Linked Open Data
at the
Vienna University of Technology
A case study concerning students

Kevin Haller[a], *Author*
Lukas Baronyai[b], *Co-Author*
Stefan Gamerith[c], *Co-Author*
{kevin.haller[a],e1326526[b],e0925081[c]}@student.tuwien.ac.at

March 2016

Abstract

The abstract goes here.

Keywords. Linked Open Data, University, Case Study

Contents

1 Introduction

While the pressure on governments and public organizations to release *Open Data* 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 as a whole.

Open Data includes non-privacy-restricted and non-confidential data. Therefore any restrictions in distribution are prohibited and data is funded only by public money. [1] The application domain for Open Data providers is not restricted by its nature in any way and ranges from traffic, weather, statistics to budgeting in the public sector. Just the publication of Open Data seems not enough but in addition the implementation of a feedback loop result in *Open Government*. This has the advantage of a constant adoption to the citizen's needs instead of just visualizing former closed data.

Despite its wide adoption of Open Data it does restrict the published Data format in any way, thus complicating the integration of heterogeneous data sets. The World Wide Web has proven great success spreading knowledge of various data sources all over the world. The building block of the Web are documents and links connecting them to form a global information space. This can be seen as the key success factor in its nearly unconstrained growth [2]. Following these principles of publishing and connecting data on the Web is known as *Linked Data*. More technically, it refers to machine-readable data which is linked to external data sets and can in turn be linked from other data sets.

[3] developed a five star rating scheme for classifying *Linked Open Data*, which combines Linked Data and Open Data. The scheme ranges from one star describing Open Data only to five stars describing Open Data in a machine-readable format using open standards with links to other data sets.

Although Linked Open Data offer universities new opportunities for providing unprecedented insight into its core activities and ease application development, a major **problem** is that **Linked Open Data has not been widely adopted by universities yet**. Even tough there are a few examples [4] of publishing university related data as Linked Open Data, there has been little knowledge of using Linked Open Data for publishing university related information.

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

1.1 Research Questions

The fundamental research questions underlying this paper is:

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

More concrete this paper concentrates on the following specific research questions:

Q1: What are best practices regarding the applicability of Linked Open Data in university settings? As of now, there are no established best practices for the use of Linked Open Data due to its little adoption in university contexts. For this very reason it is crucial to identify strengths and limitation from previous experiences [4] of using Linked Open Data as core technology.

Q2: What are major benefits and barriers for each stakeholder and what are useful use cases? We identified three different stakeholders in university context *Students*, *Researchers* and *Administration staff*. Since the success of any new technology highly depends on the acceptance of the stakeholders, the needs of each of the target groups needs to be examined. Furthermore, use cases are important to showcase profits and shortcomings to a non-technical audience.

Q3: What are major challenges for the implementation of a Linked Open Data solution? As the implementation of a Linked Open Data solution is a time consuming task, the knowledge of probable challenges from the technical perspective as well as from the management perspective is a key factor for the successful adoption.

Q4: How would a prototypical implementation of Linked Open Data look like? Among the various existing data sets available it needs to be investigated if a (semi-) automatic transformation is feasible or is the manual data provision enough. In addition, from an implementers perspective of view, critical factors regarding the storage and retrieval of Linked Open Data need to be identified.

1.2   Methodology

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

A coordinated set of semi-structured interviews To answer research questions (RQ) two to four, we interviewed a selected set of stakeholders representing *Students*, *Researchers* and *Administration staff* respectively. Semi-structured interviews were selected as the means of data collection because they are well suited for exploring the impressions and interests of the interviewees as in a discussion while still following a defined structure.

Litrature Review Undertaking a litrature review to justify scientific contributions and making sound conclusions is an established practice in any scientific community. Since our scientific work targeted in particular to the Semantic Web community, we made some pre-assumptions of a basic understanding of the technologies and concepts regarding Linked Data. More specifically, the concept of

an ontology and example knowledge descriptions languages describing these will not be covered in this paper.

Conceptual System Design The development of applications based on Linked Open Data requires a methodology which describes a common understanding of the overall system infrastructure. Therefore we designed a conceptual model of a prototypical implementation of a Linked Open Data solution.

1.3 Contributions

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

1. Identifying best practices for Linked Open Data in university context. Due to the crowing complexity and the large amount of data information systems need to process, there has been the need to efficiently handle Linked Data as well. We gave a brief overview of the already published research work regarding Linked Open Data in university context. In particular, we compared the profits and shortcomings in existing Linked Open Data solutions.

2. Finding benefits/barriers with additional use cases for stakeholders. As with every software project the very first phase of the Software Development Lifecycle (SDLC) is the *Evaluation of the Requirements*. As a Linked Open Data solution has additional requirements to the structure of the data and due to its open nature, we investigated if the overhead compared to an established technology (e.g. a database based solution) is worth the effort. A set of selected participants from the areas Research, Student Affairs and Administration are interviewed at the University of Technology in Vienna and their benefits/barriers are compared. Additionally, we proposed several use cases emphasising 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 a defined set of Linked Open Data applications which were merged together from the conducted interviews. That use cases showcased probable shortcomings which might arise before, during or after the implementation.

4. Sketching a prototypical implementation of a Linked Open Data solution. In consideration of the above mentioned obstacles of a possible Linked Open Data solution, we gave an outline of a prototypical implementation. It begins by covering the whole process of data provision and ends by applications made for end users.

1.4 Structure of this Paper

The rest of the paper is structured as follows: Section 2 provides a summary of existing efforts and works related to the research of this paper. Section 3 discusses the methodology as well as the results of the conducted case study concerning research question 2 and 3 from the perspective of students. Section 4.1 presents a model for describing how an LOD effort can evolve at a university, section 4.2 proposes a technical architecture for applying Linked Open Data principles to public resources of universities and section 4.3 discusses challenges. Whereas section 4.2 is targeted to readers concerned with information & communication technologies and software architectures, all other sections requires no deeper technical understanding of Linked Data than provided by the introduction, Finally, conclusions are drawn in section 5.

2 Related work

2.1 Linked Universities

'Linked Universities' is "'*an alliance of european universities engaged into exposing their public data as linked data*"'[1] providing help and knowledge for other universities that wish to expose their public university resources as Linked Data. 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 Linked Universities offering a portal as collaborative space with common vocabularies and practices for reusing, describing and sharing.

