensen geselecteerd die mij wat kunnen vertellen over de cultuur in de organisatie en over de organisatorische, projectmatige en technische aspecten van DS projecten.

Met behulp van onder andere ons interview probeer ik vervolgens te begrijpen wat nu de uitdagingen en KSF zijn bij DS projecten. Ons interview van vandaag zal zoals is afgesproken maximaal een uur duren. De verwachting is dat eind juli 2020 mijn onderzoek afgerond zal zijn, waar na u ook de uitkomsten van mijn onderzoek via email zult ontvangen.

Via de email heeft u reeds een toestemmingsformulier ingevuld dat aangeeft dat ik (al dan niet) uw toestemming heb om ons gesprek op te nemen. Vindt u het nog steeds goed dat ik ons gesprek vandaag opneem? \_\_\_Ja \_\_\_Nee

Zo ja: bedankt! Laat het me weten als u op enig moment wilt dat ik de recorder uitschakel of iets dat u gezegd heeft buiten de opname houdt. Zo nee: bedankt voor het laten weten. Ik zal alleen aantekeningen maken van ons gesprek.

Verder is uw deelname altijd in te trekken en kunnen gegeven antwoorden worden uitgesloten voor verdere analyse. Ons interview zal geanonimiseerd worden, uw organisatie zal dus niet bij naam genoemd worden in de uitwerking van de interviews en in het verslag zelf.

Om u een gevoel voor de opbouw van het interview te geven. De volgende thema's zullen de revue passeren:

- algemene vragen
- vragen over hoe DS projecten verlopen zijn bij jullie
- vragen over organisatie en cultuur
- procesmatige vragen
- tot slot hebben we tijd om in te gaan op eventuele zaken die niet besproken zijn

Geeft dat een beetje een idee? Heeft u nog vragen voordat we met het interview beginnen? [Vragen bespreken] Prima, als er op enig moment in dit onderzoek vragen (of andere vragen) rijzen, kunt u deze op elk gewenst moment stellen. Ik beantwoord uw vragen graag. Dan kunnen we nu beginnen.

## **Interviewvragen**

### **Algemeen:**

1. Zou u kunnen aangeven welke functie u vervult binnen uw organisatie?
2. Wat is precies uw rol bij DS (projecten) binnen jullie organisatie?
3. Wat verstaan jullie eigenlijk onder een DS project?
4. Hoe belangrijk is data voor jullie organisatie? Hebben jullie een data strategie?
5. Hoe is DS bij jullie georganiseerd?
6. Hoe belangrijk zijn DS projecten voor jullie?
7. Hoe lang zijn jullie al bezig met DS projecten
8. Hoeveel projecten hebben jullie inmiddels uitgevoerd?
9. Kun je iets vertellen over hoe lang gemiddeld DS projecten hebben gelopen?  
Mocht het zo zijn dat er veel verschil is in tijd:  
Kunt u aangeven wat nu typisch een kort project is geweest bij jullie?  
En wat is typisch een lang project geweest?
10. Kun je beknopt aangeven wat voor projecten jullie zoal gedaan hebben, zodat ik inhoudelijk een gevoel heb voor de projecten?
11. Wat is de trigger geweest om met DS projecten te starten?

