Linked **O**pen **D**ata
at the
Vienna University of Technology
A case study about research data

Lukas Baronyai[a], *Author*

Kevin Haller[b], *Co-Author*

Stefan Gamerith[c], *Co-Author*

{e1326526[a],kevin.haller[b],e0925081[c]}@student.tuwien.ac.at

March 30, 2016

Abstract

A report about Linked Open Data, how they help to improve processes in a university context and how they can successfully be applied a the Vienna University of Technology. Especially this paper aims to understand best practices regarding the applicability of Linked Open Data in university settings, major benefits, challenges and barriers of an according implementation. Further the most important stakeholders will be identified and major use cases explored in a case study. The report ends with a proposal of a generic technical architecture.

The work was done in three different shapes to match the three major stakeholders: researcher, students and administration staffs. This work focus on researcher, the other two papers are referenced in this report.

Keywords. Linked Open Data, University, Case Study

Contents

List of Figures

1 Introduction

While the pressure on governments and public organizations to release *Open Data (OD)* has significantly grown with the spread of information systems there has been also a need for *linking* these data from various sources to understand the information in a contextual sense.

As OD includes non-confidential data any restrictions in distribution are prohibited and information is funded only by public money [14]. The application domain for OD providers is not restricted by its nature in any way including from traffic, weather and public transport to name a few. However, exclusively exposing public assets is not enough, in addition establishing a feedback loop facilitates an ongoing adaption to the stakeholders concerns.

The World Wide Web has proven great success in spreading knowledge of various data sources all over the world. The building blocks of the Web are documents and links building a shared, global and connected information source. This can be seen as the key success factor in its nearly unconstrained growth [13]. *Linked Data (LD)* has adopted these principles of publishing and connecting data realized as machine-readable, structured data connected to various data sets which in turn are linked to different data sets.

Berners-Lee [1] developed a five star deployment scheme classifying OD. The scheme ranges from a one star rating covering proprietary data formats (e.g. Portable Document Format (PDF)) to a five rating including machine-readable formats using open standards with links to other data sets.

Although LOD offers universities new opportunities for providing unprecedented insight into its core activities and ease application development, a major **problem** is that **LOD has not been widely adopted by universities yet**. Even though there are a few examples [5] of publishing university related data sets as LOD there has been little knowledge of its usage for publishing university related information.

To answer the research questions stated in the next sub-section we conducted interviews at the Vienna University of Technology. In particular **the context of this work are research data**, though there exist papers investigating similar research questions but with different context: Gamerith [8] (administration data) and Haller [10](data concerning students).

The remainder of this section states the addressed research questions, describes the contributions of this work and gives an overview of the structure of this paper.

1.1 Research Questions

The fundamental research question we investigate is:

> *How can Linked Open Data help to improve processes in a university context and how can it be successfully applied?*

More concrete, this paper concentrates on the following four research questions:

RQ1: What are best practices regarding the applicability of Linked Open Data in university settings? At the time of writing this paper there are no established best practices for the use of LOD due to its little adoption in university settings. For this very reason it is crucial to identify strengths and limitations from previous experiences of using LOD as the foundation of information systems [5].

RQ2: What are major benefits and barriers for each stakeholder and what are useful use cases? We identified three different stakeholders *Students*, *Researchers* and *Administration staff*. However, focus of this work are researchers. Since the success of any new technology highly depends on their acceptance concerns of each target group needs examination. Furthermore, use cases are important to showcase profits and shortcomings.

RQ3: What are major challenges for the implementation of a Linked Open Data solution? As the implementation of a LOD solution is a time consuming task the knowledge of probable challenges from a technical perspective as well as from a management perspective is key to a successful adoption.

RQ4: How would a prototypical implementation of a publication framework based on Linked Open Data look like? Among the definition of building blocks for a publication framework of university related assets a high level picture of the overall architecture needs to be drawn to give implementers and LOD experts a common understanding of the system. Next, defining sample ontologies representing selected assets of the university domain draws concrete examples of how LOD data might look like.

1.2 Methodology

Finding an answer for the research questions above has led to the following three methodologies:

A coordinated set of semi-structured interviews To answer research questions RQ2-RQ4, we interviewed a selected set of *researcher*. Semi-structured interviews were selected as the means for data collection because they are well suited for exploring the impressions and interests of the interviewees like in a discussion while still following a defined structure.

Literature Review Undertaking a literature review to justify scientific contributions and making sound conclusions is an established practice in any scientific community. Since our scientific work targets to the Semantic Web community we made some pre-assumptions about a basic understanding of the technologies and concepts used in that respective area. More specifically, we assume a basic understanding of the concept of an ontology and some pre-knowledge about ontology description languages.

Conceptual System Design The development of applications based on LOD requires a methodology facilitating a common understanding of the overall system infrastructure. Therefore we designed a conceptual model of a prototypical implementation of a publication framework based on LOD.

1.3 Contributions

The work in this paper mainly contributes to different aspects which need to be considered when designing and implementing an application driven by LOD. More precisely, our contributions can be categorized into the following four areas:

1. Identifying best practices for Linked Open Data in university settings. Information systems needs to cope with vast amount of data nowadays growing the need to efficiently handle Linked Data as well. We gave a brief overview of existing research work in that area, in particular, we compared the benefits and shortcomings in existing LOD solutions.

2. Finding benefits/barriers including additional use cases for stakeholders. As with every software project the very first phase of the Software Development Lifecycle (SDLC) is *evaluating of requirements*. As LOD driven software has additional requirements to the accessibility of data and their organization we investigated if the overhead compared to an established technology (e.g. a database based solution) is feasible. A set of selected *researcher* are interviewed at the Vienna University of Technology. Additionally we proposed several use cases emphasizing their point of view.

3. Discovering possible obstacles for implementers of a Linked Open Data solution. As the application domain for a Linked Data is limited to the university context our work includes LOD driven applications resulting from our conducted interviews.

4. Sketching a prototypical implementation of a publication framework driven by Linked Open Data. The proposed publication framework covers the whole process of data provision, requirement analysis and application development designed for but not limited to university related assets.

1.4 Structure of this Paper

This paper is structured as follows:

- Section 2 provides a summary of existing efforts regarding Linked Data principles in university contexts to cover RQ1.
- Section 3 discusses the used methodology (subsection 3.1) and results (subsection 3.2) concerning RQ2 and RQ3 done by a case study with researchers.
- Section 4 shows a prototypical implementation for a publication framework of educational assets in subsection 4.1, while subsection 4.2 proposes a technical architecture for the application of Linked Open Data principles in educational environments.
- In the last Section 5 conclusions and future investigations are made.

Whereas section 4 is particularly intended for readers familiar with software architectures, the remaining sections do not necessarily require a deeper understanding. A general understanding of the Linked Open Data principles is necessaryand a general understanding of Semantic Web technologies for the whole paper is recommended , though.

2 Related work (RQ1)

This section gives a brief overview of existing work relating to Linked Open Data (LOD) in university contexts. In particular we investigate experiences and established practices in that area. Fortunately there has been plenty of work done especially around Linked Universities (section 2.1), an alliance of european universities. Next we introduce the LinkedUp project (section 2.2), which tries to combine several of those facilitating a uniform view of these concepts. Then we shortly describe the Austrian Open Data project (section 2.3), managed by the austrian government, and at last the Linked Data for Libraries (LD4L) project (section 2.4), which aims to interlink various university libraries.

2.1 Linked Universities

One of the most important university projects in the world of LD are the LinkedUniversities. They are "*an alliance of european universities engaged into exposing their public data as linked data*"d'Aquin et al. [6], providing help and knowledge for other universities who wants to implement LD-Systems in their infrastructure. Addressing the problem of connecting data and developing new sites by inexperienced universities, the alliance provide information so they don't have to be re-learned. For this purpose the LinkedUniversities offering a portal as collaborative space with common vocabularies and practices for reusing, describing and sharing.

Their goals are: [1]

- "Identify, support and develop common linked data vocabularies, usable accross universities for common concepts such as courses, qualifications, educational material, etc."
- "Describe reusable recipes, and share reusable tools, for exposing linked data in universities"
- "Support, through experience sharing and reuse, initiatives towards exposing university data as linked data"

Country	**University**	**Responsible**
UK	The Open University	Mathieu d'Aquin, Stefan Dietze
Germany	University of Münster	Carsten Kessler
Finland	Aalto University	Tomi Kauppinen
UK	University of Southampton	Christopher Gutteridge
Sweden	Royal Institute of Technology (KTH) / MetaSolutions AB	Hannes Ebner
Greece	Aristotle University of Thessaloniki	Charalampos Bratsas
Turkey	Ege University	Oguz Dikenelli
Czechia	Charles University	Jakub KlÃmek
Spain	Universitat Pompeu Fabra	Jorge Pantoja

Figure 1: Members of the Linked Universities [2]

2.1.1 Data sets and Endpoints of the Linked Universities

All universities participating in Linked Universities provide endpoints for all different kinds of Linked Data.

The Linked Universities project provides a list of participating universities together with endpoints offering Linked Open Data. Unfortunately up to the time of writing this paper not all endpoints listed at the member page[3] are online.

[1] d'Aquin et al. [6]

[3] `http://linkeduniversities.org/lu/index.php/datasets-and-endpoints/index.html`

The table 2.1.1 shows selected universities together with online SPARQL endpoints and data catalogs:.