Their goals are:([5])

- "'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"'

The members of this alliance are:([4])

- The Open University, UK

- University of Münster, Germany

- Aalto University, Finland

- University of Southampton

---

[1][5]

- Royal Institute of Technology (KTH) / MetaSolutions AB

- Aristotle University of Thessaloniki, Greece

- Ege University, Turkey

- Charles University in Prague

- Universitat Pompeu Fabra

2.1.1 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*"'.[6] The projects connected with other organizations through LinkedUniversities.org to gather common issues and practices. The outcome was the first university linked data platform, `http://data.open.ac.uk/`, with much impact on The Open University and the education community.

2.2 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[2], `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[3] in connection with his master thesis[4] to provide a subset of the Open Government Data provided by Vienna (`data.wien.gv.at`) as

---

[2]Member of 'Cooperation OGD Austria'

[3]`http://cweiss.net/lod/` last accessed on 15.03.2016

[4]C. Weiss, 'Transferring Open Government Data into the global Linked Open Data Cloud', 2013

Linked Open Data. In the same year 2013 the project LODPilot[5] 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`[6] 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 any more and outdated. `Open Street Maps`[7] 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 build an infrastructure to provide this collected information as Linked Open Data with the project named `LinkedGeoData`[8]. `DBPedia`[9] 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[10]. `Wikidata`[11] is a project similar to `DBPedia` and also provides knowledge about Austria in a machine-processable format.

2.3 Linked Education

Linked Education represents the idea of using Linked Data principles for educational purposes.

...

2.4 Earlier studies

...

3 Benefits and challenges of using Linked Open Data at universities

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, where *students* and persons affiliated with *student affairs* at the Vienna University of Technology were the case. The essential question of the case study was RQ2; what are the major benefits and challenges for students if certain university resources are published as Linked Open Data and which needful use cases can

---

[5] http://lodpilot.at/
[6] http://www.wegweiser.ac.at/
[7] https://www.openstreetmap.org/
[8] http://linkedgeodata.org/About
[9] http://wiki.dbpedia.org/
[10] http://dbpedia.org/page/Vienna_University_of_Technology
[11] https://www.wikidata.org/

result out of this from the perspective of the students. RQ3 played a minor role in the case study, because *students* are in general on the consuming side of information technology at a university. However, there are *students* that are operating information systems (e.g. wikis or internet forums), which collect a not insignificant amount of knowledge, which is why this research question was also of interest. The case study was conducted at the Vienna University of Technology, where teaching and research is focused on engineering and natural science. It has several faculties including Computer Sciences and Electrical Engineering & Information Technology, which is exactly why the exploration of the willingness of students to develop applications based on LOD was also a goal. **5** participants from student representatives, teaching & administration personnel to ordinary students participated in the case study.

A case study concerning *researchers* was conducted by Lukas Baronyai [7] and another one concerning *administration employees* by Stefan Gamerith [8].

The first section describes the applied methodology (see 3.1). The results are split in three sections; the first one describes the background of the interviewees (see 3.2). The second one describes the proposed use cases and the evaluated acceptance of them as well as examined additional use cases and requested datasets (see 3.3). The third section shows the result of the evaluated willingness of the interviewees to develop applications based on Linked Open Data provided by the university (see 3.4).

3.1 Methodology

Since the idea of Linked Open Data as well as the idea of applying it to the domain of universities is around for a while now (e.g. Linked Universities, 2011), the first step was to discover Linked Open Data initiatives of universities that have already been started and case studies that have already been conducted. A brief description of the results can be seen in section 2. The discovered case studies and initiatives were analysed and two common use cases, which may be the most interesting ones from the point of a student, were extracted; a map of places of interest (see 3.3.1) and a resource list management application (see 3.3.2).

Afterwards the case study was designed and semi-structured interviews were planned to answer the research questions of the case study. Both extracted use cases were part of the interviews; on the one hand to give the interviewees an idea what Linked Open Data can be used for and on the other hand to evaluate the acceptance and doubts regarding these use cases. Potential interviewees were collected by suggestions of prior interviewees and persons that are quite familiar with the domain of the university. In summary, five interviews were conducted and at the time of the conduction all interviewees were *students* or persons affiliated with *student affairs*. A more detailed description of the interviewees can be found in section 3.2. The design of the interview is further discussed in the following section.

3.1.1 Interview design

In order to answer the given research questions of the case study **five** semi-structured interviews were conducted. In a semi-structured interview the questions are planned in advance, but they are not asked in the same order as they are listed. A semi-structured interview let researchers have their freedom to decide, which question to ask next in dependence on the development of the conversation. Additionally, semi-structured interviews allow for improvisation and exploration of the studied objects [9]. This is exactly why a semi-structured approach was used, because in this case study the participant of the interview shall have the freedom to express his/her ideas without being constrained by a fixed order of planned questions. In order to have a guideline for the interviews, a questionnaire was created, which can be seen in the appendix (see A). The questionnaire is composed of five parts:

1. An introduction to the research team.

2. General questions about the background of the interviewee including for example the level of expertise in the field of Linked Open Data.

3. An introduction to Linked Open Data by using a simple example.

4. Proposal of both extracted use cases and open question about the acceptance and opinion regarding these use cases.

5. Open questions to gather ideas about potential use cases and datasets for which a more open access is requested. In addition, open questions about the willingness to develop applications based on LOD provided by the university.

The list represents in which order the parts of the questionnaire were addressed during the interview, but as mentioned earlier due to the nature of semi-structured interviews there is no need for keeping this order. First, the research team was introduced to the interviewee as well as the objective. Afterwards, the moderator asked the interviewee about their background. Due to the question about the level of expertise in the field of Linked Open Data, the moderator was able to decide whether an introduction to Linked Open Data is required or not. The maybe skipped introduction was followed by the last two parts, where the moderator asked open questions concerning the research goals of the case study.

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. An alternative is the recording of the conversation, but in this case the single researcher has to spend more time (in worst case two times more) to extract important contributions from the conversation and the interviewee may not be

in favour of it. On the other hand no second researcher, who is familiar with the study, is required and the recording can be used for later analysis. Nevertheless the first approach was used.

