# D2.3.3 Evaluation results of the LinkedUp VICI competition

**Citation for published version (APA):**

Drachsler, H. (2014). *D2.3.3 Evaluation results of the LinkedUp VICI competition*.

**Document status and date:**
Published: 31/10/2014

**Document Version:**
Peer reviewed version

**Document license:**
CC BY-NC-SA

**Please check the document version of this publication:**

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

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.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:

https://www.ou.nl/taverne-agreement

**Take down policy**
If you believe that this document breaches copyright please contact us at:

pure-support@ou.nl

providing details and we will investigate your claim.

Downloaded from https://research.ou.nl/ on date: 30 Nov. 2022

Open Universiteit
www.ou.nl

**LinkedUp: Linking Web Data for Education Project – Open Challenge in Web-scale Data Integration**

http://linkedup-project.eu/

Coordination and Support Action (CSA)

Grant Agreement No: 317620

# D2.3.3 Evaluation results of the LinkedUp VICI competition


**Deliverable Coordinator:** **Drachsler, Hendrik**

**Deliverable Coordinating Institution:** **Open University of the Netherlands** (OUNL)

**Other Authors:** Slavi Stoyanov
Eelco Herder
Marieke Guy
Maren Scheffel


| | | | |
|---|---|---|---|
| Document Identifier: | LinkedUp/2014/D2.3.3 | Date due: | 31.10.2014 |
| Class Deliverable: | LinkedUp 317620 | Submission date: | 27.10.2014 |
| Project start date: | November 1, 2012 | Version: | 1.0 |
| Project duration: | 2 years | State: | |
| | | Distribution: | Restricted |

29.04.2013 © Copyright lies with the respective authors and their institutions.

# LinkedUp Consortium

This document is part of the LinkedUp Support Action funded by the ICT Programme of the Commission of the European Communities by the grant number 317620. The following partners are involved in the project:


| | |
|---|---|
| **Leibniz Universität Hannover (LUH)**<br>Forschungszentrum L3S<br>Appelstrasse 9a<br>30169 Hannover<br>Germany<br>Contact person: Stefan Dietze<br>E-mail address: dietze@L3S.de | **The Open University**<br>Walton Hall, MK7 6AA<br>Milton Keynes<br>United Kingdom<br>Contact person: Mathieu d'Aquin<br>E-mail address: m.daquin@open.ac.uk |
| **Open Knowledge Foundation Limited LBG**<br>St John's Innovation Centre<br>Cowley Road<br>CB4 0WS, Cambridge<br>United Kingdom<br>Contact person: Marieke Guy<br>E-mail address: marieke.guy@okfn.org | **ELSEVIER BV**<br>Radarweg 29,<br>1043NX AMSTERDAM<br>The Netherlands<br>Contact person: Michael Lauruhn<br>E-mail address: M.Lauruhn@elsevier.com |
| **Open Universiteit Nederland**<br>Valkenburgerweg 177,<br>6419 AT Heerlen<br>The Netherlands<br>Contact person: Hendrik Drachsler<br>E-mail address: Hendrik.Drachsler@ou.nl | **Lattanzio Learning S.p.A.**<br>Via Domenico Cimarosa, 4<br>20144 Milano<br>Italy<br>Contact person: Elisabetta Parodi<br>E-mail address: parodi@lattanziogroup.eu |


# Work package participants

The following partners have taken an active part in the work leading to the elaboration of this document, even if they might not have directly contributed to the writing of this document or its parts:

OUNL, Hendrik Drachsler, Slavi Stoyanov, Maren Scheffel
LUH, Eelco Herder, Stefan Dietze, Ivana Marenzi
OUUK, Mathieu d'Aquin
OKF, Marieke Guy

# Executive Summary

This document D2.3.3 is the final report of Task 2.4 – Evaluation of challenge submissions. Task 2.4 is about the actual assessment of the participating projects within the LinkedUp Veni, Vidi and Vici competition on the basis of the LinkedUp Evaluation Framework (D2.2.1).

The main objective of Task 2.4 is to summarise and report the outcomes of the various competitions and analyse the practical experiences of the experts with the LinkedUp Evaluation Framework to further improve the evaluation framework.

In the current document D2.3.3 we report about the Linked Data tools and ideas that have been submitted to the third and final data competition - Vici. In total, we received 13 submissions, 10 of them have been shortlisted and invited to a poster presentation at the 13th International Semantic Web Conference (ISWC 2014), four of them have been awarded by the LinkedUp evaluation procedure. In addition, an audience price has been awarded at the ISWC.

This deliverable shortly lists the Vici submissions, explains the evaluation procedure that resulted in a short list of the best submissions, justifies the decision for the winners, and also reports the experiences collected within Vici.

31.10.2014© Copyright lies with the respective authors and their institutions.

# Table of Contents

**1. Introduction ..... 5**

**2. The VICI Competition ..... 6**

2.1. The Scoring Sheet for the Vici competition ..... 12

2.2. Evaluation results ..... 14

2.2.1. First phase evaluation ..... 14

2.2.2. Second phase evaluation ..... 16

2.2.3. Deliberation process ..... 20

2.2.4. The audience award ..... 21

2.2.5. Feedback to the LinkedUp participants ..... 22

**3. Methods for the Evaluation of LinkedUp Evaluation Framework ..... 23**

3.1. Method ..... 23

3.1.1. Participants ..... 23

3.1.2. Procedure ..... 23

3.1.3. Measurement Instrument ..... 23

3.2. Summary of findings ..... 23

3.2.1. Closed-ended items ..... 24

**4. CONCLUSIONS ..... 27**

**Appendix A. Responses of judges to open questions in Vici evaluation survey ..... 28**

# 1. Introduction

This document D2.3.3 is the final report out of three deliverables (D2.3.1 and D2.3.2 being the other two) of Task 2.4 - Evaluation of challenge submissions. Task 2.4 is about the actual assessment of the participating projects within the LinkedUp Veni, Vidi and Vici competition making use of the LinkedUp Evaluation Framework (D2.2.1).

The main objective of Task 2.4 is to summarise and report the outcomes of the various competitions and analyse the practical experiences of the experts with the LinkedUp Evaluation Framework (EF) to further improve it. This deliverable reports about the Vici competition and is the final approach to improve the EF before summarising the insights gained from the evaluation studies in D2.2.2. In the following section, we will shortly introduce the Vici competition and its specific foci.

31.10.2014© Copyright lies with the respective authors and their institutions.

# 2. The VICI Competition

The Vici Competition was the third and final in the series of competitions comprising the LinkedUp Challenge. The competition was promoted through the LinkedUp Project website and the LinkedUp Challenge website, a site dedicated purely to promoting the challenge. Vici ran from the 26th of May 2014 to the 5th of September 2014 and requested participants to submit "*advanced prototypes and tools that are driven by linked and/or open data for educational purposes*", with the remark that "*Your tool should be mature and stable; it should be used or have been used by a fair amount of users on a realistic scale*". The LinkedUp Challenge website defines "*educational purposes*" by stating that the tools and applications developed must be relevant to education – in the broadest sense of the word. This might mean that they aid learning in some way or that they support educational objectives by expanding knowledge and encouraging critical thinking.

Apart from the Open Track, the Vici competition featured two Focused Tracks, which were developed in close cooperation with the Commonwealth of Learning (CoL) and Elsevier. The focused track *Supporting Developing Countries*, developed in collaboration with CoL, looks for educational applications that target developing countries, addressing context-specific problems, issues and needs, either technical, societal or environmental. The focused track *Water Resources & Ecology*, developed together with Elsevier, involves enhancing journal article content along with related research statistics and datasets to assist in discovery, learning and interpretation of disparate content and data. A number of Elsevier datasets have been made available especially for this track.