University	**Data Set**	**SPARQL Endpoint**
Open University	`http://data.open.ac.uk`	`http://data.open.ac.uk/query`
University of Southampton	`http://data.southampton.ac.uk`	`http://sparql.data.southampton.ac.uk`
Aalto University	`http://data.aalto.fi`	`http://data.aalto.fi/endpoint`
University of Muenster	`http://data.uni-muenster.de`	`http://data.uni-muenster.de/php/sparql`
University of Pompeu Fabra	`http://data.upf.edu/en/linked_data`	`http://data.upf.edu/en/sparql`

Figure 2: SPARQL endpoints and data catalogs of selected members of Linked Universities

2.1.2 Example: The Open University and the LUCERO Project

LUCERO (Linking University Content for Education and Research Online) was a project from the Open University, funded for 1 year by the JISC Information Environment 2011 Programme under the call Deposit of research outputs and Exposing digital content for education and research. Aim of the project was to " *scope, prototype, pilot and evaluate reusable, cost-effective solutions relying on linked data for exposing and connecting educational and research content*" [4]. The projects connected with other organizations through LinkedUniversities.org to gather common issues and practices. They outcome was the first university linked data platform, `http://data.open.ac.uk/`, with much impact on The Open University and the education community.

[4] d'Aquin et al. [3]

2.2 LinkedUp project - Linking Web Data for Education

The LinkedUp project[5] is a project funded by the European Commission aimed at pushing forward the exploitation and adoption of Linked Open Data in educational organizations and institutions. This project started on the 1st November 2012 and lasted about 2 years.

More specifically, the projects main objectives are:

- `Open Web Data Success Stories`
  The identification of innovative success stories using tools and data sets in the education sector helps at spreading knowledge and awareness of Linked Data principles.
- `Web Data Curation`
  The collection of relevant educational data sets facilitate the integration of third party data and applications.
- `Evaluation Framework for Open Web Data Applications`
  Evaluation of successful large scale applications driven by Linked Open Data promotes the adoption of Linked Open Data especially in educational contexts.
- `Technology Transfer in the Education Sector`
  The promotion of Linked Open Data technologies is an explicit goal of the LinkedUp project.

The outcome of the LinkedUp project to investigate the goals listed above are several so-called *"deliverables"*[6].

Below, there is a list of the most relevant deliverables:

2.2.1 State of the art and data assessment

In this report [11] a comprehensive study on the developments in the field of Linked Data and related fields of educational data mining and learning analytics is given. The report resulted from a collaboration between the Leibniz University of Hannover, the Open Knowledge Foundation, the Open Universiteit Nederland, the Open University, Elsevier and Exact Learning Solutions. The authors first briefly describe fundamental technologies and concepts, namely Linked Data, educational data mining and learning analytics used throughout the paper. Several candidates representing each of the two basic data mining principles (*data harvesting* and *distributed search*) for heterogeneous data sets are presented. The field of learning analytics focuses on techniques targeted at analyzing and understanding educational learning processes and environments. A complete discussion on this topic though, was out of scope since this is a relatively new research area. Next, challenges and barriers concerning the heterogeneity of open educational resources are investigated. Techniques trying to solve isolated data sets include *schema mapping* and *classification and clustering*. Finally, data sets from the LinkedUp repository[7] and possible legal obstacles and solutions to these are assessed.

2.2.2 The LinkedUp Challenge(s)

The LinkedUp Challenge[8] was organized as three separate challenges *Veni*, *Vidi* and *Vici* ending on October 2013, April 2014 and November 2014 respectively. While the goal of the first competition (Veni) was developing a *"prototype or demo that uses linked and/or open data for educational purposes"*, the Vidi challenge was targeted at finding *"innovative and robust prototypes and demos for tools, which still may contain bugs, as long as it has a stable set of features and there is some proof that us can be deployed on a realistic scale"*. The rationale of the latter, yet more advanced challenge was to build *"advanced prototypes and tools that should be mature and stable; it should be used or have been used by a fair amount of users on a realistic scale"* [12]. The combined goal though was to promote creativity and innovation in ways to mash-up and link educational resources and services. In addition, companies, universities and government agencies were encouraged to share and (cross-)link to educational and non-educational assets.

[5] `http://linkedup-project.eu/`
[6] `http://linkedup-project.eu/resources/deliverables/`
[7] `http://data.linkededucation.org/linkedup/catalog/browse/`
[8] `http://linkedup-challenge.org/`

The winner for Veni was `Polimedia`[9], an application facilitating large-scale, cross-media analysis of the coverage of political events. `TuvaLabs`[10] won the first place of the Vidi competition aimed at transforming open data into opportunities for meaningful teaching and learning, equipping teachers and students with high quality, consolidated data sets. Finally, `Flax` (Flexible Language Acquisition)[11], the winner of the Vici competition, was designed to automate the production and delivery of interactive digital language collections and targeted to non-expert users (e.g. language teachers, language learners, subject specialists, instructional design and e-learning support teams).

2.2.3 Evaluation Framework

Drachsler et al. [7] proposed an evaluation framework throughout the three web open educational data competitions Veni, Vidi and Vici, described in detail in the last paragraph. The aim of the evaluation framework was to evaluate the usefulness, usability, acceptance and appropriateness of the contributions to the LinkedUp challenges. Therefore experts had to define and/or refine assessment criteria and indicators for measuring the quality of the framework. Practically the frameworks usefulness and ease of use was tested through a questionnaire and interviews. A detailed questionnaire for each of the three challenges as well as the final version of the evaluation framework are accessible at [7].

[9] `http://www.polimedia.nl/`
[10] `https://tuvalabs.com/`
[11] `http://flax.nzdl.org/`

2.3 Austrian Open Data

In Austria datasets of various governmental units from different areas of life are exposed on the open data portal `data.gv.at`. Whereas this data portal is restricted to governmental units[12], `opendataportal.at` provides a platform for anyone who wants to share data in an open way. In both cases the category *'Bildung und Forschung'* (German for education and research) is especially of interest for (Linked) Open Data initiatives at Austrian universities. Overall, the published data sets are mainly localizing points of interest (libraries, museums and universities) and apart of the Vienna University of Economics more specific information like provided courses and curricula has not been published yet about universities. In fact, the Vienna University of Economics is already active and publishes mainly course information and library collections in a machine-readable and open way on their open data portal `data.wu.ac.at` and on `opendataportal.at`. All in all the amount of data concerning education and research on the mentioned Austrian data portals is negligibly at the moment of writing.

Meanwhile data is exposed in machine-readable and open formats, there has been little effort taken to provide it as Linked Open Data. However, Christian Weiss has designed a system[13] in connection with his master thesis[14] to provide a subset of the Open Government Data provided by Vienna (`data.wien.gv.at`) as Linked Open Data. In the same year 2013 the project LODPilot[15] was initiated to create an infrastructure to provide the basic datasets published on `data.gv.at` and `data.wien.gv.at` as Linked Open Data.

Beside the open data portals, `Wegweiser`[16] is an early attempt of a set of Austrian universities to share information of public interest like course information, curricula and building plans in an open way. The information on Wegweiser is served only in a human readable format and therefore not easily processable by computers so that applications based on it can difficultly evolve. Furthermore, the service is not maintained anymore and outdated. `Open Street Maps`[17] is a global service that make spatial information of Austria (a.o.), which has been collected by an open community, available in an open way. The University of Leipzig has built an infrastructure to provide this collected information as Linked Open Data with the project named `LinkedGeoData`[18]. `DBPedia`[19] is another project of the mentioned university that extracts knowledge from Wikipedia and transforms it into Linked Open Data. For example, `DBPedia` contains a machine-processable description of the Vienna University of Technology[20]. `Wikidata`[21] is a project similar to `DBPedia` and also provides knowledge about Austria in a machine-processable format.

[12] Member of 'Cooperation OGD Austria'
[13] `http://cweiss.net/lod/` last accessed on 15.03.2016
[14] C. Weiss, 'Transferring Open Government Data into the global Linked Open Data Cloud', 2013
[15] `http://lodpilot.at/`
[16] `http://www.wegweiser.ac.at/`
[17] `https://www.openstreetmap.org/`
[18] `http://linkedgeodata.org/About`
[19] `http://wiki.dbpedia.org/`
[20] `http://dbpedia.org/page/Vienna_University_of_Technology`
[21] `https://www.wikidata.org/`

2.4 Linked Data for Libraries (LD4L)

The project "Linked Data for Libraries" (LD4L) is a collaboration of the Cornell University Library[22], the Harvard Library Innovation Lab[23], and the Stanford University Libraries[24], and is funded by a nearly $1 million two-year grant from the Andrew W. Mellon Foundation. "The project aim to create a Scholarly Resource Semantic Information Store (SRSIS) model that works both within individual institutions and through a coordinated, extensible network of Linked Open Data to capture the intellectual value that librarians and other domain experts and scholars add to information resources when they describe, annotate, organize, select, and use those resources, together with the social value evident from patterns of usage." [25].

The project build on existing previous work including:

- `VIVO`[26]
  a semantic web system and data interchange standard for describing researchers and scholars in the context of their research and scholarship
- `Project Hydra`[27]
  a software framework and community focused on creating digital repositories and collections, together with user workflows
- `BIBFRAME`[28]
  a project of the Library of Congress and Zepheira to create a linked data standard with which libraries can describe and exchange bibliographic information about scholarly resources

The result was an open source extensible LD4L ontology compatible with VIVO ontology, BIBFRAME, and other existing library LOD efforts, an open source LD4L semantic editing, display, and discovery system, implemented into the existing existing online library systems of the universities and a Project Hydra compatible interface to LD4L, using ActiveTriples to support Blacklight search across multiple LD4L instances [29]. The code sources are available at github[30].

Important: this is *not* a Linked *Open* Data project but a Linked Data project. The results are open source code projects, not endpoints like from the LinkedUniversities.

[22] `http://www.library.cornell.edu/`
[23] `http://librarylab.law.harvard.edu/`
[24] `http://library.stanford.edu/`
[25] Team [18]
[26] `http://vivoweb.org/`
[27] `http://projecthydra.org/`
[28] `http://www.loc.gov/bibframe/`
[29] Team [19]
[30] `https://github.com/ld4l`

3 Benefits and challenges of using LOD at TUWien (RQ2 & RQ3)

As mentioned in the introduction, we identified *students*, *researchers* and *administration employees* as import stakeholders at a university. This chapter describes the methodology and results of the conducted case study involing researcher. The essential question of the case study was RQ2 ("*What are major benefits and barriers for each stakeholder and what are useful use cases?*") and RQ3 ("*What are major challenges for the implementation of a Linked Open Data solution?*") A case study concerning students was conducted by Kevin Haller [10] and another one concerning administration employees by Stefan Gamerith [8].