3.2 Background of interviewees

At the beginning of the interviews, the five participants had to answer some questions about their background. They had to describe their daily work tasks and responsibilities at the university. Furthermore, the interviewees had to estimate their level of expertise in the field of Information & Communication Technologies and in the field of Linked Open Data. Table 1 shows the results and table 2 shows the used rating scales.

		Level of expertise		
ID	Assignment	ICT (1-5)	LOD (1-5)	Date of interview
A	student representative	3	1	24.11.2015
B	teaching & administration	5	2	03.12.2015
C	teaching & administration	5	1	17.12.2015
D	student	4	3	12.02.2016
E	student	4	3	18.02.2016

Table 1: Background information of the interviewees and the corresponding interviews.

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.

Table 2: Rating scales for the level of expertise in the field of ICT and LOD.

3.3 Use cases and needs

3.3.1 Map of places of interest

Description: The facilities of the Vienna University of Technology are spread over a wide area (mostly in the fourth district of Vienna) and especially, when you are new to the campus regardless of whether you are a student, an employee or a visitor, a map application that shows the location of places of interest and make it possible to search them may save some time and reduce annoyance.
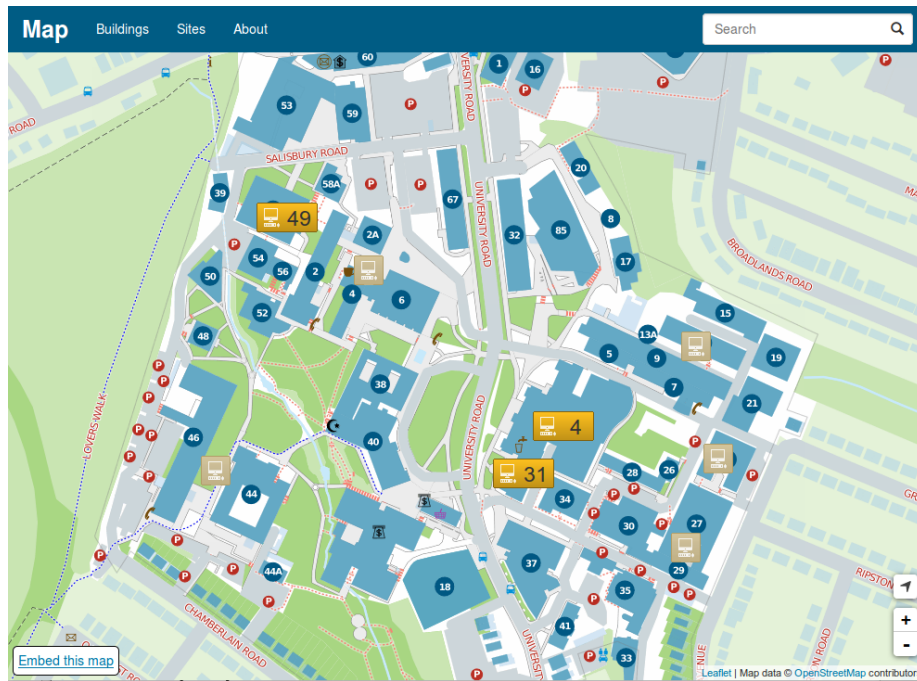
Figure 1: Map application of the University of Southampton

Not only to find them in the first place, but also for finding free learning or computer rooms as well as for the social life beyond the university; finding cafés or restaurants that are nearby.

The University of Southampton has developed a map application[12] based on their Linked Open Data initiative (see figure 3.3.1). This map application provides a detailed view of the location of the buildings and the internal structure with the ability to switch floors. The application make it possible for the user to get more detailed information like where single rooms or printers are located. Especially, the computer rooms are highlighted and the opening hours as well as an estimated occupancy rate are displayed per request. This sort of application use case of Linked Open Data seem to be very popular and is also implemented among others at the University of Münster[13] and Oxford University[14].

The Vienna University of Technology has already published the location of the lecture rooms, seminar rooms and plans of the internal structure of the buildings in an open way, but not in a format, where the content can be interpreted by machines. Additional to this, there are other sources with valuable information for such a map. TUbarrierfrei manages knowledge about the accessibility of buildings and certain rooms, Fachschaft (student representatives) provides

---

[12]http://maps.southampton.ac.uk/
[13]http://app.uni-muenster.de/Karte/
[14]https://data.ox.ac.uk/explore/science-area/

information about learning rooms and ZID about computer rooms. The project called LinkedGeoData[15] provides geographical information (e.g. cafés, restaurants and super markets that are nearby) as Linked Open Data. At the time of writing, these datasets cannot be interlinked and presented to the user in a map application.

Results: All interviewees responded positively to this proposed application use case and would use such an application. Interviewee *E* would like to have an overview of all scheduled events(exams, lectures, exercises) that take place in a certain room at a particular date. This idea corresponds to the demand of all other interviewees to find free rooms for particular tasks. *A* and *D* would like to have a service, where all free rooms that can potentially be used for meet-ups and learning like free lecture rooms are indicated on such a map. *B* and *C* would like to have a service to simplify the finding of free seminar rooms and rooms in general that can be used by tutors to mentor their students. Regarding the provision of information about the accessibility in such a map, *B* mentioned that real-time data about disfunctional elevators may be quite useful. *C* would also like to have an overview of the location of institutes. This idea corresponds to the science area application of the Oxford University[16]. *E* wanted to make the note that such a map application shall also work offline.

Challenges: The information required for such an application is separated over different data sources including diverging data owners, but this phenomena is not unique for this use case and challenges as well as an approach to overcome this is further discussed in section 4.

3.3.2 Resource management list

Description: The Vienna University of Technology consists of eight faculties, which cover the classical areas of natural and technical science and overall, there are 52 different institutes; each of them with further branches and a range of courses. TISS is the central source for actual course information over the WWW, but some of these courses have their own information system (mostly simple websites or e-learning platforms), where special details of the course are published like for example the slides of the lectures, the bibliography of learning resources, dates of events (lectures, exams, ...) or exercise sheets.

When a student wants to get an overview of the suggested resources (books, journal articles, web pages and other forms of media) for attainig the learning goals of the course, the student may have to scan web pages and lecture slides for collecting them. Furthermore the student has to search for a supplier for each resource that is not freely accessible over the WWW, like most books and journals. Resource lists management tools try to simplify this procedure.