By the closing date, we received a total of 13 submissions, which all looked very promising. Among the participants, there were some familiar names, such as FAO of the United Nations and GNOSS, who have already submitted to our earlier competitions and have extended their work. Several submissions stem from companies who aim to showcase their work. Academics from all over the world, from Brazil to New Zealand, also participated. The focused track *Supporting Developing Countries* received two submissions: an electronic repository for Russian historical statistics and a serious game that engages participants in critical analysis of text. The focused track *Water Resources & Ecology* got one submission, in which research works were visualised in various ways.

A full description of the Vici submissions can be found in WP1 deliverable *D1.3 – LinkedUp Challenge results.* In the following we shortly list the submissions to put the evaluation process in context.

Table 1: Short overview of all submissions to the Vici competition. Please note: ID is the submission ID from EasyChair and therefore does not follow a normal 1-13 count per row.

| ID | Track | Authors | Title | Abstract |
|---|---|---|---|---|
| **1** | OT | Fabrizio Celli and Johannes Keizer | AGRIS: the hub for agricultural research | AGRIS is currently a big database of scientific and socio-economic publications in the food and agricultural domain, with more than 65 active institutions continuously feeding this database with periodic updates. The AGRIS Web portal is highly visited, reaching peaks of 250,000 visits/month from more than 200 countries and territories. The current system is mostly based on RDF and allows the creation of mashup pages to integrate many different sources of information available on the Web, related to the AGRIS domain. The AGRIS objective is to make information on agricultural research globally available and to help its users finding all they need.<br><br>Website: http://agris.fao.org/<br><br>Example mash-up site: http://agris.fao.org/agris-search/search.do?recordID=JP2010001379 |
| **2** | FT | Ricardo Kawase, Patrick Siehndel and Ujwal Gadiraju | Visualizing Research Works in the Water Resources Industry | With the increasing practice of making data openly available, nowadays there is a growing amount of information easily available pertaining to water resources and ecology. Scientific works by researchers across the world contribute to the abundance in such data. Major challenges that emerge due to the volume of data include the discovery of useful and relevant content, as well as learning and interpretation of the various disparate content. In this work, we aim to aid researchers and interested stakeholders in understanding the vast landscape of scientific research in the water resources industry. We integrate different sources of data from the Web; journals from Elsevier, tweets from Twitter, and wikipedia annotations. We use interactive visualizations in order to engage the users and satisfy their information needs.<br><br>Website: http://www.l3s.de/~kawase/vici/ |
| **3** | OT | Shaoqun Wu, Alannah Fitzgerald and Ian Witten | FLAX: Flexible Language Acquisition with Linked and Open Data-Driven Learning - See more at: http://linkedup-challenge.org/vici/#sthash.8FhcW5lO.dpuf | Open Data, Open Educational Resources (OERs) and Open Access (OA) research publications provide a compelling opportunity for language learning. OERs and OA publications supply large corpora of linguistic material relevant to particular subject areas, including text, supplementary images (slides), audio and video. Such domain-specific corpora can be automatically analysed and enriched by linking to larger open linguistic datasets, including those derived from Wikipedia and Google web dumps.<br><br>Website: http://flax.nzdl.org/ |

© Copyright lies with the respective authors and their institutions.

| | | | | |
|---|---|---|---|---|
| **4** | OT | Ricardo Alonso Maturana, Ana Moreno, María Elena Alvarado and Susana Lopez | Didactalia: Building and Taking Advantage of Large Educational Knowledge Graph to Improve and Accelerate the Teaching-Learning Process | Didactalia.net is a large educational community for parents, teachers and students, supported by the social and semantic platform GNOSS. The collection offers over 85,000 educational resources, from pre-school education to high school, structured according to the standards of the semantic web (OWL/RDF) and the principles of Linked Open Data Web.<br>In 2013 Didactalia had 7,8M visits who consumed about 24,5M pages. This year we expected to reach near 10M visits. The community is growing 1,000 registered users a day.<br><br>Website: http://www.didactalia.net/en<br>Introductory video:<br>https://www.youtube.com/watch?v=UtRyq8CD6ak |
| **7** | OT | José Luis Redondo-García, Mariella Sabatino, Pasquale Lisena and Raphaël Troncy | HyperTED: Exploring TED talks at the fragment level | HyperTED offers a unique and innovative way for exploring over 1681 TED talks at the fragment level, opening a window to a more effective and efficient dissemination of the knowledge present in those inspirational conferences.<br>In this application, the concept of complete video talk as a first class citizen is further refined through the notion of chapters. Similar to paragraphs in a textual document, chapters delimit particular ideas that sequentially illustrate the context of the video. In practice, viewers generally fast browse a video without necessarily watch each particular chapter. HyperTED tackles this problem by proposing a set of automatically annotated media fragments called Hot Spots which intend to highlight the main concepts and topics discussed in a video talk.<br><br>Website: http://linkedtv.eurecom.fr/mediafragmentplayer<br>Base code: https://github.com/pasqLisena/MediaFragPlayerDemo |
| **10** | OT | Laurens De Vocht and Selver Softic | ResXplorer | Research information is widely available on the Web. Peer-reviewed research publications as well related meta data from bibliography archives offers a vast of information to investigate about related publications, events and persons for a researcher. Usually the platforms supporting this information exchange have an API that allows access to the structured content or the information is already present as Linked Data. This opens a new way to search and explore research information.<br>With the user interface of ResXplorer we help researchers to get an overview by using an approach that visualizes interactively search process in an aligned linked data knowledge base of research related resources.<br><br>Website: http://www.resxplorer.org |
| **11** | OT | Antonio Roa-Valverde | Word Bucket | Word Bucket is a vocabulary learning app that solves the common problems faced by language learners of 'forgetting and losing' the words they are learning. The application encourages language students to search for words in an inbuilt translation dictionary and save the words they find into their personalised 'Bucket'. They're then encouraged to learn the words at various levels of difficulty through tests and games. A traffic light colour system shows progression of words from white (new) to green (learned), with a specially created game at each level. |

| | | | | |
|---|---|---|---|---|
| | | | | Website: http://www.wordbucket.com |

31.10.2014© Copyright lies with the respective authors and their institutions.

| | | | | |
|---|---|---|---|---|
| **12** | OT | Navino Evans and Sean McBirnie | Histropedia – The Timeline of Everything | Histropedia is an open data project aiming to become an interactive timeline of everything in history from the big bang to the present day, allowing anyone to freely explore timescales of billions of years down to individual days. As well allowing the free exploration of any period in history the timeline will also be filterable to customise the contents of the timeline.<br><br>Company website: http://www.histropedia.com/<br>Timeline app: http://histropedia.uat.co/ |
| **13** | FT | Vyacheslav Tykhonov | Electronic repository for Russian Historical Statistics | The repository operates on the principle of a historical data-hub, bringing together data extracted from various published and unpublished sources in one place. Its principal focus is Russian economic and social history of the three centuries (18th-21st)<br>The five principal lines of inquiry aim to provide a basic set of indicators for measuring social and economic development: population, labour and employment, land, capital, production (agriculture, industry, services).<br><br>Website: http://node-146.dev.socialhistoryservices.org/datasets/indicators |
| **14** | FT | Bernardo Pereira Nunes, Luiz Guilherme A. Roland, Clara S. Ishikawa, Ângela de Araújo and Gilda Helena Bernardino De Campos | ISCOOL – A serious game powered by Semantic Web technologies | ISCOOL is an informal learning environment, which encourages participants to engage in critical analysis of text. Although applicable to many disparate demographics, two audiences in particular have been identified in the first instance: Firstly, people of low income households with limited access to educational facilities and secondly, those with disabilities (hearing impairment in particular).<br><br>Website: http://research.ccead.puc-rio.br/iscool/ |
| **16** | OT | Michael Howe | GroupMOOC | As macroeconomic trends redefine sources of competitive advantage, organisations today face a constant challenge in upskilling and reskilling their employees. While Massive Open Online Courses ('MOOCs') represent a powerful resource from top universities, they have not been appropriately tapped for organisational learning.<br><br>Website: http://www.groupmooc.com/<br>Free iPhone prototype: GroupMOOC |
| **17** | OT | Andras Benedek and Gyuri Lajos | WikiNizer™ Research: A Personal Knowledge Graph Builder Harnessing Freebase Linked Data | WikiNizer™ Research (WikiNizeR) is a visual Wiki-like knowledge orgaNizer which constructs Personal Knowledge Graphs. By enabling us to visualize "meta" levels of reflection, WikiNizeR facilitates our sense making and problem solving.<br><br>Demo: http://alpha.wikinizer.com/LinkedUpChallenge/krnl/view/arbor/demo.htm?a=G1148784<br><br>Introduction slides:<br>http://wikinizer.com |
| **18** | OT | Jaspreet Singh, Zeon | LearnWeb-OER: | The LearnWeb-OER platform enhances collaborative searching and sharing of educational resources providing specific |