### **Kernvragen (vragen over hoe DS projecten verlopen zijn en waarom)**

12. Hebben jullie te maken met uitdagingen bij ds projecten, kun je daar iets over vertellen?
13. Zijn jullie tevreden over hoe DS projecten worden uitgevoerd? Zo ja:  
Welke zaken vindt u goed gaan bij jullie DS projecten?
14. Heeft u ideeën over hoe uw DS projecten beter uitgevoerd kunnen worden?
15. Zijn er aspecten in een DS project die zich op basis van uw ervaring moeilijker laten plannen of managen?
16. Wat zijn in uw ogen de belangrijkste aandachtspunten bij de uitvoering van een DS project?  
Doorvragen:  
Is dat ook een situatie waarin jullie je bevonden hebben / aandachtspunt die jullie ondervonden hebben? Zo ja:  
Wat hebben jullie toen gedaan, toen jullie je in die situatie bevonden?  
Eventueel herformuleren:  
Zijn er andere uitdagingen die jullie gehad hebben bij DS projecten, waarover u kunt vertellen?
17. Zijn er zaken die jullie momenteel belemmeren (of belemmerd hebben) om te profiteren van DS?  
Zo ja: Zijn deze zaken onder de aandacht gebracht?  
Hebben jullie hier al oplossingen voor in beeld?  
Waar heb jij voornamelijk de behoefte aan bij DS projecten, dat nu wellicht onvoldoende aanwezig is, ontbreekt ofzo?
18. Is er een proces ingericht om te leren van zaken die goed of minder goed gingen bij reeds uitgevoerde projecten? Evaluatieproces
19. Hebben jullie al succesvolle ds projecten gehad? Wat heeft bijgedragen bij succes?
20. Hoe bepalen jullie of een DS project succesvol is geweest? Wanneer is het geslaagd?

Dank u wel voor de antwoorden zo ver, ik zou nu graag wat vragen willen stellen, over hoe DS projecten bij jullie georganiseerd zijn.

### **Organisatie en cultuur:**

21. Hebben jullie een ds team? Waaruit bestaat dat?  
Wie initieert de projecten?
22. Aan welke voorwaarden moeten er voldaan zijn, voordat gestart wordt?
23. Wie voeren de projecten uit?  
Welke afdelingen zijn bijvoorbeeld bij de projecten betrokken?
24. Welke rol speelt de directie bij DS projecten, wordt DS gestimuleerd door hun?
25. Kun je wat vertellen over de dataexpertise op bestuursniveau? Hebben jullie een CDO?  
Wie gaan over het inhoudelijke gebruik van data en het beheersen van risico's daaromtrent?  
Hebben jullie een privacy officer?
26. Heeft het personeel dat niet direct een functie heeft rondom het werken met data, ook toegang tot allerlei data? Zo ja: Kun je daar een voorbeeld van geven, makkelijk, dashboards?
27. Hoe worden beslissingen genomen in jullie organisatie? Eventueel toelichten: Is dit op basis van feiten of op basis van onderbuikgevoel, kun je daar een voorbeeld van geven?
28. Welke middelen hebben jullie (eventueel voorbeelden geven zoals gecentraliseerde data, tooling (ook open source), schaalbaarheid, cloud technologie, data-lake, data analytics platform)?

### **Procesmatig:**

Bedankt voor jouw heldere antwoorden, ik zou nu graag wat vragen willen stellen die meer inzoomen op de uitvoering van een project.

29. Maken jullie binnen de organisatie gebruik van procesmethodologieën zoals Agile, Scrum, Kanban, projectmatig werken of iets dergelijks? Zo ja, welke? Zo nee: Is dit ad-hoc en afhankelijk van het project en de mensen zelf of weer iets anders?
30. Wat zijn de voordelen en nadelen van de gebruikte methode?
31. Hoe lang mag een DS project duren?
32. Hoe plannen jullie een DS project, welke stappen hebben jullie daarbij?
33. Hoe wordt een DS project bij jullie gepland?  
Eventueel voorbeelden geven zoals de business fase, data fase.  
Doorvragen over de fases: Zou je daar wat meer over kunnen vertellen?  
Eventueel extra vragen stellen over:  
Geformuleerde doelen helder?  
Scope project helder?  
Voorwaarden project helder?  
Worden de eventuele risico's en verantwoordelijkheden in kaart gebracht?
34. Wie is degene die die communicatie verzorgt van de DS uitkomsten van een project?  
Hoe communiceren jullie de uitkomsten van DS projecten?

### **Tot slot:**

Tot slot, voordat we dit interview afronden, zijn er bepaalde zaken die niet aan bod zijn gekomen, maar waarvan u vindt dat ze wel belangrijk zijn? Zo ja:

35. Kunt u dat eens verder toelichten? Zo nee:
36. Heb je verder nog advies voor mij, hetgeen ik mee kan nemen in andere interviews?