---

[15] http://linkedgeodata.org/About
[16] https://data.ox.ac.uk/explore/science-area/

Figure 2: Reading list for 'Exploring the International 1', Aberystwyth University

> On the subject of the user experience for the student, RLM tools have often been no more than electronic replicas of the paper based list. There has traditionally been little attempt to allow students to annotate items with regard to their intended use, or interact with content. Students could not feedback whether they have found items useful in attaining their learning goals, or add personal study notes. There has been no facility to allow the students themselves to form new collections or bibliographies of resources for a particular essay or group exercise.[10]

Resource lists, that are simply exposed on web pages or in lecture slides, are not a very convenient approach from the students perspective. Resource lists management tools can take Linked Open Data principles into consideration and link the resources to the suppliers (and especially to the local academic library, if possible). The students can give feedback to a resource list of the lecturer and create their own resource lists, which can on their part be rated by other students. Students can also comment respectively mark helpful sections of resources.

Results: **B** and **C** were quite sceptical about this use case and their concerns are further discussed in the challenge section. **A**, **D** and **E** would like to have such an application.**E** said, "*Yes. It is a pain to stay up-to-date on organizational lecture information, which is often distributed over TISS[17], TUWEL[18], the institutute homepage, etc. ...*". This comment goes beyond this use case and concerns organizational information in general. Such a resource list management tool could be part of a broader set of services that simplify the organizational burden of students. **D** mentioned that (s)he was multiple times on the edge to send explored useful resources that helped him/her a lot to attain the learning goals of the course per email to the lecturer. **D** also suggested a ranking like on stackoverflow[19], where contributions can be liked or disliked and in consequence

---

[17]central information system of the Vienna University of Technology

[18]e-learning platform, a customization of moodle

[19]http://stackoverflow.com/

helpful contributions are usually at the top. **E** suggested to make it possible to link to useful contributions in the Informatikforum[20] and experience reports on VOWI[21].

Challenges: **B** and **C** mentioned that the additional administrative burden for course teams like keeping the list up-to-date may diminish the willingness of using such an application.**B** pointed out, that the use of the e-learning platform TUWEL is also not accepted throughout all course teams. While the contribution of the course team would be quite valuable, the resource list management tool could allow the students to create resource lists completely on their own. However, this approach may cause situations where solutions to exercises are contributed as well as spam, if the course team is not into it. Same as with the map application use case, the information required for such an application is separated over different data sources including diverging data owners. Challenges as well as an approach to overcome this is further discussed in section 4.

3.3.3   Explored use cases

One essential part of the interview was the exploration of the interviewee's ideas regarding needful use cases and datasets for which a more open access is requested. This section represents the results.

Course calendar: As mentioned earlier and also by some interviewees, the organizational information for courses may be separated over multiple sources. There is the central information system TISS, where basic information for mostly all courses is exposed. Especially for courses with exercises, also the e-learning platform TUWEL is often used simultaneously, where all the deadlines for the exercises are exposed and sometime more detailed information about the lecture times. And yet other courses have their own platform or own simple web pages, where more detailed organizational information is provided. In order to get an overview of all dates, a student has to collect this information manually, which is not a convenient approach. On the one hand this procedure is quite time-consuming and on the other hand prone to failures. For **D** the calendar shall have following functions.

Range of covered events: The calendar shall not only contain the lecture times of favoured courses, but also other important events like dates of course/exercise-group/exam registrations or exercise deadlines.

Semester planning: The calendar shall be able to detect intersections of events (e.g. the lecture time of one course intersect with the lecture time of another) and give a precise warning.

---

[20]internet forum for students of informatics
[21]a wiki, where students share experience and knowledge about courses

Synchronization: There shall be the ability to synchronize the course calendar with the own calendar.

TISS already provides a calendar with the dates of all favoured courses, but the calendar is restricted to the lecture times exposed on TISS.

Course information: As mentioned above, the organizational information about courses may be separated over multiple sources. Firstly, there should be a possibility to present all available organizational information about a certain course to a student. Additionally, there are also internet forums and wikis operated from students that collect information (mostly experience) about certain courses. *D* and *E* wanted to have the ability to interlink the course information on TISS with other sources like VOWI (wiki operated from students). *D* also would like to have the ability to filter courses by categories. For example, when a student is interested in artificial intelligence and (s)he want to get all courses of this field.

Event calendar: Same as with course information, the information about events that are going to take place at the university are separated over different sources in dependence on which organizational unit organizes the event. *D* would like to have a calendar for all events. It shall be possible to filter the events by categories/organizational units. This idea is similar to the event calendar of the University of Southampton[22]. At the moment of writing, the events are often announced to the students per email, flyers or by word-of-mouth recommendations including social networks. It is likely to oversee events in which you are potentially interested in. *D* would also like to have an event calendar that can be synchronized with the own calendar using the mentioned filter.

3.4 Adoption potential

The interviewees were also asked about their willingness to develop applications based on Linked Open Data provided by the university.

A is willing to develop applications based on Linked Open Data, if a comprehensive documentation is provided for application developers and would in this case see a playground for other students of informatics. *A* is a student representative and manages some valuable information for the students. *A* pointed out that a provision of some of this information as Linked Open Data may cause a disproportionate amount of work and maintenance burden, when questioned about it, but is open to it.

Student *D* would be willed to develop applications based on provided Linked Open Data, only if some incentives like rewards for innovative applications were provided, but could imagine that some of his/her student colleagues may develop

---

[22]http://www.events.soton.ac.uk/

small applications for their own joy. Student **E** would be willed to, if (s)he has a creative idea.

B saw his/her personal interest in other areas, but could imagine that students of courses that lecture software engineering, where a practical project has to be carried out (mentioned course 'Advanced Software Engineering' at the Vienna University of Technology), may design an application based on Linked Open Data provided by the university. **C** was principally interested in writing such an application.

Overall, the interviewees were more or less cautious regarding this question.