| | | | | |
|---|---|---|---|---|
| | | Trevor Fernando and Saniya Chawla | Improving Accessibility of Open Educational Resources | means and facilities for education. In the following, we provide a description of the functionalities that support users in collaboratively collecting, selecting, annotating and discussing search results and learning resources.<br><br>Website: http://learnweb.l3s.uni-hannover.de/ |

© Copyright lies with the respective authors and their institutions.

## 2.1. The Scoring Sheet for the Vici competition

A few changes have been made in the Scoring Sheet for the Vici open and focused track competitions, based on the evaluation carried out after the Vidi reviewing process. They are as follows:

- Comparison to other existing applications in the formulation of the items referring to effectiveness and efficiency was removed.
- The two items operationalising the *Data* criterion, namely 'Consuming multiple data' sources' and 'Exposing new datasets to the Linked Data cloud' were better formulated.
- A weighting was applied to the criteria *Educational Innovation* and *Data* to reflect their importance. The LinkedUp project is about linked open data, with education as its application domain. The linked open data applications should address an educational issue and try to improve learning and teaching.

The final set of criteria and indicators for the Vici competition are presented in the space below:

Table 2: Overview of evaluation items for the Vici competition per criteria.

| | |
|---|---|
| Educational Innovation (EI) | 1. Rate the extent to which the application implements an innovative educational concept (e.g. innovative ways of presenting content, innovative methods for learning or teaching).<br>2. Rate the extent to which the application is effective (e.g. leads to significant improvements in learning or teaching).<br>3. Rate the extent to which the application is efficient (e.g. saves time or efforts for learners or teachers). |
| Usability (U) | 1. Rate the extent to which the application is easy to use.<br>2. Rate the extent to which the application can quickly be learned.<br>3. Rate the extent to which the application has an attractive user interface. |
| Performance (P) | 1. How is the tool available to its target users?<br>2. How would you rate the overall quality of the tool according to the aims of the tracks?<br>Open Track: To what extent does the tool integrate open data to improve education?<br>Focused Track 1 Supporting Developing Countries: To what extent does the tool integrate open data to improve education in developing countries?<br>Focused Track 2: Water Resources & Ecology: To what extent does the tool assist in increasing knowledge and a better understanding of issues on Water Resources & Ecology OR developing countries? |
| Data (D) | 1. Does the tool consume multiple data sources?<br>2. Does the tool expose new datasets to the Linked Data cloud? |
| Legal (L) | 1. Does the tool provide background / licensing information for the used data sources? |

| | |
|---|---|
| | 2. Does the tool collect only needed personal information about the user?<br>3. Does the tool provide a statement on the terms of use? |
| Audience (A) | 1. Rate the extent to which the application addresses the needs, problems or challenges of its target group(s).<br>2. Can the application be used in various domains? |

The judges rated the individual tools according to the following instructions:

*DETAILED REVIEW PLAN:*
*Please follow the following steps during your reviews.*

*1. SCAN the submissions assigned to you – see the list at the end of this email. In each paper, there is a link to the demo site, either in the main text or in the references. Watch the demo or do a live test of the tools.*

*2. START REVIEW by selecting the button Add Review. You will see the evaluation form.*

*3. RATE your assigned submission(s) according to the criteria in the evaluation form. We apply six evaluation criteria with some specific indicators.*
*These are:*

- *EI. Educational Innovation*
- *U. Usability*
- *P. Performance*
- *D. Data*
- *L. Legal & Privacy*
- *A. Audience*

*Once you have provided your ratings for the criteria, please give a short justification of your ratings in the Review text box by addressing each of the six criteria. You can use the following text template for your feedback.*

*Template for justification of the rating:*

*EI. Educational Innovation:*
*[YOUR REVIEW COMMENTS]*

*U. Usability of the tool:*
*[YOUR REVIEW COMMENTS]*

*P. Performance:*
*[YOUR REVIEW COMMENTS]*

*D. Data:*
*[YOUR REVIEW COMMENTS]*

*L. Legal & Privacy:*
*[YOUR REVIEW COMMENTS]*

31.10.2014© Copyright lies with the respective authors and their institutions.

*A. Audience:*
*[YOUR REVIEW COMMENTS]*

*4. SUBMIT the review when ready in EasyChair.*

## 2.2.Evaluation results

In the final Vici competition, we applied a two-phase evaluation process to the submissions. Similar to the two-phase process of the EC for research proposals, we first looked at all submissions according to their ratings and reviewer comments and whether they fulfil the requirements of the Vici competition call. Then, in the second phase, we applied weighting of criteria to explore strengths and weaknesses of the Vici competition participants according to the specific call objectives. In the following sections we will summarise the experiences gained and actions taken within this two-phase approach.

### 2.2.1. First phase evaluation

In the first stage of the evaluation we needed to select a shortlist of max. 10 submissions that were to be invited to the ISWC conference. Apart from the core objectives in regard to 'Educational Innovation' and 'Linked Open Data', another condition for making the shortlist was the requirement for being a mature tool and also worth to be presented to the ISWC audience. Figure 1 shows the total ratings given to each submission by summing up all single rating values per submission.