### **Afsluiting**

Dan zijn we hierbij aangekomen bij het einde van het interview. Ik wil u danken voor uw tijd en moeite. Ik vond het reuze interessant om te horen hoe jullie bezig zijn met DS. Uiterlijk 3 weken na dit interview zal ik u de uitwerking van dit interview ter goedkeuring aanbieden. Mijn indicatie is dat

ik eind juli gereed zal zijn met mijn onderzoek. Zodra dit het geval is, zal ik per email de uitkomsten van mijn onderzoek opsturen. Nogmaals bedankt en ik zal nu de opname stopzetten.

## Bijlage 4: Lijst codes

| Afkorting | Organisatie   | Functie                                     |
|-----------|---------------|---|
| GMO       | Organisatie 1 | Groupmanager Operations Research & Big Data |
| PO        | Organisatie 1 | Product Owner                               |
| MDS       | Organisatie 2 | Manager DS                                  |
| SDS       | Organisatie 2 | Senior Data Scientist                       |
| DI        | Organisatie 3 | Director Insights                           |
| MADS      | Organisatie 3 | Manager Analytics & Decision Support        |

| Code                                 | Grounded | Code Groups   |
|--------------------------------------|----------|---------------|
| Organisatie 1                        | 76       | Organisatie 1 |
| Organisatie 2                        | 54       | Organisatie 2 |
| Organisatie 3                        | 49       | Organisatie 3 |
| Groupmanager Operations & Big Data   | 39       |               |
| Product Owner                        | 37       |               |
| Director Insights                    | 25       |               |
| Manager DS                           | 22       |               |
| Senior Data Scientist                | 32       |               |
| Manager Analytics & Decision Support | 23       |               |
| Algemeen                             | 71       |               |
| Presentatie                          | 13       | Algemeen      |
| Data toegankelijkheid                | 11       | Algemeen      |
| Aanleiding voor DS projecten         | 11       | Algemeen      |
| Middelen voor DS projecten           | 9        | Algemeen      |
| Data governance                      | 9        | Algemeen      |
| Bezig met DS projecten               | 9        | Algemeen      |
| Type DS project                      | 9        | Algemeen      |
| Urgentie data kwaliteit              | 9        | Algemeen      |
| Volwassenheid (optiek organisatie)   | 8        | Algemeen      |
| Belangrijkheid van DS projecten      | 8        | Algemeen      |
| Belangrijkheid data                  | 8        | Algemeen      |

|   |    |                       |
|---|----|-----------------------|
| Duur van DS projecten                           | 6  | Algemeen              |
| Cultuur & organisatie                           | 69 |                       |
| Bedrijfscultuur                                 | 25 | Cultuur & Organisatie |
| Organisatie DS                                  | 14 | Cultuur & Organisatie |
| Management support                              | 12 | Cultuur & Organisatie |
| DS expertise op bestuursniveau                  | 11 | Cultuur & Organisatie |
| Samenstelling team bij DS projecten             | 8  | Cultuur & Organisatie |
| Uitvoerders DS projecten                        | 8  | Cultuur & Organisatie |
| Initiatoren DS projecten                        | 7  | Cultuur & Organisatie |
| DS community                                    | 2  | Cultuur & Organisatie |
| Ervaringen met DS projecten                     | 45 |                       |
| Relatie met de business                         | 7  | Ervaringen            |
| Senioriteit                                     | 7  | Ervaringen            |
| Participatie van de business                    | 6  | Ervaringen            |
| Resultaten DS project                           | 4  | Ervaringen            |
| Zelfstandig DS oplossingen in productie brengen | 3  | Ervaringen            |
| IT support                                      | 2  | Ervaringen            |
| DS oplossing in productie brengen               | 2  | Ervaringen            |
| Data scientists                                 | 2  | Ervaringen            |
| Standaardisatie in technologie                  | 1  | Ervaringen            |
| Nieuwe IT technologie                           | 1  | Ervaringen            |
| Uniformiteit in manier van werken               | 1  | Ervaringen            |
| Vrijheid in tooling                             | 1  | Ervaringen            |
| Leidend worden                                  | 1  | Ervaringen            |
| Betrouwbaarheid DS oplossing                    | 1  | Ervaringen            |
| Het DS proces                                   | 38 |                       |
| Start DS project                                | 10 | Het DS proces         |
| Communicatie uitkomsten DS projecten            | 8  | Het DS proces         |
| Evaluatie DS projecten                          | 6  | Het DS proces         |
| Bepalen van succes                              | 4  | Het DS proces         |
| Niet besproken                                  | 5  |                       |
| Hybride DS organisatie                          | 1  | Niet besproken        |
| Best practice process for DS                    | 1  | Niet besproken        |
| Softe of randvoorwaardelijke elementen          | 1  | Niet besproken        |