4   Technical architecture and challenges

Today's universities usually have to manage a significant amount of information and additional one is produced day by day. Due to the complexity of the domain of a university, there may be multiple services that handle a particular part of the domain. For a university that commits both to science and science education, the domain includes areas like the management of an academic library or the provision of an education respectively research platform. Furthermore, larger universities tend to spread their services across several departments to serve their stakeholders and satisfy their specific requirements. This phenomena of spreading services can not only be observed at the Vienna University of Technology, but also at the Open University[23] (Fouad Zablith, Mathieu d'Aquin, Stuart Brown and Liam Green-Hughes, 2011)[11].

It is likely for such an environment to result in independent services, which encapsulate their collected information into disconnected data sources with different formats and/or data owners. As a result of this, the information that has been collected by all these multiple services can not easily be interlinked, which blocks the full potential. A simple example for such two isolated services at the Vienna University of Technology are the e-learning platform called TUWEL[24], which is a customization of moodle[25], and the central information system called TISS. Both of these services provide course information and material, but they are intended for different purposes. TISS focuses mainly on administrative functionality like course or exam registrations, whereas TUWEL focuses on a broader interaction between teacher and student. Providing for example an additional functionality, which synchronizes important dates from deadlines of exercises to the time of lectures and exams with your own calendar might be quite useful, but the development is costly, due to the fact that the information is separated over this different isolated data sources and at that point not freely accessible.

Section 4.2 proposes a technical architecture using a Linked Data approach to overcome this environment of data silos and to support the evolvement of a university-wide data space, which can be queried. In fact this university-

---

[23]http://www.open.ac.uk/
[24]https://tuwel.tuwien.ac.at/
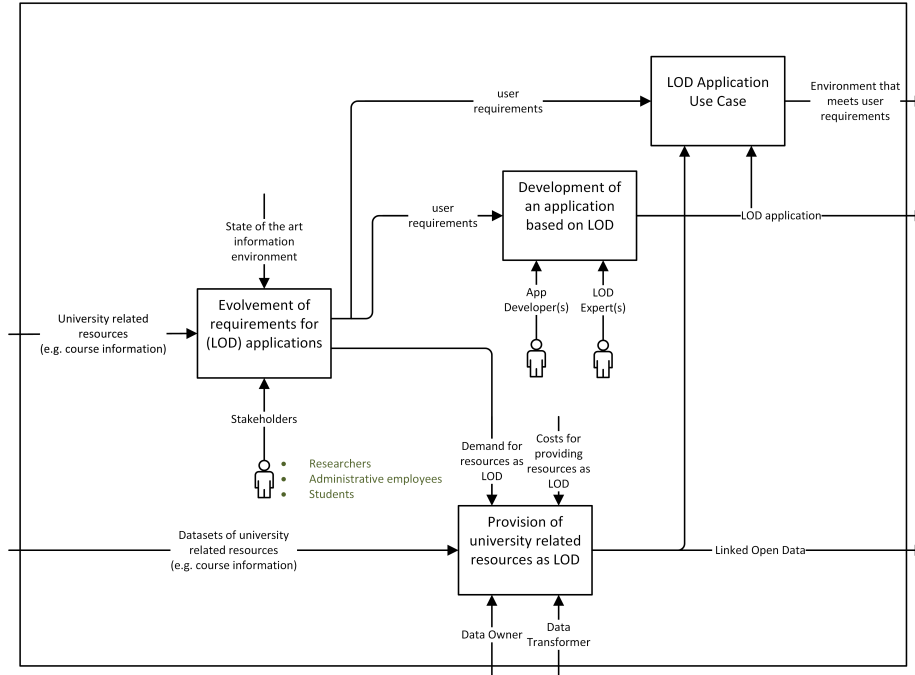[25]https://moodle.org/

Figure 3: IDEF0: The evolvement of a LOD effort at a university

wide data space can be interlinked with the LOD of other entities like DBPedia or LinkedGeoData to enrich it. Section 4.1 describes the stakeholder driven evolvement of a Linked Open Data effort at a university. The last section discusses the challenges (see 4.3).

4.1 Linked Open Data effort at universities

This section discusses our model (see figure 3) for describing the evolvement of a Linked Open Data effort at a university driven by the stakeholders. The model visualizes the issue on an organizational level in IDEF0 notation.

The evolvement of a requirement for a LOD application is driven by the needs of the *stakeholders* and the current information environment at a university and around this university. The requirement for a Linked Data application is rising, when the needs of a stakeholder are not sufficiently met by the information environment. Such a rising demand for a LOD application must not only be caused by an unsatisfactory situation at the university, but also by external developments like Linked Science, Open Education or the Semantic Web in general. The evolvement of such a requirement for a LOD application leads to user requirements, which shall be met by the application, and the

demand for providing university resources, which will be required for such an application, as LOD.

- *Stakeholders*
  As stated in the introduction (see 1), we identified students, administration employees and researchers as important *stakeholders* for a Linked Open Data effort at a university.

The provision of a university resource as LOD is controlled by the demand on publishing the known datasets of this resource as LOD and the given costs for the process. If the demand is low and the costs are high, the provision is pointless from an economic point of view. This process includes the *transformation* of the datasets into Linked Data as well as the exposure of the transformed Linked Data in an open way. The transformation process can be tricky. Especially, if the datasets consists of information that change dynamically, it takes extra thought to keep the data fresh. Also *data owners* of the datasets can diverge or be unknown in advance. An issue, which is not that obvious, is the problem of copyright licenses, if your approach is data mining on public web pages of the university. This three issues are further discussed in the section of technical challenges (see 4.3).

- *Data owner*
  The owner of a dataset is the organization or person that is responsible for the dataset and manages it. The exploration of the data owner is important for transforming the dataset into Linked Open Data, because the owner can authorize access to it and has specific expertise about it.

- *Data transformer*
  Dependent on the format of the dataset, the data transformer can either be a program or a semi-automatic approach, where a human prepares the dataset for a program. A complete automatic approach is possible, if the datasets are available in a machine-processable format and if the content can be interpreted (e.g. CSV, XML). Whereas a semi-automatic approach is required, if the content of the dataset can not be interpreted (e.g. images, PDF or hidden in the text) or if the dataset is not available in a machine-processable format (e.g. human expertise). Then a human has to interpret and store the result in a appropriate format so that a automatic approach is applicable. The data sources and the corresponding transformation techniques are further discussed in Section 4.2.