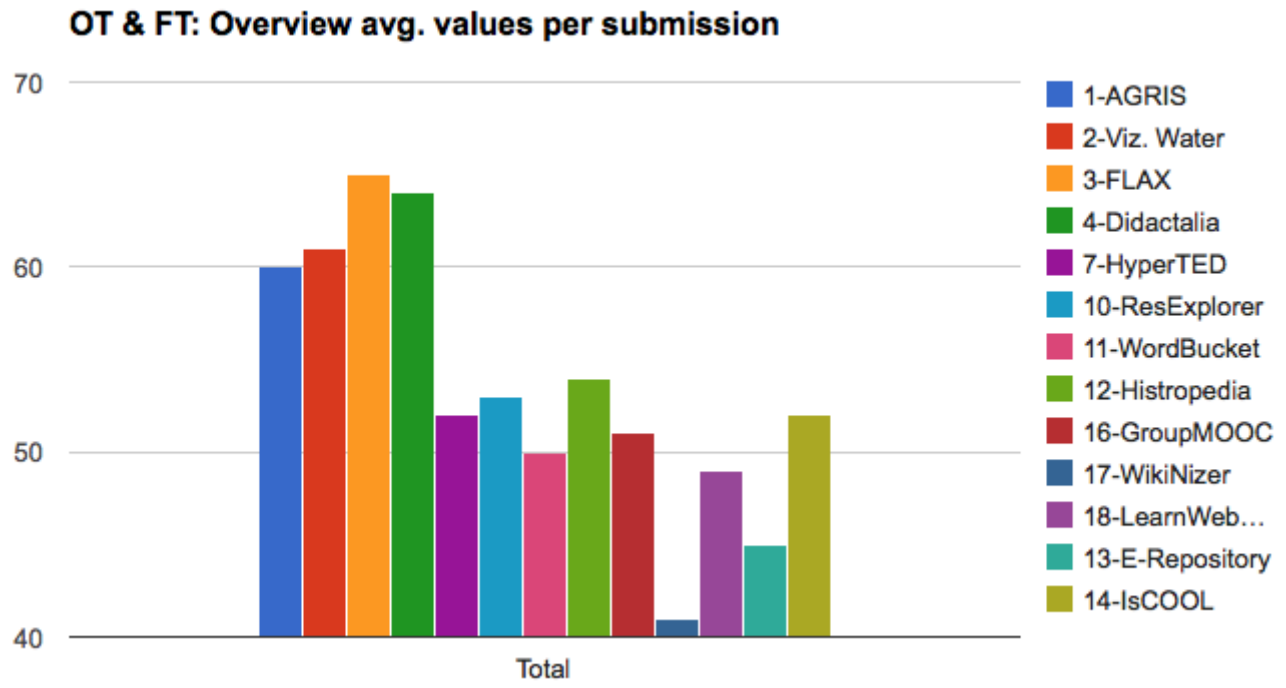



**Figure 1:** Overview of ratings per criteria given to submissions Focused Tracks and Open Track.

From Figure 1 it becomes clear that submission *17 - WikiNizer* (41 points) and *13 - Electronic Repository* (44 points) are far behind all other submissions. After an additional check, the members of the LinkedUp Team came to the conclusion that both tools are not as mature as their competitors.

*WikiNizer* was very difficult to use and looked like an early prototype. The *Electronic Repository* submission was more a dataset than a mature tool.

A more difficult decision concerned the 10th place of the shortlist. Submission *18 - LearnWeb OER* (49 points), *16 - GroopMOOC* (50 points), and *11 - Wordbucket* (51 points) have been rated almost equally and required a thorough inspection by the LinkedUp consortium. The review ratings of all these applications are quite comparable (the sum of ratings as well as the ratings in each category), which made the decision a difficult one. Moreover, all submissions are mature and working applications.

However, we agreed that there were some good reasons for excluding *Word Bucket* from the short list. The reviewers rated the *Educational Innovation* of *WordBucket* very low. They made clear that there were already many vocabulary trainers available and some of them are even more innovative (e.g. Duolingo) than the *Word Bucket* approach.

Another critical aspect of *Word Bucket* was the lack of Linked Data sources used (*Word Bucket* also received the lowest score on this aspect). Even though *Word Bucket* is available and working, and a nice app, it does not demonstrate the use of linked or open data. Both *LearnWeb-OER* and *GroupMOOC* use Linked Open Data sources and enrich those with additional functionalities. For *Word Bucket* it is not really clear which data is used and how it is used. Furthermore, the reviewers emphasised the fact that the provided vocabulary to learn is not a very good and reliable source. One reviewer commented that this makes the submission out of scope for the competition – and we tend to agree with this. All reviewers reflected this also in their recommendation for acceptance in EasyChair (*Word Bucket* received the lowest recommendation of all submissions).

The scores on indicators of the EF and the deliberation process indicated to include the following 10 submissions in the shortlist and invited them to the ISWC conference:

1. ID 3: FLAX (OT score: 65)
2. ID 4: Didactalia (OT score: 64)
3. ID 2: Vis. of Water Resources (FT score: 61)
4. ID 1: AGRIS (OT score: 60)
5. ID 12: Histropedia,(OT score: 54)
6. ID 10: ResXplorer (OT score: 53)
7. ID 14: ISCOOL (FT total score: 52)
8. ID 7: HyperTED (OT total score: 52)
9. ID 16: GroupMOOC, (OT score: 51)
10. ID 18: LearnWeb-OER (OT score: 49)

Not on the shortlist:

11. ID 11: WordBucket (FT score: 50 points)
12. ID 13: Electronic Repository (FT score: 44 points)
13. ID 17: WikiNizer (FT score: 41 points)

31.10.2014© Copyright lies with the respective authors and their institutions.

### 2.2.2. Second phase evaluation

After the first stage evaluation was completed and the notifications were sent to the shortlisted candidates, we started the second evaluation phase that needed to identify the winning submissions.

As suggested in the review meeting of the LinkedUp Project in January 2014, we applied *weighting* of criteria to the submissions in order to highlight the submissions that have addressed in the best possible way the central objectives of the Vici call. This initial plan became even more important as we had very few submissions for the FTs (one submission for each of the FTs *Water Resources & Ecology* and *Developing Countries*). It meant no real competition. The consortium decided to evaluate the FTs not independently from the OT.

The weighting aimed at amplifying the core objectives of the Vici call - *linked and open data for educational purposes*. Therefore, the criteria *1. Linked Open Data* and *2. Educational Innovation* have been multiplied by a factor of 1.5. We selected a weighting value of 1.5 after investigating several approaches to weighting. A factor of 1.5 seems to be strong enough to make a distinction in the ranking (a factor between 1.1 to 1.4 might be too weak), whereas a higher factor (1.6 to 1.9) might be too strong and even diminish the effect of other criteria in the EF. Figure 2 shows the initial rating of the submissions and Figure 3 the amplified results after the weighting.