The first section of this chapter describes the applied methodology (see section 3.1). The second section evaluates and analyzes the results of the case study, investigating the possible benefits, barriers and disadvantages of some use cases and a general view at the interviewees (see section 3.2).

3.1 Methodology

In this study the data were acquired by a coordinated set of semi-structured interviews. As mentioned the stakeholders were classified into three groups (*administrative staff*, *students* and *researchers*) and therefore three different versions of the questionnaire but with joint parts for statistically evaluation were worked out. For each version exists an according paper, in this work only the category "researcher" will be described.

3.1.1 Design of questionnaire

The main purpose of the interviews were the collecting of the stakeholders thoughts, needs and knowledge, so the method of of a *semi-structured* interview was chosen. A *fully structured interview* would not be adequate because of its strict character allowing only predefined answers and a *unstructured interview* would be top difficult to analyze. In contrast a *semi-structured* interview let researchers have their freedom to decide, which question to ask next in dependence on the development of the conversation - this allows the freedom to express and evaluate new ideas without being being constrained by a fixed order of planned questions.

After choosing the method the questionnaire was defined to use as guidline during the interviews. To allow a general, generic shared analyze of the interviewees the team decided to mix open questions from the semi-structured model with closed questions with fixed, predefined questions. The result had four parts:

1. General question about the interviewee for classification, about his/her work
2. General question about the interviewee's knowledge in general technical and LOD context.
3. Explanation of LD, followed by a specific set of questions targeting the thoughts and opinions of the interviewee about presented use cases and example application. Motivation of this part is to introduce the interviewee to LOD if it is an unknown topic and let him/her start to think about LOD to prepare the next part
4. Wide open Questions to explore and find use cases and existing data sources for LOD application at the university.

Part 1 and 2 are used for a general statistical evaluation (see section 3.2.1). In Part 3 were two concrete use cases proposed: publishing (meta-)data from the library similar to the LD4D project (see 2.4) and publishing the publication data of the university from the publication database, according to the similar "Open Research Online" [31] endpoint from the Open University.

The list above represents indeed the order of the parts addressed during the interview, but due the semi-structured design the questionnaire allowed variance in part 3 and 4. E.g. could the length and depth of the LOD introduction be adopted to the level of expertise of the interview partner.

[31] `http://data.open.ac.uk/page/context/oro`

3.1.2 Data Validity and Quality

To ensure both a continuous conversation flow and a high quality recording of the spoken words, the interviews were conducted in teams of two researchers. One researcher, who was the moderator and one, who had the task to collect data by writing valuable contributions onto a logging sheet. This approach was chosen, because the attempt to moderate a interview and to simultaneously log it may lead to the lose of valuable contributions or a frequently interrupted conversation. Further to avoid too detailed notes all interviews were additionally audio recorded so no information was lost and the second researcher could easier follow the interview and make more valuable notes. As result the data are available as interview notes and audio records and further the evaluation could be done without replay the whole interview or transcribe it.

3.1.3 Description of interviewed people

As mentioned the interviewees of this study were chosen according to the category "*Research*", so the interview partner were active researcher in various fields. Altogether four interviews were done. Because of the technical character of LOD the chosen people are all technically experienced so they are able to imagine use cases at the university - for a detailed describtion of the interviewees background see section 3.2.1.

3.2 Results

After giving a detailed description of the used methodology in the previous subsection this subsection contains the interviews outcome. First the general questions about the interviewees background are evaluated and interpreted. Because this part was included in all three questionnaires, this part also includes data from the other reports for comparing the data, but only the data from the researcher interviews are interpreted. In the next parts, the use case of the library data and publication data are evaluated and their potential benefits, barriers and disadvantages analyzed. In the last part, the ideas and mentions from the open questions are evaluated.

3.2.1 General statistically evaluation of the interviewees background

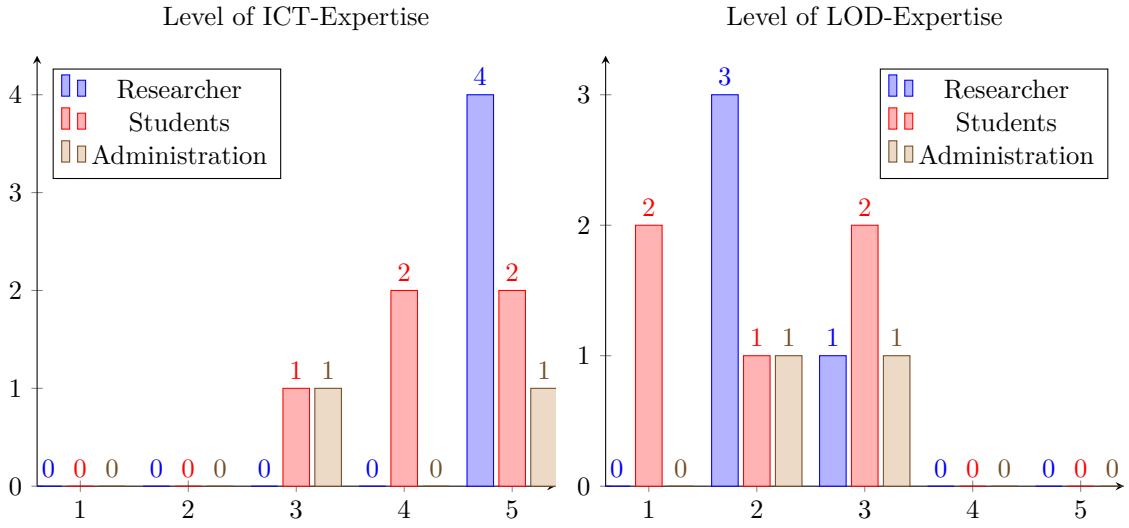


Figure 3: Level of Experience

The first part of the interview aimed to categorize the interview partners and to understand their background. They had to estimate their level of expertise in the field of Information & Communication Technologies and in the field of Linked Open Data in a formal way and then describe their daily work and responsibilities at the university. This part was identical in all three interview series (administration, students and researcher), so the result can be compared. The levels of expertise can be seen in figure 3.2.1 and the according rating scales in table 3.2.1. For a detailed description of the administration and student interviews see the corresponding reports Gamerith [8] (administration data) and Haller [10](data concerning students).

Tough all four research-interviewees had a high expertise in ICTs, their expertise in LOD are mainly unincisive, although everyone already knew the concept of LOD. This can easily explained by their research fields (see table 3.2.1): everyone works in a technical context.

This could be seen as advantage as well as disadvantage. On the one hand the interviewed persons needed lesser effort to understand the benefits of LOD and could easier imagine further use-cases and possible data sets for LOD. But on the other hand, their perspectives were in some way restricted by their profession - a more divergent angle of view may be interesting for further studies to explore more “non-technical” use-cases. But for an initial study like this one it may be enough.

3.2.2 General results

In this section we describe general results which not applied to one of the described use cases. This mainly involved the common attitudes, thoughts and doubts for LOD of the interviewed persons.

3.2.2.1 General benefits

All interview partners strongly welcomed the idea of LOD and LOD application at the university. Everyone of them could imagine a LOD project despite of possible challenges or problems, they rated clearly the benefits over them. The opportunity to develop new ideas based on an access to

Value	ICT	LOD
1	Fundamental	I never heard of Linked Open Data.
2	Novice	I heard of Linked Open Data, but never used it.
3	Intermediate	I used Linked Open Data in a not intense way. E.g. as part of a workshop or home project.
4	Advanced	I used Linked Open Data in a practical project.
5	Expert	I used Linked Open Data in several practical projects and consider myself an expert in Linked Open Data.

Figure 4: Rating scales for the level of expertise in the field of ICT and LOD.

ID	Assignment	Date of interview
A	research & teaching in HCI	15.12.2015
B	research in information retrieval	04.12.2015
C	research transfer	23.11.2015
D	research in audio and video analysis	24.11.2016

Figure 5: Background information of the interviewees and the corresponding interviews.

open data was very common liked across all questions. One important point was to divide and free the data from their original purpose and application to create context-independent ideas. It was suggested to publish as many data as possible as LOD and then let (master) students in courses and projects make application based on this data without strict restrictions and goals.

3.2.2.2 General challenges

One of the most frequently mentioned challenges was costs in terms of money and work, followed by data quality and administration.

Costs in term of money may be a small challenge if a LOD project stays at a local university context but may raise proportional if the project move to a higher level like Austria- or Europe-wide. In terms of work the challenges are mainly bureaucracy to manage the project and in the maintenance of a living system. This has also a strong connection to the challenge of data quality, a crucial point of real time LOD. If they get outdated, they get useless and therefore a huge investigation of time and work has be done in this area to ensure a high quality of the data.

Another mentioned challenge was the communication between technically experienced stakeholders and lesser experienced stakeholders which hold the domain knowledge. Therefore a proper ontology must be found to fit both needs and simultaneously describes the data sufficiently. Also this ontology has to be maintained to always fit the needs and data appropriately.

Further some of the interviewees pointed the way of making the LOD available out. To stimulate and motivate other people to work with the data interfaces, they have to know that they exists. Furthermore it won't be successful to just put the data to an endpoint, instead the students, other stakeholder and interested people need to actively be informed about the benefits and availability (e.g. in a way described in the previous section 3.2.2.1.

3.2.3 Library data

Use case This question aimed for a use case similar to the project Linked Data Libraries (described in Section 2.4). The proposed scenarios was to publish the data or meta-data of the university library (and all of its specialized libraries) as LOD and provide an application to access the data. A further option of this scenario would be an interlinking to other LOD data sets, e.g. from the publisher "Springer" or to other universities like LD4L do it.

