A case study on Linked Data for University Courses

Barnabás Szász[1], Rita Fleiner[2], András Micsik[3]

[1] University of Debrecen, Debrecen, Hungary
bszasz@gmail.com
[2] Óbuda University, Budapest, Hungary
fleiner.rita@nik.uni-obuda.hu
[3] MTA SZTAKI, Budapest, Hungary
micsik@sztaki.mta.hu

Abstract. Óbuda University wanted to build a linked dataset describing their courses in the semester. The concepts to be covered included curricula, subjects, courses, semesters and educators. A particular use case needed the description of lecture rooms and events as well. Although there are several ontologies for the mentioned domains, selecting a set of ontologies fitting our use case was not an easy task. After realizing the problems, we created the Ontology for Linked Open University Data (OLOUD) to fill in the gaps between re-used ontologies. OLOUD acts as a glue for a selection of existing ontologies, and thus enables us to formulate SPARQL queries for a wide range of practical questions of university students. OLOUD integrates data from several sources and provides personal timetables, navigation and other types of help for students and lecturers.

Keywords: Ontology · Linked Open Data · Linked Open University Data · SPARQL

1 Introduction

In this paper we focus on a special segment of open data at the university domain: university courses. We aim to facilitate the implementation of Smart Universities [1] by defining a common data model for course information. Ontological representation as the most modern description method for the problem domain was chosen. Originally our objective was to develop a generic data model for university course related data. During our work we noticed that though the Bologna Process ensures a certain level of compatibility for education systems in the EU, this does not reach deeper constructs regarding the educational model. We found that the meaning of the main concepts (like course, subject and study programme) is quite different in currently available educational models in Europe.

Presenting course related information requires a lot of data originating from multiple information systems at a typical university. As these systems are usually not fully integrated and the access to the data is limited, significant effort is necessary to suc-

adfa, p. 1, 2011.

cessfully navigate through the potential difficulties. Foreign students, who are not aware of the local specialties can find it even more cumbersome. With our data model we would like to support the generation and the management of integrated university data and also the appearance of future mobile and web applications building on the use of this data.

In Section 2 use cases are explored for the planned course dataset. Section 3 is about existing work related to our goals. Section 4 clarifies the terminology and describes how our new "glue" ontology was built and how it re-uses other existing ontologies. Finally, we summarize results in Section 5.

2 Use Case

In the following we list some of the tasks we aim to support with our Linked Data approach.

Courses. Courses are organized into a series of lectures and lab exercises, either in a weekly or in a custom cadence within a semester. There might be multiple labs advertised for a course, so students can choose the most suitable to their circumstances. This creates the challenge of assembling a personal timetable for students and lecturers avoiding conflicts and considering personal preferences and requirements.

Students could benefit from an integrated view containing course description (title, identifier, abstract, dependencies), course time, location, and learning resources. A personal information service may provide students with on-demand information about their daily schedule, navigation to the next lecture, overlaps of classes, etc. We also consider accessible way-finding to course locations.

Curriculum fulfillment. There is a need for long term planning of studies as well. Quite often there are no predefined course timetables at Hungarian universities, just a list of courses to be completed, and a dependency graph among the courses, which defines the prerequisites for each. Some courses are advertised in every second semester only. Some universities recommend a specific order of courses, but following such an order breaks easily if for example a single course is not completed in the suggested semester. Thus, students face a kind of constraint satisfaction problem to solve at each semester start.

For this purpose, students need a personal advisor recommending the best way for them to fulfill the curriculum requirements. This advisor needs to consider where the student is on his roadmap, what courses they should focus on, what are the personal preferences (e.g. preferred number of courses or credits per semester) and what courses are being advertised.

Resource reservation. University resources (rooms, equipment) are used by multiple faculties. They can be booked for regular courses, exams in the exam period and other events. Different types of events may have separate registries, thus blocking an overall view of anticipated resource usages. One needs an overall list of reservations by reserving person, location and date at least.

3 Related Work

Linked Universities[1] and Linked Education[2] are two European initiatives created to enable education with the power of Linked Data. Linked Universities is an alliance of European universities engaged into exposing their public data as linked data. LinkedEducation.org is an open platform aimed at further promoting the use of Linked Data for educational purposes. However, we did not find any existing solution or vocabulary at these websites fulfilling our needs for the data model.