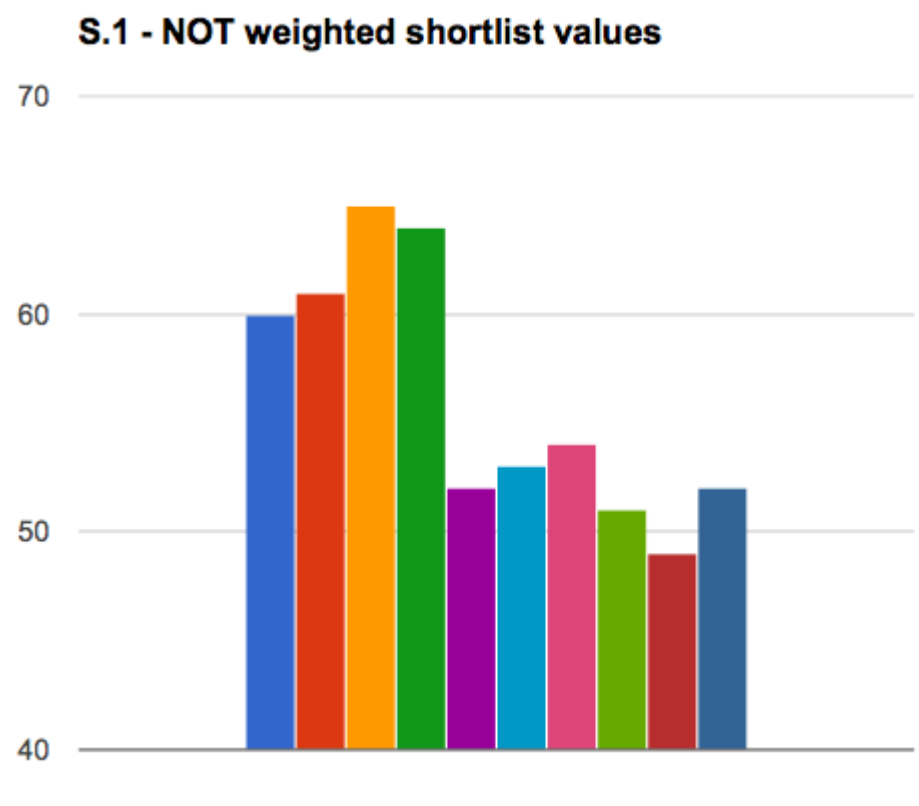



**Figure 2:** Stage 1- All shortlisted submissions NOT weighted

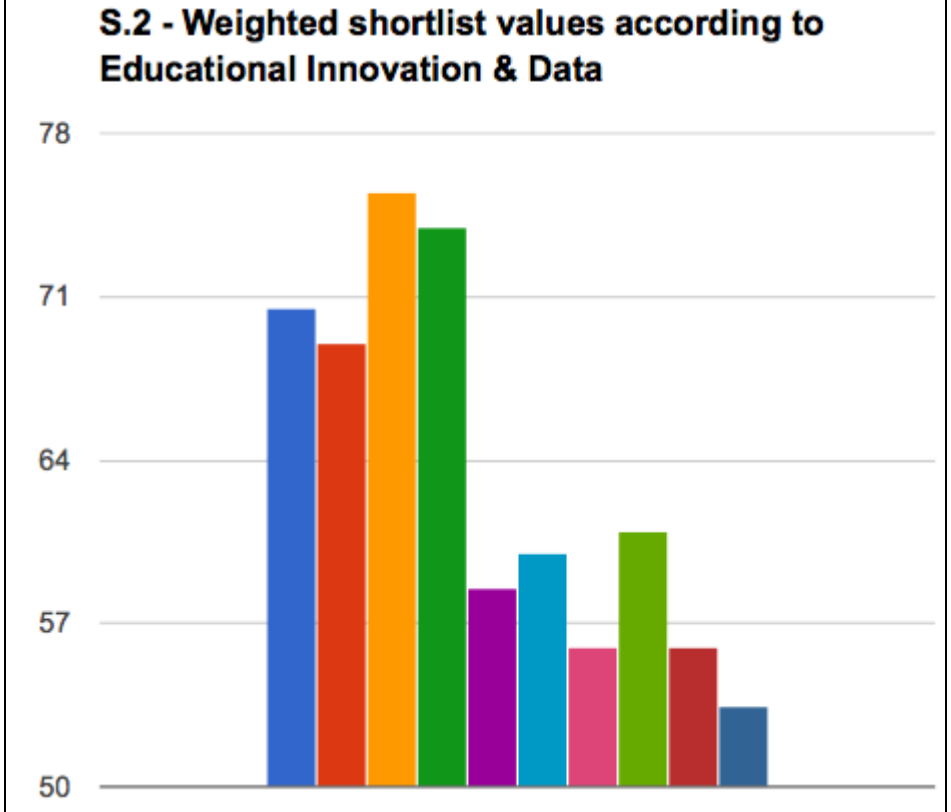
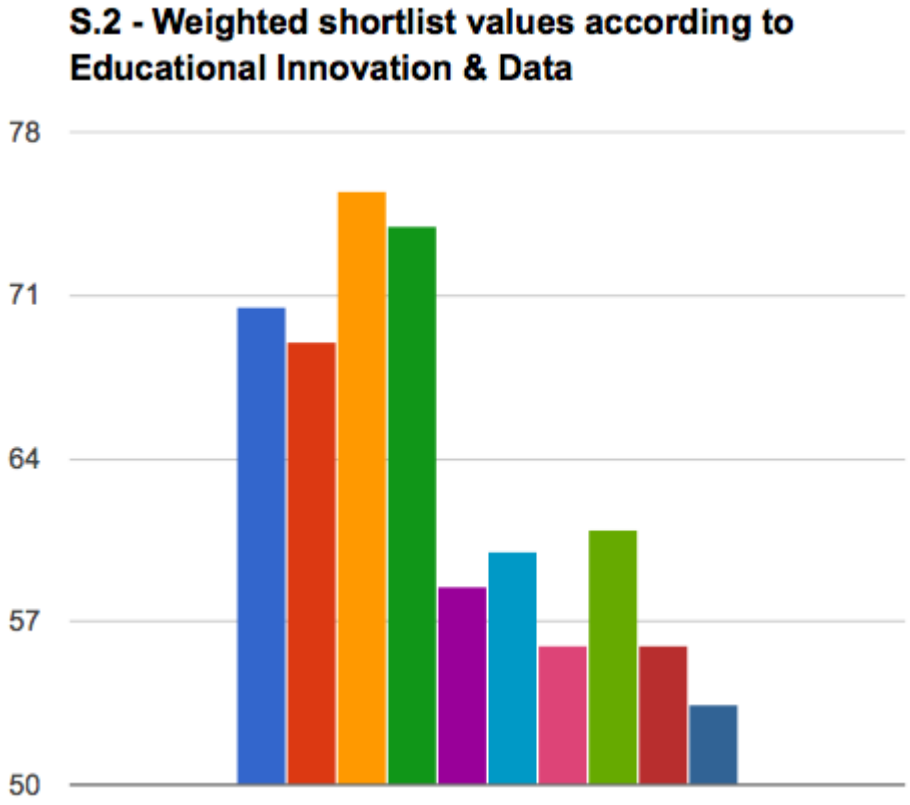



**Figure 3:** Stage 2 - All shortlisted submissions weighted according to *1. Educational Innovation* & *2. Data*.

© Copyright lies with the respective authors and their institutions.

As we can see, the weighting of the criteria (*1. Educational Innovation* & *2. Data*) affected the ranking of the submissions especially for the lower rated submissions. However, in both Figures 2 and 3, submission *3 - FLAX* remains the highest rated submission followed by number *4 - Didactalia* and number *1 - AGRIS* from the OT. The weighting of criteria actually had a negative effect for the FT submissions number *2 - Water Visualization* and number *14 - ISCOOL*. By amplifying the core evaluation criteria, the submission *1 - AGRIS* from the OT could take over the third place from submission *2 - Water Viz.* of the FT. Although there have been some changes to the low rated submissions, the top four submissions, including one FT submission, remain well above the other submissions.

As the FT submissions have been developed addressing different objectives than the OT submissions, we decided to balance the disadvantage by *weighting* the item - *How would you rate the quality of the tool according to the track aims? To what extent does it integrate open data: -for education (Open Track), to improve education in developing countries (FT1), to better understand issues of Water Resources (FT2)* by the factor 1.5 to compensate the disadvantage to be evaluated with submissions that have been designed for different purposes. Figure 4 shows the adjusted ranking of the submissions after the weighting of the FT.




Here we present again Figure 3 to make it easier for the reader to see the changes of the weighting.

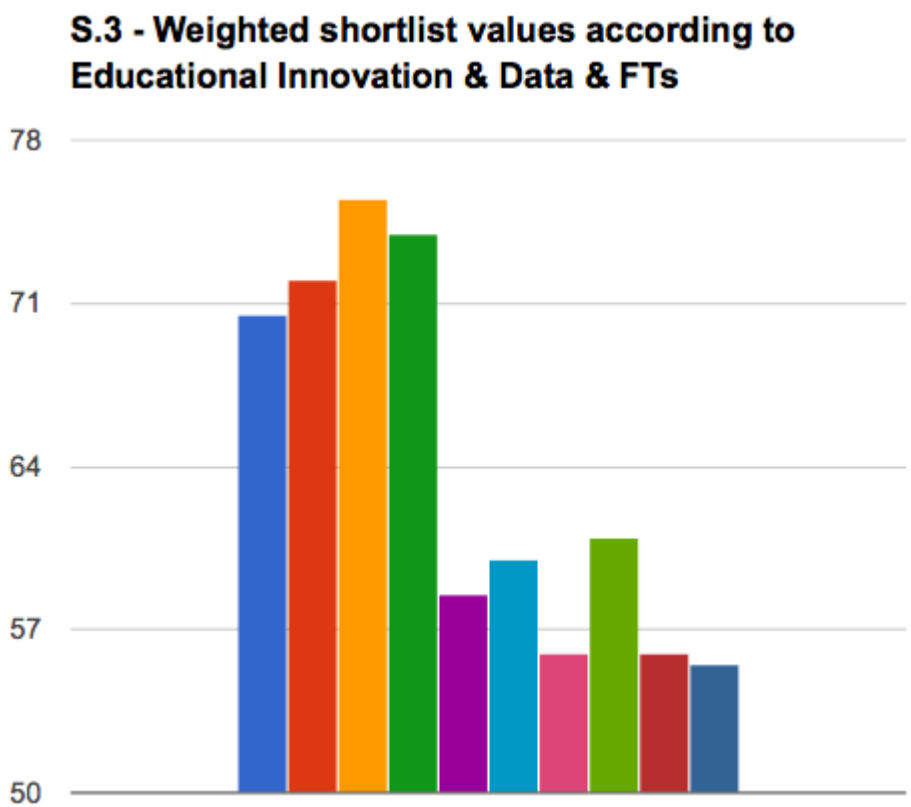



**Figure 4:** Weighted shortlist with FT receive additional weighting only for item P2 due to their disadvantage being in evaluation of OT.