The development of LOD applications can emerge at a university, if potential initiators see a requirement for their idea of an application (meets user requirements that are not satisfied by the current information environment) and if the datasets of the resources required for such an application have already been exposed or they are going to be exposed in near future. The evolvement of an environment, where new LOD applications are initiated, assumes incentives

for the *development teams.* At a university with researchers and students in the field of informatics as potential *developers*, joy and acknowledgement may be an incentive, beside the monetization of the work like through university aid, donors or advertisements. This particular issue was part of the interviews with the students and the results can be seen in section 3.4.

- *Application developer*
  The application developer has the essential expertise for building software applications that meet the predefined specifications, which are influenced by the user requirements. The developer can also be a stakeholder. Computer science researchers can for example program solutions to simplify some work tasks for themselves or the whole research community at the university.

- *Linked Data expert*
  An expert, who has expertise about Linked Data and the supporting frameworks of the used programming language, is inevitable, when developing an application based on Linked Data. Whereas the idea of Linked Data is around for while now, common application developers may not be familiar with this technology to an extent that they can develop a reliable application based on it. The level of the required expertise depends on how the exposed Linked Data shall be used. If the application uses the exposed Linked Data only for querying data, the level is low and an application developer can learn it with spending some extra time. If the application itself shall use the Linked Data approach to store data (e.g. triple store) and/or do more advanced operations on Linked Data like link it with other (linked) datasets, manipulate it or reason over it, a higher level is needed and it is not done by spending some extra time. In this case inviting an expert to the development team, is a better approach.

A LOD application use case is the interaction of users with a developed application that makes use of LOD provided by the university. The application use case is the attempt to satisfy user requirements and becomes part of the information environment at the university. The exploration of interesting use cases from the student perspective was an essential part of the conducted interviews and the results can be seen in section 3.3.

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. Students of a university are usually on the consuming side of information systems. However, there are students that are operating information systems like wikis and internet forums. Furthermore, some course teams have their own web pages, where detailed information about the course is provided. In earlier studies of Linked Data applied to the domain of universities little attention has been
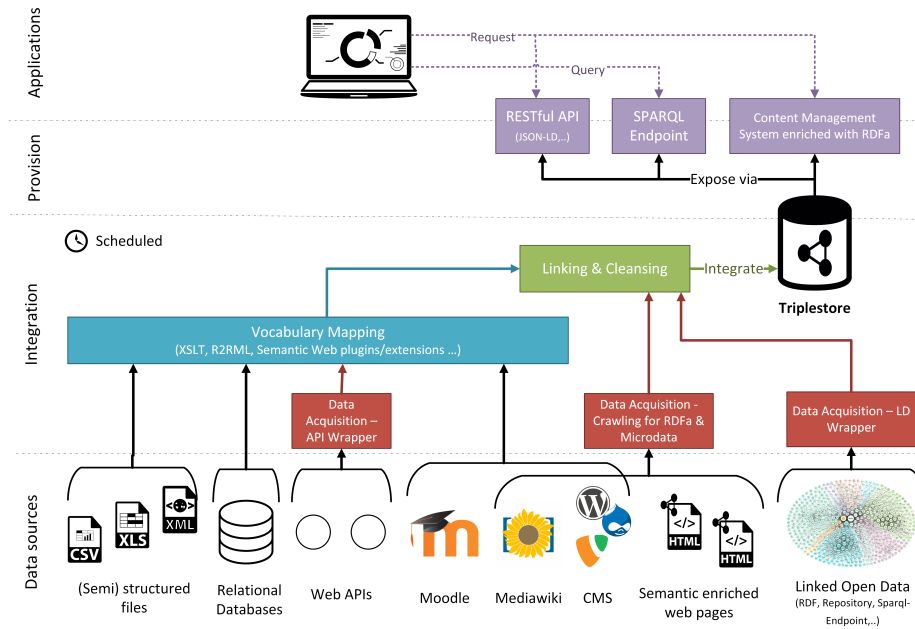
Figure 4: High level architecture for providing linked open data

paid to this issue. The proposed technical architecture in contrast suggests a solution.

The technical architecture is composed of three tiers. The first tier represents the diverging forms of **data sources** that can occur in an information environment at a university (see 4.2.1). The second tier shows the **integration** of such data sources into a central data storage for linked data also called triple store (see 4.2.2). The third tier shows the **provision** of the stored linked data so that developers can implement application based on it (see 4.2.3).

4.2.1 Data sources

As mentioned earlier, the information environment of a university may consist of multiple independent services that evolved over time powered by data sources that do not met the principles of Linked Data like relational databases. Nevertheless, there are techniques to integrate these diverging data sources into a Linked Data environment without reconstructing legacy services and work-flows, which is described in the **integration** tier. In this section, the different forms of data sources that may occur in the information environment of a university are listed and described.

Not machine-processable: Not machine-processable sources of knowledge like the expertise of a human (e.g. which rooms are in which way accessible)

or digital data that is hard to interpret (e.g. images, PDF files) have to be transformed by a human into data formats that can be processed and interpreted by computers in order to be managed by the proposed technical architecture.

Semi-structured files

Relational databases

Web APIs are one way to publish the underlying data of a service in a structured format so that it can be processed by applications. It is common to provide results in formats such as XML and JSON. At the Vienna University of Technology, the central information system TISS has a RESTful API, which gives access to some data like course information or the address book.

According to the star rating system of Tim Berners-Lee, Web APIs achieve a rating of $\star\star\star$, if the API uses standardized structured formats like XML and JSON. However, this Web APIs have some limitations in the perspective of Linked Data.

> Data returned from Web APIs typically exists as isolated fragments, lacking reliable onward links signposting the way to related data. Therefore, while Web APIs make data accessible on the Web, they do not place it truly in the Web, making it linkable and therefore discoverable. [12]

...

...

Semantic enriched web pages

4.2.2 Integration

...

4.2.3 Provision

...

4.3 Challenges

The technical challenges.

Data freshness

Data ownership ...

License (of content that has been publicly exposed, sharing possible?)

  ...

Comfort zone of the stakeholders

  ...

5 Conclusion

The conclusion goes here.

6 Acknowledgement

A Questionaire

This questionnaire is intended to gather new ideas about useful use cases and applications as well as to figure out datasets that may be interesting for publishing as Linked Open Data. The target group of this interview guideline are students, student's representatives and interest groups. The whole interview is planned to last not more than 40 minutes.