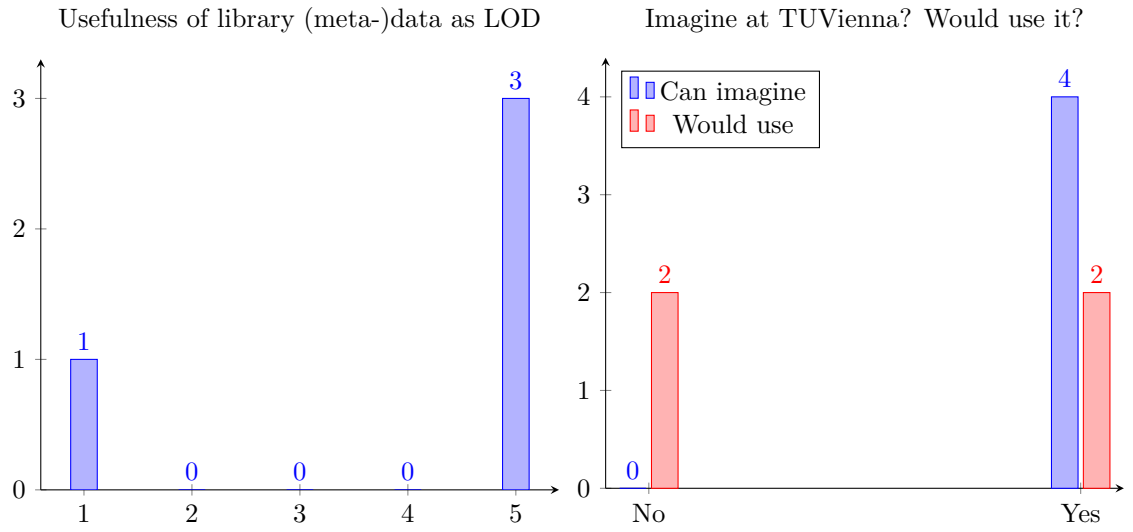


Figure 6: Usefulness of library (meta-)data as LOD

Statistically evaluation It can be seen in Figure 3.2.3 that the interviewed persons strongly agreed to the scenario (found it "extremely useful") and could imagine a similar project at the TU Vienna. Only one interviewee found it difficult to see advantages and therefore argued that he wouldn't use it. Another one found it indeed useful in a general context but not for his own work.

Needs, potential benefits As stated, all of the interviewed persons could imagine a similar project. One of the main reasons of the strong acceptance was the current interface of the library website [32], which only allows a search with only a few, specified parameters (see Figure 3.2.3). Also the physical search in the library itself was claimed due to a lack of orientation and knowledge about the position of e.g. a searched book. Both point of criticism are expected to vanish by an open access to the data and appropriate applications, which provides a detailed and personalizable search interface.

Furthermore an open access to the data was seen as a chance for everyone to interact with it and as opportunity to stimulate creativity of the people.

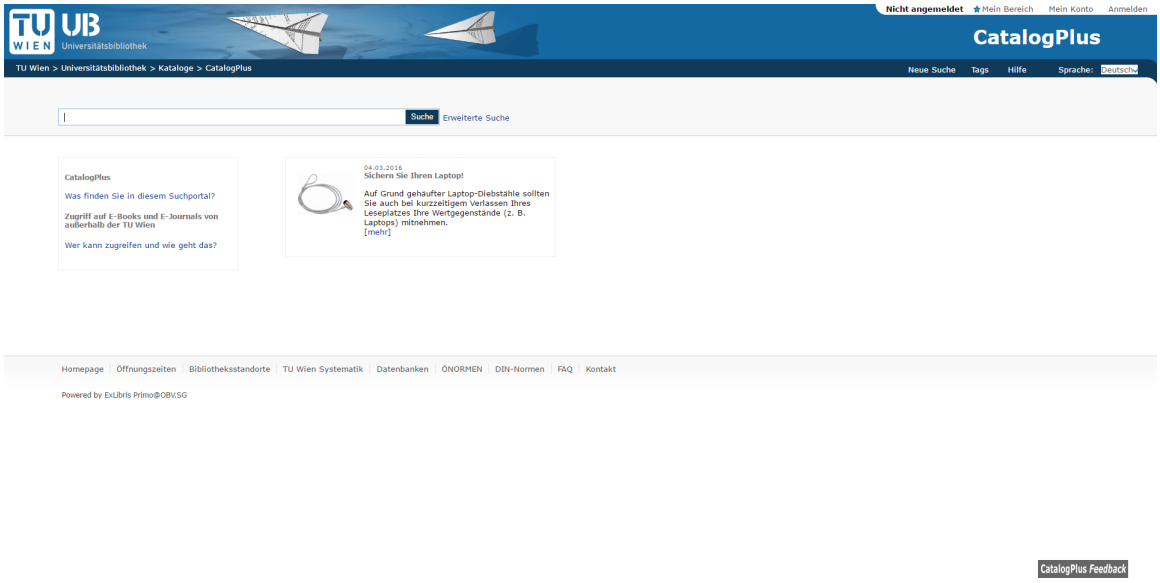
Barriers, potential disadvantages The library institution was very conservative perceived with skepticism, refusals and resentment against new ideas, especially in a technical term. It was stated by some interviewees, that this kind of project would only be seen as an extra amount of work in the library and not as an opportunity.

Beside the expected opposition there was also a real amount of work estimated by the interviewees. In particular there would be an effort to invest in digitizing books and keep this data up-to-date. This would be an important part of the project, otherwise the data would lose there value - not digitized data could not be distinguishable from not available data. Therefore it would be essential to have a complete and current database.

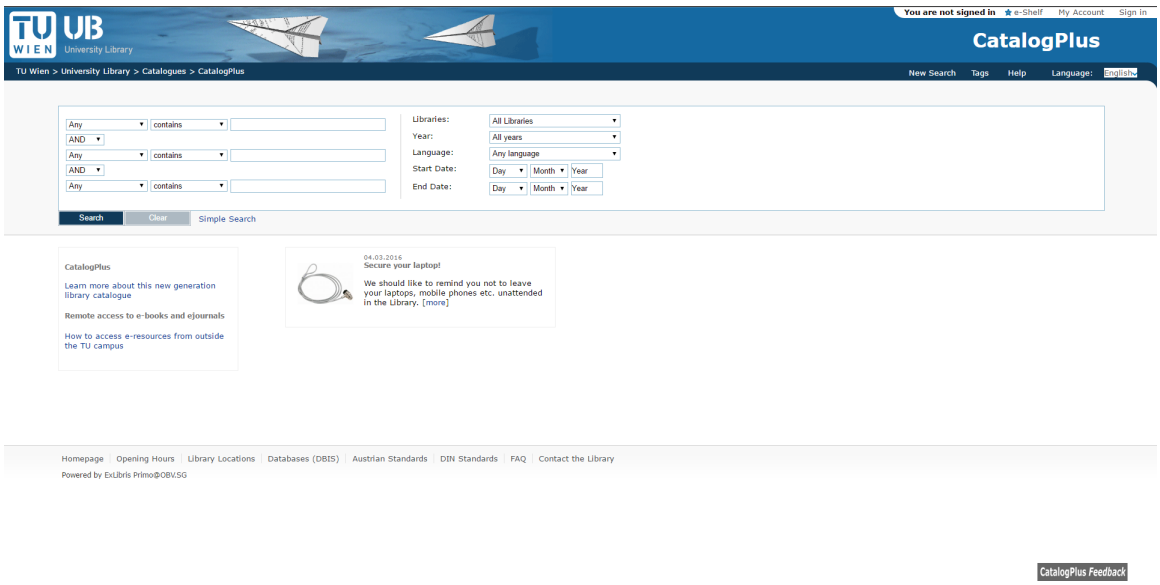
Another concern of the interviewed people were the copyright of the data. However, this problem was seen as handleable, because only meta-data would be open accessible.
At last, one interviewee declined the idea and usefulness of the LD4L and similar projects because of the existing platform Google Scholar [33]. He stated, that it already provides similar data and

[32] www.ub.tuwien.ac.at/
[33] www.scholar.google.at/

therefore everything he needs for his work. Further an additional platform would be too intricate to use.



(a) Homepage and simple search



(b) Extended search

Figure 7: Website of the TUVienna library

3.2.4 Data from the Publication database

Use case The proposed use case applied to the already existing publication database of the university. In the current state the database can be accessed via its website [34], where an interface for search exists. For other search parameters a request to the administration has to be made. The introduced LOD approach provides an LOD interface to the existing database, so the data can be accessed over e.g. SPARQL. This use case follows the principle of the existing "Open Research Online" [35] endpoint from the Open University, which expose information about publications originating from OU researchers using the BIBO Ontology (see section 4.2.2 for details).

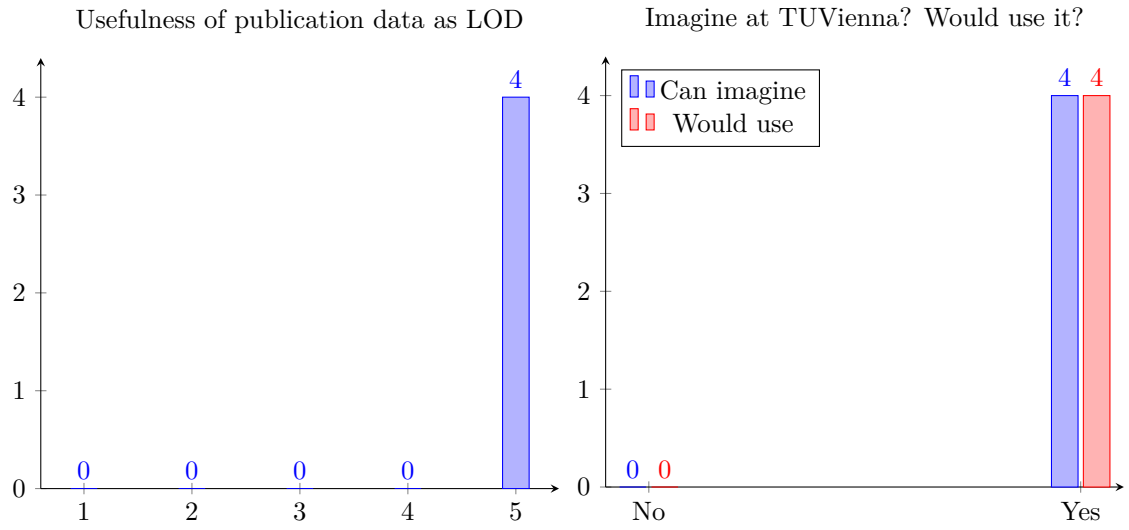


Figure 8: Usefulness of publication data as LOD

Statistically evaluation Similar to the library use case the interviewees liked the idea and found it all "extremely useful". Everyone could imagine such a project at TUVienna and would also use it.