© Copyright lies with the respective authors and their institutions.

As Figure 4 shows, the weighting of the FT submissions naturally let them take advantage in the overall ranking of the submissions. Balancing their disadvantage by multiplying their FT evaluation item however did not change their overall ranking too strongly. They just could win a little bit through the amplification of their target group item. However, submission *2 - Water Viz.* took over the third place from submission *1 - AGRIS* again.

### 2.2.3. Deliberation process

As in the previous competitions, all Vici submissions were thoroughly discussed by the LinkedUp deliberation committee to make sure that the best applications made the final list and that all received a fair treatment. The deliberation process was conducted face-to-face during the project consortium meeting in Milton Keynes (October, 2014). The deliberation committee consisted of representatives from all partner organisations. During the deliberation all reviews comments, feedback from the shortlisted submissions, and the expert judgement have been considered to make the final decision.

We entered into the deliberation meeting with the following ranking of submissions:

1. Place: Submission *3 - FLAX*
2. Place: Submission *4 - Didactalia*
3. Place: Submission *1- AGRIS* OR Submission *2 - Water Viz.* depending on the weighting.

It was not decided so far how to proceed with the other FT submission *14 - ISCOOL* that also got a place in the shortlist but was not ranked as one of the top 3 submissions.

The deliberation committee took two important decisions. First of all it was decided, that there would be no FT winners as the FT merged with the OT. Second, the weighting of the evaluation indicators was used to get additional views on the data and explore the strengths and weakness of the submissions. The consortium decided finally to give four awards: a first and second place and a shared third place.

The winners of the Vici competition therefore are:

- 1st place: ID 3: *FLAX* (OT score: 65)
  FLAX is an outstanding open-source software system designed to automate the production and delivery of interactive digital language collections. It provides data for language learning in a new and innovative way by combining corpora and support teachers in reacting automated exercises and assessments with these corpora. Exercise material comes from digital libraries (language corpora, web data, open access publications, open educational resources) for a virtually endless supply of authentic language learning in context. FLAX has been designed for language teachers, learners studying languages, subject-matter specialists, instructional designers and e-learning support teams. The tool supports building own language collections. It therefore, can also be used as a tool to generate datasets for its community. The FLAX software can be freely downloaded to build language collections with any text-based content and supporting audio-visual material, for both online and classroom use.

- 2nd place: ID 4: *Didactalia* (OT score: 64)
  Developed by the SME GNOSS, who also participated in the previous Veni competition. It is a powerful system that has millions of users on a regularly basis in Spain and explicitly addresses educational stakeholders. The Didactalia system is a lighthouse example for the use of Semantic Web and Linked Data and helps you to browse, find and use learning material on many different topics, for different age groups and from various educational repositories. The reviewers found it an outstanding initiative and a great example for other countries apart from to Spain.

- 3rd shared place:
  As both submissions are very strong, received almost the same scores, and where changing places during the weighting procedure we decided to award both with a third place.
    - ID 2: *Viz. of Water Resources* (FT score: 61)
      The Visualization of Water Resources & Ecology provides rich means to search journals, tweets and Wikipedia annotations. The interactive visualizations address the Focus Track - Water Resource, proposed and supported by Elsevier, to see how linked data can be used for making the learning experience more effective and appealing.

    - ID 1: *AGRIS* (OT score: 60)
      AGRIS is a system to support Agricultural Science and Technology worldwide. It contains a bibliographic database, collecting metadata from every side of the world, in various languages. AGRIS can rely on over 150 data providers located in more than 100 different countries. The application has a huge collection of bibliographic references in agriculture attracting thousands of visitors on a monthly basis. The systems tremendously improved since the last Vidi competition according to the reviewer comments esp. by completely exposes its database as Linked Open Data.

## 2.2.4. The audience award

The People's Choice Vote is run in parallel to the EF assessment and gives competition entrants an opportunity to promote their submission, and external people an opportunity to be involved in the voting process. Through running such a vote we hope to be able to share our experiences, which may prove to be a useful resource for others planning to run a similar open vote in the future. For these reasons we have been keen to explore various different voting options.

For the Vici Competition we encountered problems with the software selected resulting in scores that are effectively null and void. The decision was taken to replace the online People's Choice Vote with an offline vote. The People's Choice winner was identified during a physical People's Choice Vote that took place at the poster session at ISWC. Tokens were issues to ISWC attendees and they were encouraged to vote throughout ISWC up to the end of the poster session.

31.10.2014© Copyright lies with the respective authors and their institutions.

### 2.2.5. Feedback to the LinkedUp participants

Based on the evaluation collected from the judges in the EasyChair system we provided each submission with an overview of their performance according to the evaluation criteria. We took the sum of the assessment scores of the judges for each submission and report it with the average value for the evaluation criteria in the following template.

**[PROJECT ACRONYM] : [FULL TITLE]**
**Rating given by the judges ordered by 1. Your rating | 2. Average rating:**

| | | |
|---|---|---|
| Educational Innovation: | [INDIVIDUAL RATING] \| AVG Educational Innovation: | 29 |
| Usability: | [INDIVIDUAL RATING] \| AVG Usability: | 25 |
| Performance: | [INDIVIDUAL RATING] \| AVG Performance: | 18 |
| Data: | [INDIVIDUAL RATING] \| AVG Data: | 20 |
| Legal: | [INDIVIDUAL RATING] \| AVG Legal: | 20 |
| Audience: | [INDIVIDUAL RATING] \| AVG Audience: | 24 |

**Comments provided by the judges:**
[WRITTEN COMMENTS MADE BY THE JUDGES]

# 3. Methods for the Evaluation of LinkedUp Evaluation Framework

Apart from reporting the outcomes of the Vici competition, one of the main objectives of this deliverable is to capture the experience of the judges in applying the EF to the Vici submissions. Given the anticipated shortage of time (due to the submission deadline after the summer holidays, which left us with less than four weeks for the evaluation) we decided to apply a questionnaire-based survey. It allowed us to collect information from the reviewers quickly and in a structured way. The evaluation survey was hosted on Google forms.

## 3.1.Method

### 3.1.1. Participants

All 29 reviewers of the LinkedUp project Vici competition submissions were invited to participate in the study. The majority of them have a strong background in technology-enhanced learning with special interest in semantic web and linked open data. All of them have a PhD degree and are active on Technologies for Education and esp. data driven tools. 18 of them, which is more than 60%, responded positively to the invitation and filled out the questionnaire.

### 3.1.2. Procedure

The reviewers received an invitation to participate in the survey through the conference management system EasyChair. They got 5 days to complete the survey. A reminder was sent on day 4 after the announcement of the survey. It is a short period of time but they were informed about that a month before the reviewing process.