## Bijlage 5: Groepering van de codes per organisatie

| Codes   | • Organisatie 1Gr=76 | • Organisatie 2Gr=54 | • Organisatie 3Gr=49 |
|---|----------------------|----------------------|----------------------|
| • AlgemeenGr=71   | 30                   | 21                   | 20                   |
| • Cultuur & organisatieGr=69                                    | 28                   | 23                   | 18                   |
| • Het data science procesGr=38                                  | 13                   | 13                   | 12                   |
| • Ervaringen met data science projectenGr=45                    | 20                   | 16                   | 9                    |
| • BedrijfscultuurGr=25  | 12                   | 7                    | 6                    |
| • Organisatie data scienceGr=14                                 | 6                    | 6                    | 2                    |
| • Proces management methodologieGr=14                           | 5                    | 5                    | 4                    |
| • PresentatieGr=13  | 5                    | 4                    | 4                    |
| • Management supportGr=12                                       | 2                    | 6                    | 4                    |
| • Data toegankelijkheidGr=11                                    | 8                    | 1                    | 2                    |
| • Aanleiding voor data science projectenGr=11                   | 6                    | 3                    | 2                    |
| • Data science expertise op bestuursniveauGr=11                 | 4                    | 5                    | 2                    |
| • Start data science projectGr=10                               | 3                    | 5                    | 2                    |
| • Urgentie data kwaliteitGr=9                                   | 6                    | 1                    | 2                    |
| • bezig met DS projectenGr=9                                    | 5                    | 2                    | 2                    |
| • Type data science projectGr=9                                 | 4                    | 3                    | 2                    |
| • Data governanceGr=9   | 3                    | 4                    | 2                    |
| • Middelen voor data science projectenGr=9                      | 3                    | 3                    | 3                    |
| • Volwassenheid (optiek organisatie)Gr=8                        | 5                    | 3                    | 0                    |
| • Belangrijkheid dataGr=8                                       | 4                    | 3                    | 1                    |
| • Uitvoerders data science projectenGr=8                        | 4                    | 2                    | 2                    |
| • Belangrijkheid van data science projectenGr=8                 | 3                    | 2                    | 3                    |
| • Samenstelling team bij data science projectenGr=8             | 3                    | 2                    | 3                    |
| • Communicatie uitkomsten data science projectenGr=8            | 2                    | 3                    | 3                    |
| • SenioriteitGr=7   | 5                    | 2                    | 0                    |
| • Initiatoren data science projectenGr=7                        | 3                    | 2                    | 2                    |
| • Relatie met de businessGr=7                                   | 2                    | 3                    | 2                    |
| • Duur van data science projectenGr=6                           | 2                    | 2                    | 2                    |
| • Evaluatie data science projectenGr=6                          | 2                    | 2                    | 2                    |
| • Participatie van de businessGr=6                              | 1                    | 1                    | 4                    |
| • Niet besprokenGr=5  | 3                    | 2                    | 0                    |
| • Resultaten data science projectGr=4                           | 3                    | 0                    | 1                    |
| • Bepalen van succesGr=4  | 1                    | 1                    | 2                    |
| • Zelfstandig data science oplossingen in productie brengenGr=4 | 2                    | 1                    | 0                    |
| • Data science oplossing in productie brengenGr=2               | 2                    | 0                    | 0                    |
| • Data science communityGr=2                                    | 1                    | 1                    | 0                    |
| • Data scientistenGr=2  | 2                    | 0                    | 0                    |
| • IT supportGr=2  | 1                    | 0                    | 1                    |
| • Leidend wordenGr=1  | 0                    | 1                    | 0                    |
| • Standaardisatie in technologieGr=1                            | 0                    | 1                    | 0                    |
| • Uniformiteit in manier van werkenGr=1                         | 0                    | 1                    | 0                    |
| • Vrijheid in toolingGr=1                                       | 0                    | 1                    | 0                    |
| • Hybride data science organisatieGr=1                          | 0                    | 1                    | 0                    |
| • Betrouwbaarheid data science oplossingGr=1                    | 1                    | 0                    | 0                    |
| • Nieuwe IT technologieGr=1                                     | 1                    | 0                    | 0                    |
| • Best practice process for data scienceGr=1                    | 1                    | 0                    | 0                    |
| • Softe of randvoorwaardelijke elementenGr=1                    | 1                    | 0                    | 0                    |