Needs, potential benefits In contrast to the previous use case, the work amount of work was significant lesser stated, because the data are already there, just as the need of being up-to-date of the database.
Further some of the interviewed persons (independent to each other) come up with the idea of building an application based on the Linked Open Data interface to provide an overview of a researcher references as a widget or similar for a personal website. The data could be interlinked e.g. with the LOD interface of the Springer publisher [36] and complemented by including the Journal Impact Factor (JIF).

Barriers, potential disadvantages The interviewees identified uniformly a problem with the data ownership. Although there is an administration of the database, the *ownership* of the data itself is very unclear and therefore a contact and responsible person would be hard to find for such a project - there needs to be more investigation.
Another point, some of the interviewed persons were concerned of, was quality assurance - to use the data, they need to be synchronized with the original database. To taking care of this point, the implementation of a similar project has to ensure this.

[34] `www.publik.tuwien.ac.at/pubstart.php`
[35] `http://data.open.ac.uk/page/context/oro`
[36] `www.lod.springer.com/`

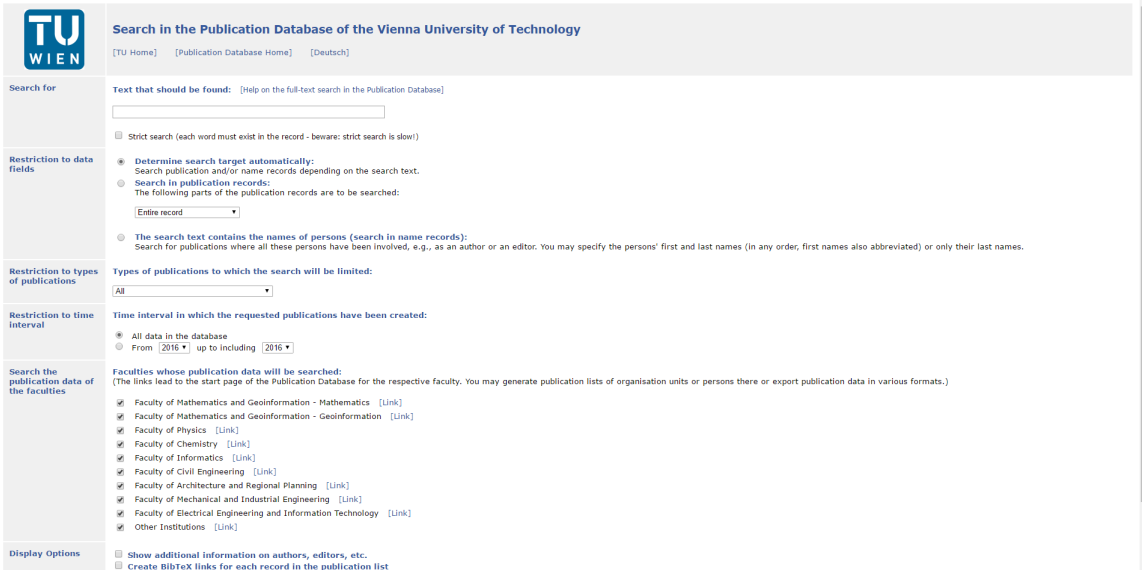


Figure 9: Publication database

3.2.5 Further use cases

In this subsection we describe some other use cases mentioned by the interviewees. Beside the two main use cases described above, some of the interviewees came up with their own ideas. This use cases are results of the flow of conversation, so they may not be fully reasoned, so consider this subsection as some kind of wish list or a collection of what *might* be possible.

3.2.5.1 "Sideproducts" of research

During the most research work a lot of "sideproducts" are produced. This can be all kind of statistics, survey data, project descriptions, work documents or sample data. The majority of this data get lost or somewhere stored after the project is done. Though they are evaluated and analyzed during the project, the data may be interesting from other, not intended perspectives - but they are not or only restricted-by-knowledge [37] available. If the data sets would be accessible through a LOD interface, a cross-disciplinary use are imaginable, e.g. a Visual Computing research project could use model data from a architecture research project without detailed knowledge of the original source (which could be gained through links in the LODs). To offer a sufficient data base, this kind of project may be more meaningful in an European context.

Besides the obvious benefits of this use case, there are many challenges and problems. First there has to be implemented a highly flexible LOD-system, which can simultaneously handle a high range of formats (e.g. documents, tables, raw data) and interlink them probably. Further, to make them interlinkable and searchable, there has to be invested a big amount of *manual* work to label the data. The result will be a huge list of keywords, which, again, has to be searchable. Another crucial point are the quality of the labeling: they had to be specific as possible to specify them accordingly as well as general as possible to free them from the original context and make them interdisciplinary understandable. And, at last, such a system must be as little as possible work to input the data from the project and anonymize them to avoid unnecessary frustration (which could easily overrule the benefits).

3.2.5.2 Statistics from TISS

The university management system TISS already provides a lot of statistically data, e.g. inscription data or general data about the students, so this data are optimal for a LOD use. The co-work of Haller [10] and Gamerith [8] explore and evaluate this use case.

3.2.5.3 Make Curricula comparable

One big problem of the communication between universities are the missing comparability of their curricula. To determine which university offers which course they have to manually be searched, listed and additionally matched with the corresponding, similar regarding to the content but different named courses at other universities. In particular this process is important for reworking the content of study programs to raise the coverage of the research field. Publishing the curricula of the universities as LOD would strongly improve this process and make them comparable.

Again, this use case makes only sense in an European context, e.g. in corporation with the Erasmus program to simplify the process of course crediting. Though it may be easy to publish single curricula of an university, it adds more profitableness to implement an Europe-wide system.

3.2.5.4 Overview of university projects and research focuses

Similar to the use case about the curricula described above, an Europe-wide publishing of university research projects (if they are not under restriction) and research focuses as LOD would be very useful to link researcher with other teams working on similar areas. Such a system could provide a research field based search and allow queries like "Who is currently working on non-classical logics in Europe?" or "Which university has a research focus on artificial social intelligence?".

This information may not only be interesting for researcher but also for companies, which looking for corporations, or scientific funding.

[37] only available if the existence are known

4 Proposed technical architecture and challenges (RQ4)

As an organization covering areas like research, student affairs and administration a university has to manage a significant amount of knowledge, adding new information on a daily basis. Such an application domain is complex and includes areas like management of an academic library and provision of educational resources which have to be conform to stakeholders requirements. Traditionally the *Service Oriented Architectures (SOA)* have been used to meet these needs. However, as the application domain grows many small and similar services tend to emerge. That phenomena cannot only be observed at the Vienna University of Technology, but also at the Open University[38] Zablith et al. [21].

A major problem of evolving similar, independent services are diverging data formats and service owners. Thus, knowledge and administrative information that has been collected by multiple services can not be easily interlinked. An example for such isolated services is the e-learning platform called TUWEL[39], combining moodle[40] and the central information system called TISS[41]. These services provide course information and material, but are intended for different purposes. Whereas TISS focuses mainly on administrative functionality, TUWEL supports the interaction between teacher and student. Adding additional services which, for example, synchronize deadlines and registration dates is costly due to the fact that the information is separated over different isolated sources and not easily accessible.

4.1 The big picture: a publication framework

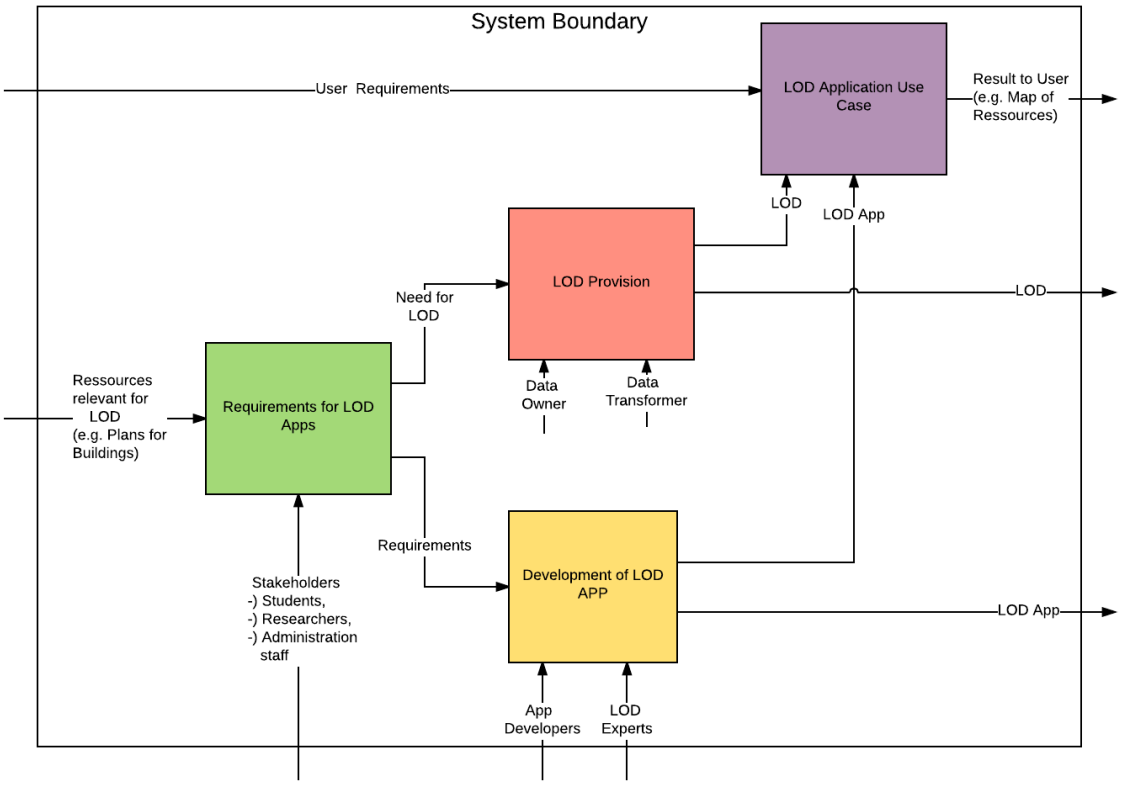


Figure 10: High level framework for LOD publishing in IDEF0 notation

[38] http://www.open.ac.uk/
[39] https://tuwel.tuwien.ac.at/
[40] https://moodle.org/
[41] https://tiss.tuwien.ac.at/