A.1 About me

At that point, where I wrote this interview guideline, this is a short description of myself. My Name is Kevin, I am a student of the Vienna University of Technology and I am attending the bachelor program Software and Information Engineering. I am working on a scientific work about Linked Open Data and the possible advantages of publishing the education related data of the university as Linked Open Data so that the datasets can easily be interlinked with other datasets to enrich them. The focus of this work are useful use cases, applications as well as potential interesting datasets from the student's perspective.

A.2 Introduction to Linked Open Data

In the situation, in which we are now, a lot of information/data is published on web pages in a human readable format. So if we want to search for special information, the default procedure may be the following. Let's say, we want to get information about the learning rooms of the Vienna University of Technology and the way it can be entered accessibly. Then we will use our favourite search engine, scan web pages and follow links to gather the required pieces of information to join it and thereby answering the initial question of our interest. This information cannot be found on a single web page (e.g. Fachschaft Informatik[26] (student representatives), TUbarrierefrei[27] , Wegweiser[28]). For humans this procedure is no problem, but think of computers. If you want to write an application, that shall answer exactly this initial question, you may have to face a lot of problems and there may be no sufficient way to gather this information. At that point, there is generally no efficient way, that computers/machines understand natural language and furthermore interlink the information about several web pages.

And here Linked (Open) Data enters the game.

> Linked Data describes a method of publishing structured data so that it can be interlinked and become more useful. It builds upon standard Web technologies, but rather than using them to serve web pages for human readers, it extends them to share information in a way that can be read automatically by computers. This enables data from different sources to be connected and queried [13].

---

[26]https://fsinf.at/lernen
[27]https://www.tuwien.ac.at/dle/barrierefrei/willkommen/
[28]http://www.wegweiser.ac.at/tuwien/

Linked Open Data extends Linked Data with the idea of publishing the data in an open way, so that anyone can freely access, use, modify, and share for any purpose.

To come back to the previous example, the information was separated over different sources. Beyond that, the information provided by this sources was only published in a human readable format. To take the idea of Linked Open Data into consideration, the information can also be provided in a structured format so that it is machine-readable. E.g. the Fachschaft Informatik can provide the list of learning rooms as Linked Open Data and TUbarrierefrei can share their information about the accessibility of the rooms respectively the buildings as Linked Open Data. Now an application designer can use both datasets and interlink them to answer the initial question of finding an accessible learning room. Furthermore the designer can interlink these datasets with other (external) datasets like Open Street Maps [29] or the list of computer rooms of the ZID [30]. The result of this may be a map of interesting places for students, employees and visitors.

As not to hang on a single application or use case, one of the incitements of publishing the information as Linked Open Data is the hope, that there is an audience, that come up with creative ideas, of which you and I are currently not thinking about.

A.3   Proposal of use cases

A.3.1   Map of places of interest

The buildings of the campus of the Vienna University of Technology are spread over a wide area. Especially, if you are new to the campus regardless of whether you are a student, an employee or a visitor, a map application of the places of interest would be very helpful. Not only to get stress less and punctual to lectures, meetings or conferences, also for the social life afterwards – finding cafes or restaurants that are nearby as well as computer rooms or learning rooms for using the gap time between lectures.

The University of Southampton has developed a map application for their campus including the location of computer rooms (see live[31] and figure 3.3.1) as well as the University of Münster (see live[32] or figure 5). Both universities make use of their Linked Open Data projects.

The Vienna University of Technology has already published the location of the lecture rooms, seminar rooms and the plans of the buildings in an open way on Wegweiser, but not in a machine-readable format. As already mentioned in the introduction to Linked Open Data, there are different sources of interesting information regarding the campus of the university. TUbarrierfrei hosts information for the accessibility, the Fachschaft provides information about the

---

[29]http://www.openstreetmap.org/
[30]http://www.zid.tuwien.ac.at/en/student/internet_raeume/
[31]http://maps.southampton.ac.uk/
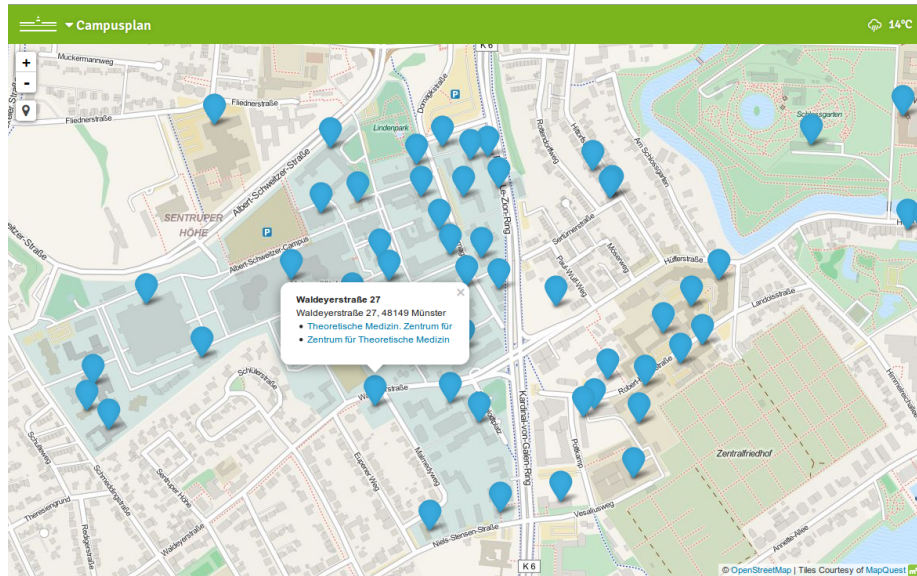[32]http://app.uni-muenster.de/Karte/

Figure 5: Map application of the University of Münster

learning rooms, the ZID about the computer rooms. The project called Linked-GeoData[33] provides geographical information (e.g. cafés, restaurants and super markets that are nearby), which has been collected by the community on OpenStreetMaps, as Linked Open Data. In the situation, in which we are now, these datasets cannot be interlinked although an application that offer this information to us in a useable way may be valuable for students and the university itself.