## Bijlage 6: Groepering van de codes per functie

| Codes  | • Director Insights<br>Gr=25 | • Groupmanager Operations & Big Data<br>Gr=39 | • Manager Data Science<br>Gr=22 | • Product Owner<br>Gr=37 | • Manager Analytics & Decision Support<br>Gr=23 | • Senior Data Scientist<br>Gr=32 |
|--|------------------------------|---|---------------------------------|--------------------------|---|----------------------------------|
| • BedrijfscultuurGr=25                                       | 3                            | 3   | 2                               | 9                        | 3   | 5                                |
| • Management supportGr=12                                    | 2                            | 1   | 3                               | 1                        | 2   | 3                                |
| • PresentatieGr=13   | 2                            | 3   | 2                               | 2                        | 2   | 2                                |
| • Belangrijkheid van data science projectenGr=8              | 2                            | 1   | 2                               | 2                        | 1   | 0                                |
| • Communicatie uitkomsten data science projectenGr=8         | 2                            | 1   | 1                               | 1                        | 1   | 2                                |
| • Participatie van de businessGr=6                           | 2                            | 1   | 0                               | 0                        | 1   | 1                                |
| • Proces management methodologieGr=14                        | 1                            | 2   | 1                               | 3                        | 3   | 4                                |
| • Samenstelling team bij data science projectenGr=8          | 1                            | 2   | 0                               | 1                        | 2   | 2                                |
| • Data science expertise op bestuursniveauGr=11              | 1                            | 1   | 3                               | 3                        | 1   | 2                                |
| • Data governanceGr=9  | 1                            | 2   | 2                               | 1                        | 1   | 2                                |
| • Middelen voor data science projectenGr=9                   | 1                            | 1   | 2                               | 2                        | 1   | 1                                |
| • Start data science projectGr=10                            | 1                            | 1   | 1                               | 2                        | 1   | 4                                |
| • Aanleiding voor data science projectenGr=11                | 1                            | 1   | 1                               | 5                        | 1   | 2                                |
| • Type data science projectGr=9                              | 1                            | 1   | 1                               | 3                        | 1   | 2                                |
| • Bezig met DS projectenGr=9                                 | 1                            | 2   | 1                               | 3                        | 1   | 1                                |
| • Uitvoerders data science projectenGr=8                     | 1                            | 1   | 1                               | 3                        | 1   | 1                                |
| • Initiatoren data science projectenGr=7                     | 1                            | 1   | 1                               | 2                        | 1   | 1                                |
| • Duur van data science projectenGr=6                        | 1                            | 1   | 1                               | 1                        | 1   | 1                                |
| • Bepalen van succesGr=4                                     | 1                            | 1   | 1                               | 0                        | 1   | 0                                |
| • Evaluatie data science projectenGr=6                       | 1                            | 1   | 0                               | 1                        | 1   | 2                                |
| • Data toegankelijkheidGr=11                                 | 1                            | 4   | 0                               | 4                        | 1   | 1                                |
| • Belangrijkheid dataGr=8                                    | 1                            | 1   | 2                               | 3                        | 0   | 1                                |
| • IT supportGr=2   | 1                            | 0   | 0                               | 1                        | 0   | 0                                |
| • Organisatie data scienceGr=14                              | 0                            | 3   | 3                               | 3                        | 2   | 3                                |
| • Urgentie data kwaliteitGr=9                                | 0                            | 2   | 1                               | 4                        | 2   | 0                                |
| • Relatie met de businessGr=7                                | 0                            | 2   | 0                               | 0                        | 2   | 3                                |
| • Resultaten data science projectGr=4                        | 0                            | 2   | 0                               | 1                        | 1   | 0                                |
| • Volwassenheid (optiek organisatie)Gr=8                     | 0                            | 2   | 2                               | 3                        | 0   | 1                                |
| • SenioriteitGr=7  | 0                            | 3   | 1                               | 2                        | 0   | 1                                |
| • Leidend wordenGr=1   | 0                            | 0   | 1                               | 0                        | 0   | 0                                |
| • Standaardisatie in technologieGr=1                         | 0                            | 0   | 1                               | 0                        | 0   | 0                                |
| • Uniformiteit in manier van werkenGr=1                      | 0                            | 0   | 1                               | 0                        | 0   | 0                                |
| • Zelfstandig data science oplossingen in productie brengenG | 0                            | 2   | 0                               | 0                        | 0   | 1                                |
| • Data science communityGr=2                                 | 0                            | 1   | 0                               | 0                        | 0   | 1                                |
| • Hybride data science organisatieGr=1                       | 0                            | 0   | 0                               | 0                        | 0   | 1                                |
| • Vrijheid in toolingGr=1                                    | 0                            | 0   | 0                               | 0                        | 0   | 1                                |
| • Data scientistenGr=2                                       | 0                            | 1   | 0                               | 1                        | 0   | 0                                |
| • Best practice process for data scienceGr=1                 | 0                            | 0   | 0                               | 1                        | 0   | 0                                |
| • Softe of randvoorwaardelijke elementenGr=1                 | 0                            | 0   | 0                               | 1                        | 0   | 0                                |
| • Data science oplossing in productie brengenGr=2            | 0                            | 2   | 0                               | 0                        | 0   | 0                                |
| • Betrouwbaarheid data science oplossingGr=1                 | 0                            | 1   | 0                               | 0                        | 0   | 0                                |
| • Nieuwe IT technologieGr=1                                  | 0                            | 1   | 0                               | 0                        | 0   | 0                                |