4.1.1 Parts

In this sub-section we provide an overview of a high level architecture, illustrated in Figure 10, for a university wide publication framework which includes:

- **Requirements for LOD-Applications**
  At the beginning stand the existing resources and the needs of the stakeholder. Combining these leads to the requirements for an application, the first step of the process. According to this requirement a decision has to be made whether they are realizable depending on the cost-value ratio and whether the solution has to be a LOD application. These process has to actively involve all stakeholders.

- **Data Provision**
  After defining the requirements, the existing data (e.g. a publication database) must be transformed in a appropriate, machine readable LOD format. These can happen in a manually (only for small data sets) or a (semi-)automated (for big data set) way. Ideally there is a automated transformer based on an existing, well maintained and up-to-date database so there has to be less cared about the of the data (for a more technical description see Section 4.2.1). Key roles for this process are the original data owner and the data transformer.

- **Development of an Application**
  Based on the requirements definitions from the previous step a proper application (e.g. a browser of publication data) now can be constructed considering the stakeholder needs and existing resources (transformed to LOD). To support this process and to not obtaining all knowledge from zero it is recommended to access the knowledge of LOD experts (e.g. provided by the LinkedUniversities [42]). The development can simultaneously be done with the data provision if proper interface between the application and its data are made.

- **Application Use Case**
  Combining the data, transformed in LOD, the LOD application and user requirements (not the application requirements from the first step) result in the actual application use case, representing the environment or the domain.

4.1.2 In- and outbound interfaces

There were several indirectly interfaces mentioned above - we define them now in a more formal way and divide them in inbound (arrows pointing from outside the system boundary) and outbound interfaces (arrows pointing from inside the system boundary).

The inbound interfaces are:

- **User Requirements**
  Understanding user requirements is an essential part of the software development process. Due to the open nature of Linked Open Data privacy concerns and legal issues should be already considered in the requirements as misunderstandings are hard to fix in later phases.

- **Existing Resources**
  The starting point of every LOD application are resources, ideally already existing (to reduce the amount of work). It can be everything from relational databases, simple Excel files or other Linked (Open) Data sets. For more details see section 4.2.1.

- **Stakeholders**
  Stakeholders are defined as *"a person, group or organization that has interest or concern in an organization. Stakeholders can affect or be affected by the organization's actions, objectives and policies."* Post et al. [15] Their needs and demands have to be considered in a LOD project as similar as in every other software project.

- **Application Developers**

- **LOD Experts**
  To avoid unnecessary redundancy in acquiring knowledge of LOD implementation, it is highly recommended to involve either LOD experts in the development or access their accumulated know-how e.g. by platforms like LinkedUniversities [43] (see section 2.1)

[42] `www.linkeduniversities.org/`
[43] `www.linkeduniversities.org/`

- **Data Owner**
  Every data set has its owner, therefore this role has to be considered in the development process to avoid organizational conflicts and copyright issues. Ideally he is directly involved to access his specific know-how about the data set.

- **Data Transformer**
  To use a data set in a LOD approach, the data have to be transformed either manually or (semi-)automatically into a proper format (see Berners-Lee [1] for the 5 star model of data format and section 4.2.1 for details about transformations).

The outbound interfaces are:

- **Linked Open Data**
  As a result of the LOD provision the actual data are provided e.g. as SPARQL endpoint, so others can easily access and use them in other applications. For technical details about endpoints see section 4.2.4 or as example the SPARQL endpoint of The Open University `www.data.open.ac.uk/query`.

- **Application**
  The outcome of the development phase are applications for end-users to access and interact with the data e.g. in form of a web platform. For technical details about endpoints see section 4.2.4 or as example the application list of The Open University `www.data.open.ac.uk/applications.html`.

- **End-User Result**
  Finally the main result is the actual interaction of the users with the applications and endpoints, (hoepfully) acting in the boundaries of the defined use cases.

4.2 Proposal of a technical architecture

In order to overcome the environment of data silos and to support the evolvement of a university-wide data space for public resources, a technical architecture following Linked Data principles is proposed and discussed in this section. Berners-Lee suggested four principles for publishing data on the web, which can be considered as best practise. The compliance with this best practise leads to Linked Data; "the idea is that the more people follow these principles, the more their data will be usable by others."[17]. The four principles stated by Berners-Lee (2006) [1] are:

1. Use URIs as names for things
2. Use HTTP URIs so that people can look up those names.
3. When someone looks up a URI, provide useful information, using the standards (RDF, RDFS, SPARQL)
4. Include links to other URIs. so that they can discover more things.

Linked Open Data extends Linked Data with the idea of publishing it in a open way.

> Open means anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness).[20]

In order to transform demanded datasets of the university into Linked Data and publish it in an open way, a system architecture is required that faces the issues of diverging data sources and data freshness."An important architectural pattern used in system development is the multitier architecture. It logically separates the functionality of the system in a number of layers and specifies the communication between those layers."[17].

.

The proposed architecture is mainly inspired by the toolkit "`Tabloid`" ("Toolkit ABout Linked Open Institutional Data") by the LUCERO Project [44] - a collection of tools, examples and documentations. The general principle is a system, which extract RDF data from existing data sources (section 4.2.1), load them into a triple store (section 4.2.3) and finally expose them to the web (section 4.2.4). For illustration see the LUCERO workflow in figure 12 (in this part only the generic parts are described). In this workflow there are way more components than we will talk about (e.g. mechanism of detecting data changing) than shown in the figure. Additional a more generic architecture can be seen in figure 11.

4.2.1 Collect & extract data from sources

The first step must be to collect and extract the data from the source and transform them into a proper LOD format, e.g. RDF. By now there are basically for every data format tools for transforming them into RDF, LinkedUniversities made a collection of them [46], here are the main tools:

- `From Relational Databases`
  - `Triplify` [47] is a tool which use SQL queries to generate RDF data from a relational database.
  - `D2RQ` [48] is a tool which also use SQL queries but with the use of a mapping that relates the structure of a database to RDF triples. It transformers SPARQL queries at run-time into SQL queries using this mapping.
- **From XML and RSS**
  In terms of syntax, XML, RSS and RDF sharing the same base, so there have to be fewer effort to be done to transform them syntactic, commonly using XSLT. W3C recommended the `GRDDL` [49] language for this purpose.

[44] d'Aquin et al. [4]
[46] `http://linkeduniversities.org/lu/index.php/tools/index.html`
[47] `http://triplify.org/`
[48] `http://www4.wiwiss.fu-berlin.de/bizer/d2r-server/`
[49] `http://www.w3.org/TR/grddl/`

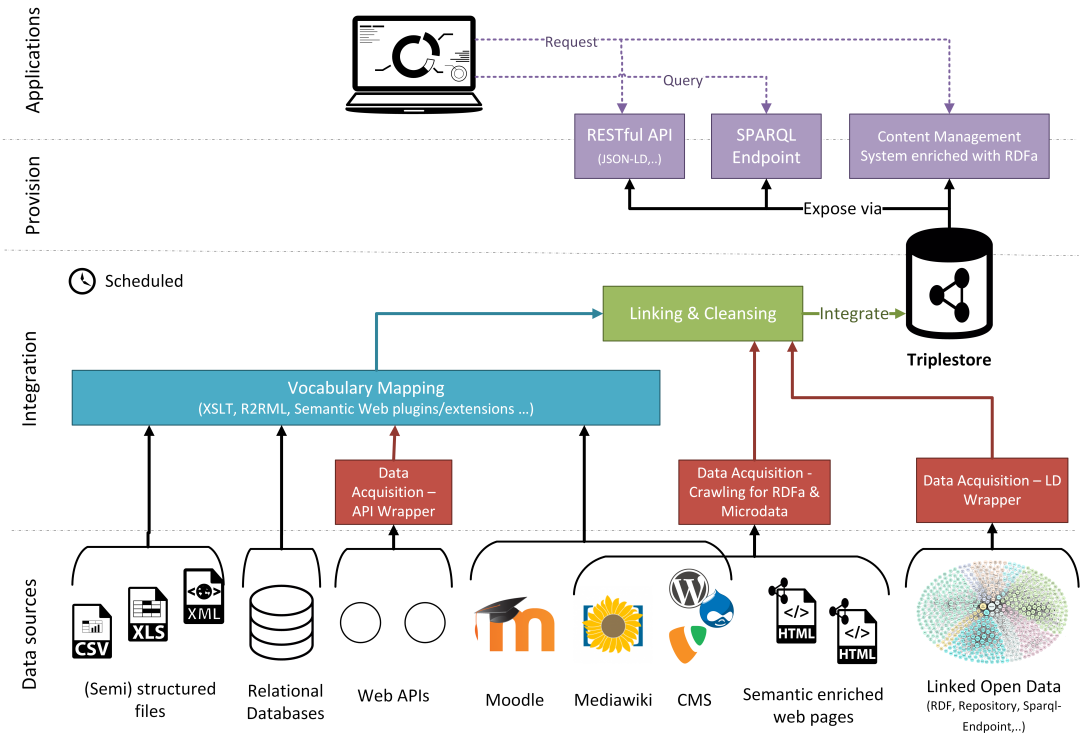


Figure 11: Generic high level technical architecture for providing Linked Open Data (adapted from EUCLID [17])