A.3.2 Course information / Resource management

The Vienna University of Technology consists of eight faculties, which cover the classical areas of natural and technical science and overall, there are 52 different institutes - each of them with further branches and a range of courses. TISS is the central source for actual course information over the WWW, but some of these courses have their own information system (mostly simple websites or e-learning platforms), where special details of the course are published like for example the slides of the lectures, the bibliography of learning resources, dates of events (lectures, exams, ...) or exercise sheets.

When a student wants to get an overview of the proposed resources (books, journal articles, web pages and other forms of media) for studying the content of a course, the student may have to scan web pages and lecture slides for collecting them. Furthermore the student has to search for a supplier for each resource that

---

[33]http://linkedgeodata.org/About

is not freely accessible over the WWW, like most books and journals. Resource lists management tools try to simplify this procedure.

> On the subject of the user experience for the student, RLM tools have often been no more than electronic replicas of the paper based list. There has traditionally been little attempt to allow students to annotate items with regard to their intended use, or interact with content. Students could not feedback whether they have found items useful in attaining their learning goals, or add personal study notes. There has been no facility to allow the students themselves to form new collections or bibliographies of resources for a particular essay or group exercise.[10]

Resource lists, that are simply exposed on web pages or in lecture slides, are not a very convenient approach from the students perspective. Resource lists management tools can take Linked Open Data principles into consideration and link the resources to the suppliers (and especially to the local academic library, if possible). The students can give feedback to the resource list of the lecturer and create their own resource lists, which can on their part be rated by other students. Students can also comment respectively mark helpful sections of resources.

Talis[34] is such an application (see figure 2). It is used by various universities of the United Kingdom among others.

A.4   About the interviewee

The questions of this section are intended to get a deeper understanding of the background of the interviewee regarding Linked Open Data, information technology in general and the role inside of the university.

What is your area of responsibility? How would you characterize your daily work tasks?

_____

_____

How would you classify your level of expertise in the field of information and communication systems?
    ☐   Fundamental
    ☐   Novice

---

[34]https://talis.com/

- ☐ Intermediate
- ☐ Advanced
- ☐ Expert

How would you classify your level of expertise in the field of Linked Open Data?

- ☐ I never heard of Linked Open Data.
- ☐ I heard of Linked Open Data, but never used it.
- ☐ I sued Linked Open Data in a not intense way. E.g. as part of a workshop or home project.
- ☐ I used Linked Open Data in a practical project.
- ☐ I used Linked Open Data in several practical projects and consider myself an expert in Linked Open Data.

A.5   Places of interest and resource list management

Do you have the desire for a map application like the proposed one in the introduction? Do you see a general desire for such an application?

_____

_____

Can you imagine some further datasets, additional to the ones mentioned in the introduction, that you want to interlink in such a map application? If yes, I am pleased to hear them.

_____

_____

Do you have a desire for such a resource management application like the proposed one in the introduction? Do you see a general desire for such an application?

_____

_____

A.6 Datasets

Which datasets/information that could be public do you want to access more efficient and freely to deskill/beautify your course of the day at the Vienna University of Technology?

_____

_____

Do you know such datasets that may be potentially interesting for your colleagues or other organizations/interest groups inside of the Vienna University of Technology?

_____

_____

If there will be a Linked Open Data initiative at Vienna University of Technology, would you be willed to publish your datasets that may be potentially interesting for the community as Linked Open Data? If there will not be an initiative, would you still be willed to?

_____

_____

A.7 Linked Open Data use cases / applications

Do you have the desire for a concrete application (that may be feasible with Linked Open Data) for your daily work at the university or the social life around the university? If yes, I am pleased to hear from it.

_____

_____

Can you imagine to use Linked Open Data that is provided by the Vienna University of Technology to develop new applications? Do you see a playground for students (your colleagues)? Which playground do you see?

_____

_____

References

[1] M. Janssen, Y. Charalabidis, and A. Zuiderwijk, "Benefits, adoption barriers and myths of open data and open government," *Information Systems Management*, vol. 29, no. 4, pp. 258–268, 2012.

[2] I. Jacobs and N. Walsh, "Architecture of the World Wide Web, Volume One," W3C, W3C:rec, Dec. 2004, http://www.w3.org/TR/2004/REC-webarch-20041215/. [Online]. Available: http://www.w3.org/TR/2004/REC-webarch-20041215/

[3] T. Berners-Lee, "Linked Data - Design Issues," *W3C*, no. 09/20, 2006. [Online]. Available: http://www.w3.org/DesignIssues/LinkedData.html

[4] M. d'Aquin, C. Kessler, and T. Kauppinen, *Members of Linked Universities*, 2014, [accessed 8-February-2016]. [Online]. Available: http://linkeduniversities.org/lu/index.php/members/index.html

[5] ——, *Linked Universities*, 2014, [accessed 24-February-2016]. [Online]. Available: http://linkeduniversities.org/index.html

[6] M. d'Aquin, F. Zablith, E. Motta, O. Stephens, S. Brown, S. Elahi, and R. Nurse, *LUCERO - Aims, Objectives and Final Outputs of the Project*, 2010, [accessed 29-February-2016]. [Online]. Available: http://lucero-project.info/lb/2010/06/lucero-aims-objectives-and-final-outputs-of-the-project/index.html

[7] L. Baronyai, "Publishing TUWien Data as Linked Open Data," 2016.

[8] S. Gamerith, "Linked Open Data in Administration at the Vienna University of Technology," 2016.

[9] P. Runeson and M. Höst, "Guidelines for conducting and reporting case study research in software engineering," vol. Empir Software Eng (2009), no. 14, pp. 131–164, 2008.

[10] C. Clarke and F. Greig, "Resource List Management Tool for Undergraduate Students based on Linked Open Data Principles." in *The Semantic Web: Research and Applications*, 2009.

[11] F. Zablith, M. d'Aquin, S. Brown, and L. Green-Hughes, "Consuming Linked Data within a Large Educational Organization," Bonn, Germany, 2011.

[12] T. Heath and C. Bizer, *Linked Data: Evolving the Web into a Global Data Space (1st edition)*. Morgan & Claypool, 2011, vol. Synthesis Lectures on the Semantic Web: Theory and Technology.

[13] C. Bizer, T. Heath, and T. Berners-Lee, "Linked Data - The Story So Far," *International Journal on Semantic Web and Information Systems*, 2009.