## Bijlage 7: Matrix met combinaties van subcodes met aantallen

|  | Aanleiding voor data science projecten<br>Gr=11 | Bedrijfscultuur<br>Gr=25 | Belangrijkheid data<br>Gr=8 | Belangrijkheid van data science projecten<br>Gr=8 | Bezig met DS projecten<br>Gr=9 | Urgente data kwaliteit<br>Gr=9 | Volwassenheid (optiek organisatie)<br>Gr=8 | Type data science project<br>Gr=9 | Organisatie data science<br>Gr=14 | Data governance<br>Gr=9 | Data science community<br>Gr=2 | Data science expertise op bestuursniveau<br>Gr=11 | Data toegankelijkheid<br>Gr=11 | Management support<br>Gr=12 | Middelen voor data science projecten<br>Gr=9 | Proces management methodologie<br>Gr=14 | Samenstelling team bij data science projecten<br>Gr=8 | Senioriteit<br>Gr=7 |
|--|---|--------------------------|-----------------------------|---|--------------------------------|--------------------------------|--|-----------------------------------|-----------------------------------|-------------------------|--------------------------------|---|--------------------------------|-----------------------------|--|---|---|---------------------|
| Aanleiding voor data science projecten<br>Gr=11        | 0   | 0                        | 3                           | 0   | 2                              | 2                              | 0  | 1                                 | 0                                 | 0                       | 0                              | 0   | 1                              | 3                           | 0  | 0                                       | 0   | 0                   |
| Bedrijfscultuur<br>Gr=25                               | 0   | 0                        | 2                           | 3   | 0                              | 2                              | 2  | 0                                 | 1                                 | 3                       | 1                              | 3   | 5                              | 4                           | 1  | 1                                       | 0   | 1                   |
| Belangrijkheid data<br>Gr=8                            | 3   | 2                        | 0                           | 1   | 0                              | 1                              | 2  | 0                                 | 0                                 | 0                       | 0                              | 0   | 2                              | 2                           | 1  | 0                                       | 0   | 0                   |
| Urgentie data kwaliteit<br>Gr=9                        | 2   | 2                        | 1                           | 1   | 0                              | 0                              | 2  | 1                                 | 1                                 | 0                       | 0                              | 2   | 2                              | 2                           | 0  | 1                                       | 1   | 0                   |
| Senioriteit<br>Gr=7                                    | 0   | 1                        | 0                           | 0   | 0                              | 0                              | 2  | 0                                 | 2                                 | 0                       | 0                              | 0   | 0                              | 0                           | 0  | 0                                       | 1   | 0                   |
| Type data science project<br>Gr=9                      | 1   | 0                        | 0                           | 0   | 1                              | 1                              | 0  | 0                                 | 3                                 | 0                       | 0                              | 0   | 0                              | 0                           | 0  | 0                                       | 0   | 0                   |
| Volwassenheid (optiek organisatie)<br>Gr=8             | 0   | 2                        | 2                           | 1   | 0                              | 2                              | 0  | 0                                 | 1                                 | 0                       | 0                              | 1   | 0                              | 0                           | 1  | 0                                       | 1   | 2                   |
| Belangrijkheid van data science projecten<br>Gr=8      | 0   | 3                        | 1                           | 0   | 3                              | 1                              | 1  | 0                                 | 2                                 | 0                       | 0                              | 0   | 0                              | 1                           | 1  | 0                                       | 0   | 0                   |