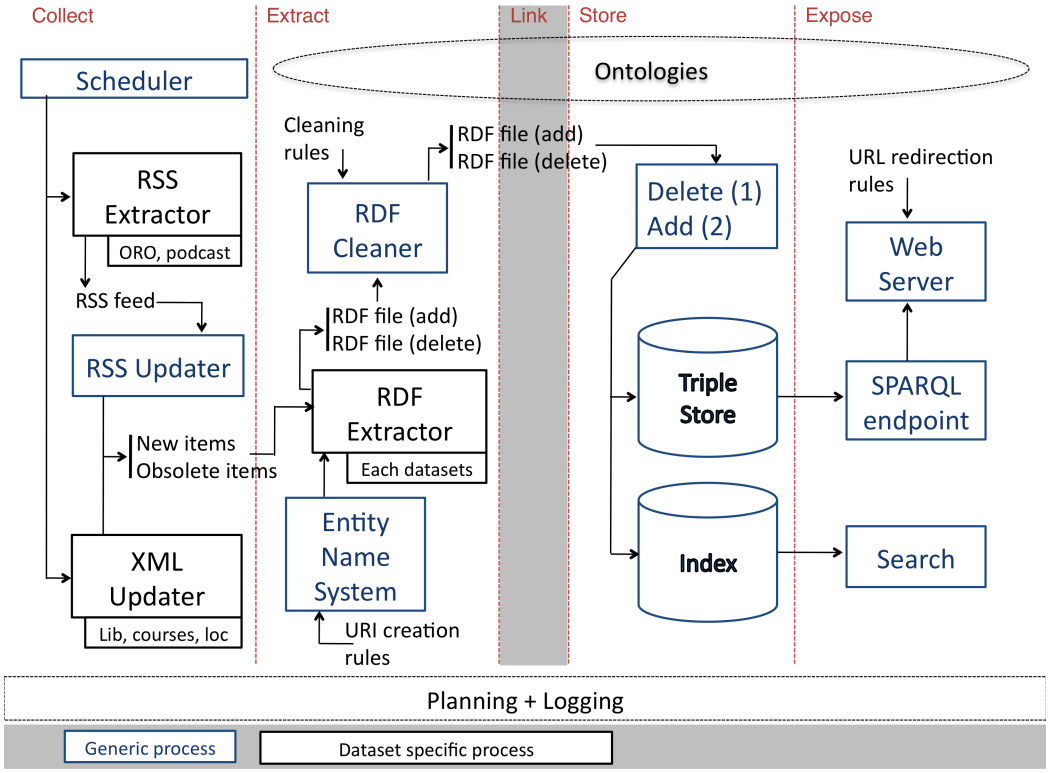


Figure 12: Example: the LUCERO workflow[45]

- **From Tables and Spreadsheets**
  - `Google Refine` [50] with the RDF Extension [51] is an easy way to clean, transform and explore data in a tabular format including MS Excel, Google Spreadsheet and CSV.
  - Other tools like `Any23` [52] or `QUIDICRC` [53] providing simple transformation from CSV to RDF

After extracting the data as RDF from there source, they need to be cleaned and interlinked with themselves and other data.

4.2.2 Vocabularies and Ontologies

To represent and store the collected data, they must be mapped into an ontology and a vocabulary.

Gruber [9] defines an *ontology* as "*an explicit specification of a conceptualization*". *Conceptualizations* are objects, entities or concepts that may or may not exist in the universe. In addition to that, the *vocabulary* defines the relationships between those objects. In other words, the vocabulary defines the conceptual model of what can be represented. Bock et al. [2] describe ontologies from a more practical point of view defining the three conceptual components of an ontology - *classes*, *instances* and *properties*. Regardless of what concrete implementation of an ontology is used graphical representations (e.g. graph) are preferred over textual ones to give a high level overview of the concepts used in an ontology.

Again, LinkedUniversities listed a lot of useful vocabularies and ontologies on their website [54]. As an example for such an ontology let's take a look at the use case "publication database" from section 3.2.4. To mapping the data from this database a bibliographic ontology is needed. LinkedUniversities recommend the `BIBO` (Bibliographic Ontology) [55]. It can be used as a citation ontology, as a document classification ontology, or simply as a way to describe any kind of document in RDF, so it is ideally for this purpose. You can see the `BIBO` graph in figure 13.

4.2.3 Store the data

Considering the performance of such a system and to avoid too much communication with the original data source, the center of the system is a triple store, where the extracted data are stored and from there exposed to the web. A triple store (or RDF store) is a purpose-built database, "*designed to store and retrieve identities that are constructed from triplex collections of strings (sequences of letters). These triplex collections represent a subject-predicate-object relationship that more or less corresponds to the definition put forth by the RDF standard.*" Rusher [16]. There are a lot of implementations of such a triple store, like `Sesame` [56], `Jena TDB` [57], `4Store` [58] or `SwiftOWLIM` [59]. The LUCERO project settled for the `SwiftOWLIM`, because it is "*free, scalable and efficient, and includes limited reasoning capabilities, which might end up being useful in the future*" d'Aquin et al. [4].

4.2.4 Expose the data to the Web

The last step is obviously to expose the data from the triple store to the web. For this purpose a SPARQL endpoint is commonly used, but there are also other way, e.g. resolvable URIs, the linked data API [60] or an Open RESTful API. All of them make the data available so others can query and access them to use it for their application or to serve themselves as resource for linking it with other data.

[50] `http://code.google.com/p/google-refine/`
[51] `http://lab.linkeddata.deri.ie/2010/grefine-rdf-extension/`
[52] `http://lab.linkeddata.deri.ie/2010/grefine-rdf-extension/` (at the moment of the work unavailable)
[53] `http://any23.org/`(at the moment of the work unavailable)
[54] `http://linkeduniversities.org/lu/index.php/vocabularies/index.html`
[55] `http://bibliontology.com/`
[56] `http://rdf4j.org/`
[57] `http://openjena.org/TDB/`
[58] `http://4store.org/`
[59] `http://www.ontotext.com/owlim/`
[60] `http://code.google.com/p/linked-data-api/`

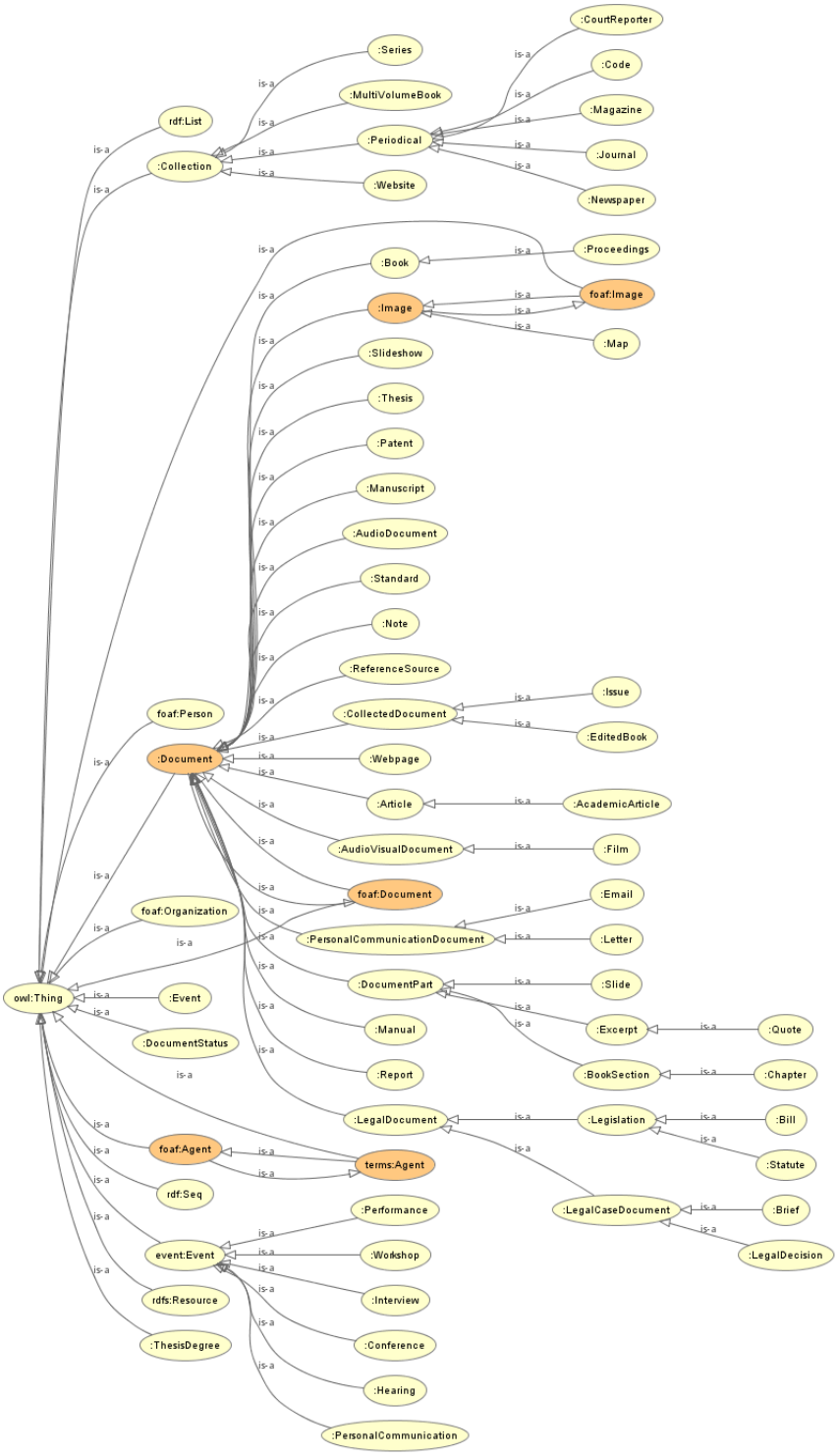


Figure 13: Graph of the Bibliographic Ontology

4.3 Challenges

In this section we will briefly describe some of the most crucial factors which has to be faced when implementing a system similar to the proposed one.