### 3.1.3. Measurement Instrument

The questionnaire for data collection consisted of closed-ended items regarding the evaluation criteria. The reviewers were asked to rate on a scale of 1-to-5, the extent to which it was easy / difficult (1 = very difficult; 5 = very easy) for them to apply a particular indicator. The judges were encouraged to make comments in a space added to each of the closed-ended items. The questionnaire was expected to take no longer than 15 minutes.

## 3.2.Summary of findings

We first explored the data to identify potential outliers. These are reviewers who significantly differ in their rating when compared to the group and it could affect the survey's results. There were less outliers compared to the previous VIDI survey. Outliers were detected for several items, namely:

- The application is easy to use (reviewer ID1 - low)
- The application can quickly be learned (ID1 - low)
- The application has an attractive user interface. (ID 1 and ID 13 – low)
- The tool provides a statement on the terms of use (ID1 - low)

31.10.2014© Copyright lies with the respective authors and their institutions.

- Rating the quality of the tool according to the track aims (ID 1, 8, 9, and 17 - extremely low; ID7 and ID10 - extremely high).

It seems that the reviewer with ID1 was consistent in giving a low score on some indicators. 'Rating the quality of the tool according to the track aims' produced some polar results as four of the reviewers scored extremely low and two scored extremely high. This seems to be a trend in the results of the VICI reviewer survey – a relatively large variety in scoring. It can be seen also in the range of the minimum and maximum values (for a half of the items it is 1-5; for the other half, it is 2-5) and the relatively high standard deviation for the majority of the items. Before going further with the analysis we needed to check whether the outliers could affect the study outcomes. We compared the mean of each of the items with outliers to its trimmed mean. No significant difference between the two means on any of the items with outliers was found, meaning there was no need to exclude any data from the analysis.

### 3.2.1. Closed-ended items

A quick eye inspection of the items in Figure 5 indicates that most of the items score higher than 3.

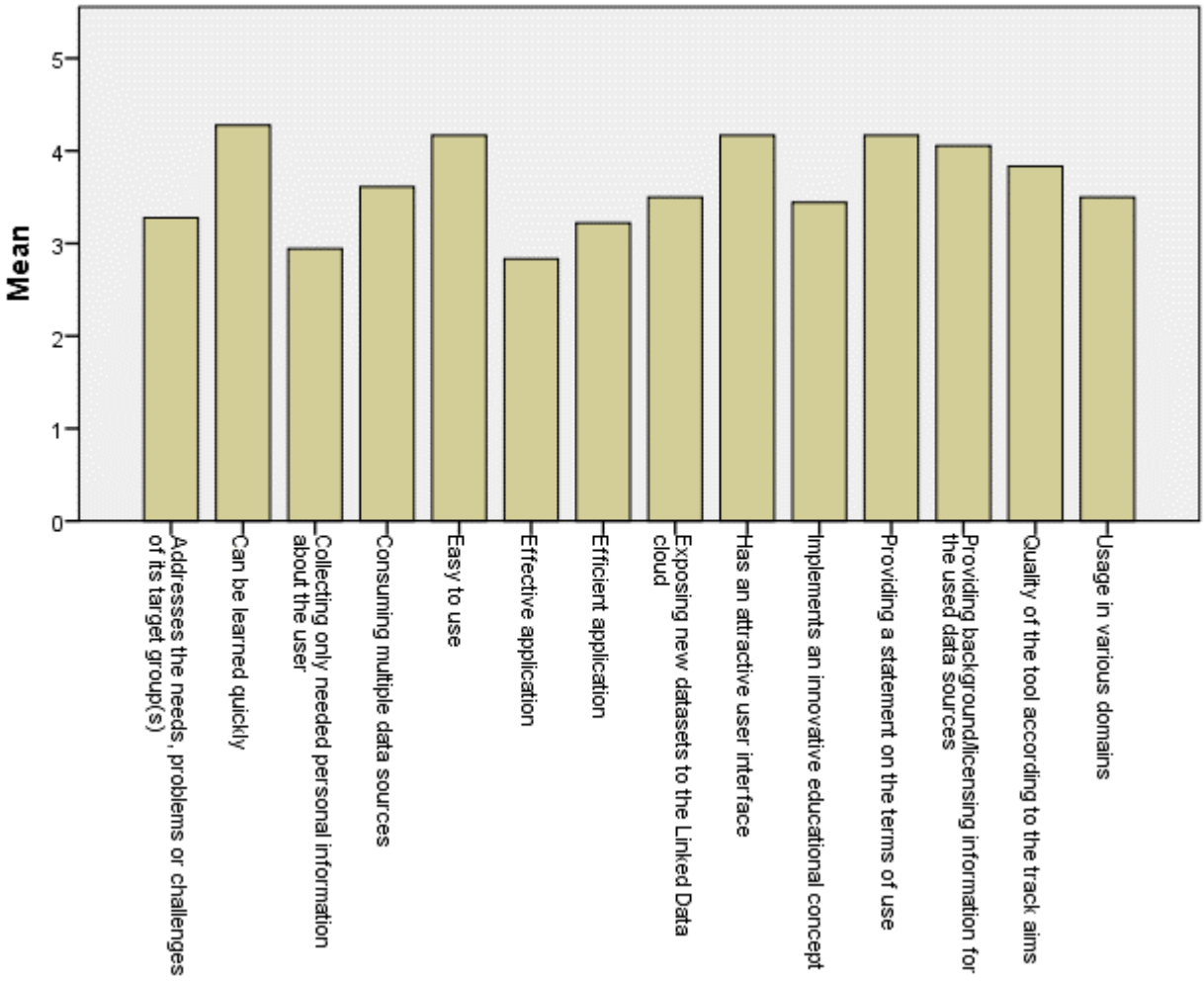



**Figure 5:** Visualisation of the scores on the evaluation framework indicator.

As the Table 3 shows the highest score got the item: 'The application can quickly be learned' (M = 4.28; SD = 0.83), followed by 'The application is easy to use' (M = 4.17; SD = 0.92), 'The tool provides a statement on the terms of use' (M = 4.17; SD = 0.99), 'The application has an attractive user interface' (M = 4.17; SD = 1.1), and 'The tool provides background / licensing information for the used data sources' (M = 4.06; SD = 1.06).

The items that scored relative low are: 'The application is effective. (e.g. leads to significant improvements in learning or teaching)' (M = 2.83; SD = 0.99), 'The tool collects only needed personal information about the user' (M = 2.94; SD = 1.26), 'The application is efficient (e.g. saves time or efforts for learners or teachers' (M = 3.22; SD = 0.8), and 'The application addresses the needs, problems or challenges of its target group(s)' (M = 3.28; SD = 0.75).

Some of the reasons why the items in regard to effectiveness and efficiency of the tool scored relatively low were given in the comments (open-ended items) of the reviewers to these indicators (see Appendix A). The reviewers did not have enough empirical ground to make an informed judgement as the submission did not provide sufficient information in this respect. Some of the reviewers claimed that the applications they reviewed were not about education. The indicators operationalising the criteria Educational Innovation, including those referring to effectiveness and efficiency, were meant to be about *perceived* effectiveness and efficiency and the potential of the tool to improve learning and save time and efforts. We had the intention to provide some more explanation in the instruction to those items but the technical affordance of EasyChair did not allow for that. At least predicate 'perceived' should have been added.

According to the reviewers' comments, it is not easy to find out whether the tool collects implicitly personal information (apart from data entered by the users themselves). There was also inclarity what was meant by 'domain' ('The application can be used in various domains'). "Educational domain or knowledge domain, could a similar approach work in another knowledge/education domain without alteration"? In addition, the authors not always provided explicit information about the target group of their application. It seems from the comments of the reviewers that the usability indicators were the easiest to apply.