| Bezig met DS projecten<br>Gr=9                         | 2   | 0                        | 0                           | 3   | 0                              | 0                              | 0  | 1                                 | 3                                 | 0                       | 0                              | 0   | 0                              | 0                           | 1  | 0                                       | 0   | 0                   |
| Communicatie uitkomsten data science projecten<br>Gr=8 | 0   | 1                        | 0                           | 0   | 0                              | 0                              | 0  | 0                                 | 0                                 | 0                       | 1                              | 0   | 0                              | 0                           | 0  | 0                                       | 0   | 0                   |
| Data governance<br>Gr=9                                | 0   | 3                        | 0                           | 0   | 0                              | 0                              | 0  | 0                                 | 0                                 | 0                       | 0                              | 3   | 4                              | 0                           | 0  | 0                                       | 0   | 0                   |
| Data science community<br>Gr=2                         | 0   | 1                        | 0                           | 0   | 0                              | 0                              | 0  | 0                                 | 0                                 | 0                       | 0                              | 0   | 0                              | 0                           | 0  | 0                                       | 0   | 0                   |
| Data science expertise op bestuursniveau<br>Gr=11      | 0   | 3                        | 0                           | 0   | 0                              | 2                              | 1  | 0                                 | 0                                 | 3                       | 0                              | 0   | 1                              | 3                           | 0  | 0                                       | 0   | 0                   |
| Management support<br>Gr=12                            | 3   | 4                        | 2                           | 1   | 0                              | 2                              | 0  | 0                                 | 0                                 | 0                       | 0                              | 3   | 1                              | 0                           | 0  | 0                                       | 0   | 0                   |
| Organisatie data science<br>Gr=14                      | 0   | 1                        | 0                           | 2   | 3                              | 1                              | 1  | 3                                 | 0                                 | 0                       | 0                              | 0   | 0                              | 0                           | 1  | 1                                       | 2   | 2                   |
| Proces management methodologie<br>Gr=14                | 0   | 1                        | 0                           | 0   | 0                              | 1                              | 0  | 0                                 | 1                                 | 0                       | 0                              | 0   | 0                              | 0                           | 0  | 0                                       | 2   | 0                   |
| Samenstelling team bij data science projecten<br>Gr=8  | 0   | 0                        | 0                           | 0   | 0                              | 1                              | 1  | 0                                 | 2                                 | 0                       | 0                              | 0   | 0                              | 0                           | 0  | 2                                       | 0   | 1                   |
| Data toegankelijkheid<br>Gr=11                         | 1   | 5                        | 2                           | 0   | 0                              | 2                              | 0  | 0                                 | 0                                 | 4                       | 0                              | 1   | 0                              | 1                           | 0  | 0                                       | 0   | 0                   |