4.3.1 Data ownership

When exposing data as LOD the ownership of the original data has to be cleared without any doubt, so there are no legal concerns to be cared about when implementing the system. An unclear ownership may otherwise lead to a complete failure of the project, if the worst comes to the worst at a time where already a lot of time and work are spent. Further, to ensure a stable system, the ownership has be clear over a long time in the future - otherwise other applications, built on the data, loose their source.

4.3.2 Data quality

Publishing a LOD system there must be a compromise found between actuality of the data and a proper system load of the interface to the data source (e.g. an active used database), so existing system may not be (to heavily) effected by the LOD system. According to the data the actuality may range between weeks and days (e.g. for location data) or minutes and seconds (e.g. public transport data).

Further it must be ensured that the linking of the data are valid at every time, linking to no longer existing sources or sources with sinking data quality will lower the quality of data itself.

4.3.3 Response time of the LOD interface

To achieve not only a providing of the data but also a use of the data in applications, the interface (e.g.

5 Conclusions and future work

Revisit each research question and give a condensed answer derived from possibly multiple methods (e.g., benefits: as identified in interviews, as seen at other universities)

The opportunity to develop new ideas based on an access to open data was very common welcome across all questions (though no concrete ideas came up).

Costs

6 Acknowledgments

The authors would like to thank...

References

[1] Tim Berners-Lee. Linked data - design issues. *W3C*, (09/20), 2006. URL http://www.w3.org/DesignIssues/LinkedData.html.

[2] Conrad Bock, Peter Haase, Rinke Hoekstra, Ian Horrocks, Alan Ruttenberg, Uli Sattler, and Mike Smith. OWL 2 web ontology language structural specification and functional-style syntax. *W3C Recommendation, 2nd edn.(December 11, 2012)*, 2008.

[3] Mathieu d'Aquin, Fouad Zablith, Enrico Motta, Owen Stephens, Stuart Brown, Salman Elahi, and Richard Nurse. LUCERO - Aims, Objectives and Final Outputs of the Project. http://lucero-project.info/lb/2010/06/lucero-aims-objectives-and-final-outputs-of-the-project/index.html, 2010. [accessed 29-February-2016].

[4] Mathieu d'Aquin, Fouad Zablith, Enrico Motta, Owen Stephens, Stuart Brown, Salman Elahi, and Richard Nurse. The LUCERO Project: Tabloid. http://lucero-project.info/lb/tabloid/index.html, 2010. [accessed 07-March-2016].

[5] Mathieu d'Aquin, Carsten Kessler, and Tomi Kauppinen. Members of Linked Universities. http://linkeduniversities.org/lu/index.php/members/index.html, 2014. [accessed 24-February-2016].

[6] Mathieu d'Aquin, Carsten Kessler, and Tomi Kauppinen. Linked universities. http://linkeduniversities.org/index.html, 2014. [accessed 24-February-2016].

[7] Hendrik Drachsler, Slavi Stoyanov, Marieke Guy, and Maren Scheffel. Final Version of the LinkedUp Evaluation Framework. http://linkedup-project.eu/files/2014/11/LinkedUp_D2.2.2.pdf, 2014. [accessed 28-March-2016].

[8] Stefan Gamerith. Publishing TUWien Data as Linked Open Data. 2016.

[9] Thomas R. Gruber. A translation approach to portable ontology specifications. *Knowl. Acquis.*, 5(2):199–220, 1993.

[10] Kevin Haller. Linked Open Data at the Vienna University of Technology - A case study concerning students. 2016.

[11] Eelco Herder, Stefan Dietze, Besnik Fetahu, Mathieu D'Aquin, and Constantin Graf von Rex. State of the art and data assessment. http://linkedup-project.eu/files/2013/09/LinkedUp-D1.1-SotA.pdf, 2013. [accessed 25-March-2016].

[12] Eelco Herder, Marieke Guy, and Hendrik Drachsler. LinkedUp Challenge Results. http://linkedup-project.eu/files/2014/11/LinkedUp_D1.3.pdf, 2014. [accessed 27-March-2016].

[13] Ian Jacobs and Norman Walsh. Architecture of the World Wide Web, Volume One. W3c:rec, W3C, 15 dec 2004. URL http://www.w3.org/TR/2004/REC-webarch-20041215/. http://www.w3.org/TR/2004/REC-webarch-20041215/.

[14] Marijn Janssen, Yannis Charalabidis, and Anneke Zuiderwijk. Benefits, adoption barriers and myths of open data and open government. *Information Systems Management*, 29(4):258–268, 2012.

[15] James E. Post, Lee Preston, and Sybille Sachs. *Redefining the Corporation - Stakeholder Management and Organizational Wealth*. Stanford University Press, Stanford, new. edition, 2002. ISBN 978-0-804-74310-5.

[16] Jack Rusher. Triplestore, semantic web advanced development for europe (swad-europe), workshop on semantic web storage and retrieval - position papers. https://www.w3.org/2001/sw/Europe/events/20031113-storage/positions/rusher.html, 2003. [accessed 07-March-2016].

[17] E. Simperl, B. Norton, M. Maleshkova, J. Domingue, A. Mikroyannidis, P. Mulholland, and R. Power. *Using Linked Data Effectively*. The Open University, 2013. ISBN 978-1-78007-926-4. URL http://www.euclid-project.eu/.

[18] LD4L Project Team. Linked data for libraries. http://www.ld4l.org/, 2015. [accessed 29-March-2016].

[19] LD4L Project Team. Ld4l - expected outcomes. http://www.ld4l.org/node/15, 2015. [accessed 29-March-2016].

[20] LD4L Project Team. Ld4l - expected outcomes. http://opendefinition.org/, 2015. [accessed 30-March-2016].

[21] Fouad Zablith, Mathieu d'Aquin, Stuart Brown, and Liam Green-Hughes. Consuming Linked Data within a Large Educational Organization. Bonn, Germany, 2011.

.1 Questionaire

QUESTIONNAIRE
LINKED OPEN DATA

Name:_____

Organization:_____

Position (Role):_____

General Questions

1. **What are is your area of responsibility? How would you characterize your *daily* work tasks?**

_____

_____

_____

2. **How would you say classify your level of experience with Information Systems?**

Fundamental	Novice	Intermediate	Advanced	Expert
☐	☐	☐	☐	☐

3. **How would you classify your level of expertise with Linked Open Data?**

never heard	heard but never used	used in small example	used in practice	Expert in LOD
☐	☐	☐	☐	☐
1	2	3	4	5

LOD in Research

4. How useful do you find the improvements with Linked Data in libraries? E.g. in context of an use case of my presentation

not useful		useful		extremely useful
☐	☐	☐	☐	☐
1	2	3	4	5

Why?

_____

_____

5. Can you imagine a similar project at the Vienna University of Technology?

Yes ☐
No, because:

_____

_____

What benefits do you see?

_____

_____

What disadvantages or barriers do you see? E.g. copyright?

_____

_____

Would you use it?

Yes ☐
No, because:

Can you imagine/recommend other roles/persons?

6. Do you think that publishing the scientific publications or their metadata of TU as LOD might be useful?

not useful		useful		extremely useful
☐	☐	☐	☐	☐
1	2	3	4	5

Why?

_____

_____

7. Can you imagine a similar project at the Vienna University of Technology?

Yes ☐
No, because:

_____

_____

What benefits do you see?

_____

_____

What disadvantages or barriers do you see?

_____

_____

Would you use it?

Yes ☐
No, because:

_____

Can you imagine/recommend other roles/persons?

LOD Applications

8. Which kinds of applications based on LOD could be useful in a research context (e.g., single researcher, research workgroup, faculty, scientific community)?

What benefits do you see?

What disadvantages or barriers do you see?

Would you use it?

Yes ☐
No, because:

Can you imagine/recommend other roles/persons?

9. Which kinds of applications based on LOD could be useful in a general university context (e.g., teaching, administration, orientation)?

What benefits do you see?

What disadvantages or barriers do you see?

Would you use it?

Yes ☐
No, because:

Can you imagine/recommend other roles/persons?

LOD Data

10. Which kinds of data could be useful to publish as LOD in a research context (e.g., single researcher, research workgroup, faculty, scientific community)? Do you have any data that may be useful?

What benefits do you see?

What disadvantages or barriers do you see?

Would you use it?

Yes ☐
No, because:

Can you imagine/recommend other roles/persons?

11. **Which kinds of data could be useful to publish as LOD in a general university context (e.g., teaching, administration, orientation)?** Do you have any data that may be useful?

_____

_____

_____

_____

What benefits do you see?

_____

_____

What disadvantages or barriers do you see?

_____

_____

Would you use it?

Yes ☐
No, because:

_____

_____

Can you imagine/recommend other roles/persons?

_____

_____

End

12. Do you know any other person who might be relevant for this interview?

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

13. Do you want the final report of the outcome of the study?

Yes	☐
No	☐

.2 Contact persons for further investigations