Table 3: Descriptive statistics in regard to rating of the evaluation framework indicators

| Items | N | Mean | Std. Deviation |
|---|---|---|---|
| Can be learned quickly | 18 | 4,28 | ,826 |
| Easy to use | 18 | 4,17 | ,924 |
| Providing a statement on the terms of use | 18 | 4,17 | ,985 |
| Has an attractive user interface | 18 | 4,17 | 1,098 |
| Providing background/licensing information for the used data sources | 18 | 4,06 | 1,056 |
| Quality of the tool according to the track aims | 18 | 3,83 | ,707 |
| Consuming multiple data sources | 18 | 3,61 | 1,335 |
| Usage in various domains | 18 | 3,50 | ,786 |

31.10.2014© Copyright lies with the respective authors and their institutions.

| | | | |
|---|---|---|---|
| Exposing new datasets to the Linked Data cloud | 18 | 3,50 | 1,505 |
| Implements an innovative educational concept | 18 | 3,44 | 1,247 |
| Addresses the needs, problems or challenges of its target group(s) | 18 | 3,28 | ,752 |
| Efficient application | 18 | 3,22 | ,808 |
| Collecting only needed personal information about the user | 18 | 2,94 | 1,259 |
| Effective application | 18 | 2,83 | ,985 |
| Valid N (listwise) | 18 | | |

# 4. CONCLUSIONS

This deliverable provides the final evaluation of the LinkedUp EF in the Vici state. The EF that has been developed over the past two years has been reviewed for its effectiveness and accuracy by using five different quantitative and qualitative methods such as:

1. An expert validation over the Group Concept Mapping method (D2.1)
2. A literature review study (D2.1)
3. A pilot study of the EF within the LinkedUp tutorial at LAK13 conference (D2.2.1)
4. Expert interviews and semantic analysis of the interview text (D2.3.1)
5. Survey studies at the Vidi competition (D2.3.2)

We are very satisfied with the final state of the EF in Vici, as it could successfully identify the shortlist candidates that have been invited to the ISWC2014 conference as well as the 4 winning submissions according to their final ranking.

The experience with the weighting has been a very fruitful one. Weighting criteria can provide different perspective on strong and weak aspects of a set of submissions. Thus, weighting can be an approach to highlight specific objectives of a data competition. However, it also opens the floor for various discussions such as:

- Which factor for weighting should be applied?
- Why you weight factor A vs. factor B?
- Weighting of factors should be transparent to the data competition participants and the judges.

In general we suggest having a balanced set of evaluation criteria that are all equally important. Nevertheless, having a set of weightings at hand can be supportive to explore the weak and strong points of your submissions.

The reviewer survey was overall very positive and identified only very minor improvements for the EF, namely:

- A better formulation of the indicators educational effectiveness and educational efficiency. It should read ‘perceived effectiveness’ and ‘perceived efficiency’. Some more explanation to the items is needed.
- Some more clarification is also needed to define what is meant by ‘domains’ in the item ‘The application can be used in various domains’.
- The item about collecting personal information should explicitly refer to only data entered by the user in the form of the tool.

31.10.2014© Copyright lies with the respective authors and their institutions.

# Appendix A. Responses of judges to open questions in Vici evaluation survey

| **Educational Innovation** |
|---|
| □ It's just a tool, not a pedagogical concept ... therefore I am not able how efficient or effective the tool might be |
| □ To really grasp the innovation potential, one should be able to play with the app for some time... e.g., with your own data... |
| □ Effectiveness and efficiency should ideally be evaluated with a user study. Best I could provide was an educated guess. |
| □ Efficiency and effectiveness was difficult to judge because some tools which I evaluated didn't really show the learning process. Example: one tools support the download of datasets for education, but it remains unclear how they can be used for education. |
| □ Only if it is implied to judge perceived efficiency and efficacy, papers do not provide experimental data. |
| □ Evaluating efficiency and effectiveness requires to see the application on use in a real context. Just trying it a bit around is not sufficient to get an idea. It would help it authors are asked to provide information evaluations carried out on their own experiments regarding that criteria. |
| □ Without any clear description about an educational concept which makes purpose of the new application it is different to judge about it. |
| □ It is not straightforward to deduce whether the application really leads to significant improvements in learning or teaching without testing by ourselves the application inside the classroom. Sometimes the application presents some kind of social evaluation that can be useful in this matter, but this is not always the rule. |
| □ Lack of a precise definition. Also, this is just very hard to evaluate in isolation, i.e., without actually using it in an educational context. It would be very unreasonable to give a score to this just by looking at a demo, knowing that other disciplines spend months to evaluate this. |
| □ It seems hard to assess the effectiveness and efficiency of the application if we only get a chance to test the application once or twice, since these metrics might be more appropriately be measured in association with specific learning tasks. |
| □ Probably the most difficult and the most subjective indicator. It requires good knowledge of the method; furthermore, the indicator is multifaceted: reviewers will take their overall judgement of the innovativeness of the method regardless of the particular implementation into account. Overall, I'd say that innovation is a too complex concept to operationalise it in just two numbers. |

| **Usability** |
| --- |
| □ An (unwritten?) rule for web interfaces is that one should be able to use them without prior training. I think that the time I spent using the interface was enough to judge its usability and learnability. Attractiveness is somewhat subjective, but still there are (written and unwritten) conventions about this. |
| □ The criteria is ok. More criteria from the technology acceptance model makes things more difficult to answer. |
| □ I don't see much difficulty in this criterion regarding the application. But I feel it is very subjective and might differ from one evaluator to another. |
| □ “attractive user interface” is highly subjective and actually seems a non-issue. |
| □ Unlike the previous set of questions, this section of usability testing is more easier. |
| □ Usability is a well defined concept, therefore the indicators were easy to apply. |
| **Audience** |
| □ It depends on how the creator sells the tool, but in most cases it is relatively easy to judge whether a tool fits the given use-case or target group or not. To what extent an application can be used in various domains can be a bit harder. |
| □ The first criteria can only be judged when the authors provide enough information about the target user and their problems.<br>□ What do you mean by domains? Learning domains, topics? This question is ambitious. |
| □ It is not clear what is meant - could a similar approach work in another knowledge domain? - Should the same application be fit to be used in another domain without alteration? - Educational domain or knowledge domain? |
| □ The first questions has to be told by the authors. You can only check if that fits or not. The second one might be easy to measure if you can identify other domains of interest. |
| □ Without any clear description about the target group which makes purpose of the new application it is different to judge about it. |
| □ To assess this, the intended audience should be described. |
| □ This is challenging to answer as the questions lead to assumptions about the target group and the domains. |
| □ Quite easy, but hard to say when the knowledge of the audience and/or the domain by the reviewer is not as good. |

31.10.2014© Copyright lies with the respective authors and their institutions.

| **Legal aspects** |
| --- |
| □ The background/licensing information can be easily checked: either it is available or not. Whether an application (implicitly) collects personal information about the user is not always easy to see (c.f. Google Analytics). However, I assume that an application that does not (visibly) exploit personal information also doesn't collect it. |
| □ Collection of personal information: it is impossible to judge whether more data is gathered than the data entered by the user. I would say "entered by the user in the forms of the tool".<br>□ Terms of use: Where should it be mentioned to tick "yes"? In the tool or the descriptions provides besides the tools (e.g., paper, slides) ? |
| □ We do not receive source codes, so we do not know what information is actually collected. |
| □ Question 1 and 3 require the applications informs about it, and you can check. For question 2, it is difficult to know just by using the tool. |
| □ What constitutes essential personal information is not very clear and may depend on the needs of the application. |
| □ If not described in the paper, the application does not normally make it explicit what if any personal data is collected. |