The Open University in the UK was the first university that created a linked data platform to expose information from its departments [2]. The Open University datasets can be classified in the following six groups: open educational resources, scientific production, social media, organizational data, research project output, publication metadata. The main difference with Óbuda University is the lack of navigation and timetable data at the Open University.

The general process for building linked open university data and a use case at Tsinghua University are described in [3]. Procedures like choosing datasets and vocabularies, collecting and processing data, converting data into RDF and interlinking datasets are studied. The datasets unfortunately are not available through public SPARQL endpoint.

The Lucero project analyzed open educational datasets in 2012 [4]. Linked Open Datasets in four universities and four broader educational projects were studied and the most commonly used vocabularies, classes and properties were described. In this case no representations for course, semester or lecture room concepts were found. The general state of linked data for education is studied in [5] containing statistics on vocabulary re-use. We found another very useful review of ontologies for modelling course information in higher education [6].

AIISO (Academic Institution Internal Structure Ontology) [7] provides classes and properties to describe the structure of an academic institution. It is designed to be used in conjunction with the Participation ontology [8] which stands for describing the roles that people play within groups.

TEACH [9] is a lightweight vocabulary providing detailed properties in order to describe a course, but it doesn't model the provider of the course. The concepts in TEACH lack some important features that are essential for our purposes. For example, in order to describe university courses the concept of Subject is necessary, which doesn't exist in TEACH.

XCRI-CAP [10] is the abbreviation for eXchanging Course Related Information, Course Advertising Profile. It is the UK standard for describing course marketing information. XRI-CAP doesn't make a distinction between a module, a course, a subject or a study program.

The Metadata for Learning Opportunities (MLO) Advertising ontology [11] aims to standardize the specifications for describing and exchanging information about courses and learning opportunities. It can be considered the European equivalent of the British Standard XCRI-CAP for advertising courses. MLO-Adv contains very abstract and general concepts and misses many properties for courses and curricula.

[1] http://linkeduniversities.org

[2] http://linkededucation.org

After reviewing the above ontologies, we realized that there are many ontologies defining Course, Subject, Curriculum and Degree, but neither provides a full coverage and there are missing relations between the concepts. We concluded that the AIISO ontology provides the best coverage, and its structure fits our concrete use case, thus it can be reused and extended with the necessary terms.

Table 1. Possible mapping of the basic terms

AIISO	**Teach**	**XCRI-CAP**	**MLO-Adv**
Programme	Study Program	Course	Learning Opportunity Specification
Subject	Course		Learning Opportunity Specification
Course	Course		Learning Opportunity Instance

4 Implementation

The major concepts for university students and teachers in our use case are depicted in Figure 1. **Table 1** demonstrates how different ontologies use different labels for more or less similar concepts. For this reason a short summary about how these terms are interpreted in our use case is given in the following.

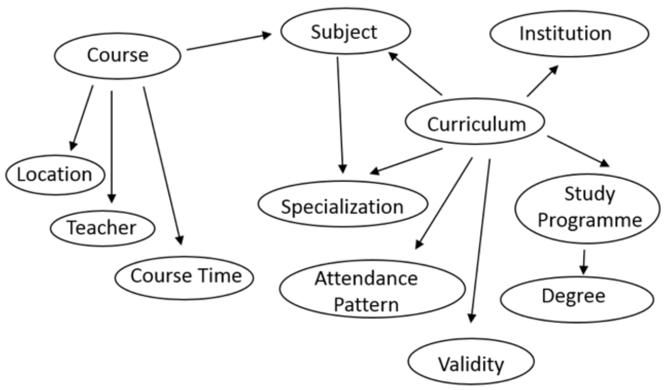

Fig. 1. Main concepts for our use cases

After university enrollment each student is assigned to a **Curriculum**, which is a set of **Subjects** and their relations showing the dependencies among the subjects. A curriculum might specify **Specializations** which consist of optional subjects. A Curriculum stands in many-to-one relationship with a **Study Programme** (e.g. Applied Mathematician), offered by the university. Each Curriculum has a specific **Attendance Pattern** (full-time, part-time, correspondence, etc.). A Study Programme determines the qualification and a specific **Degree** (BSc, MSc, BA, MA, PhD, etc.) that students will get after the successful completion of their studies. A Study Programme must be accredited by an external body. The curriculum is the specification how the Study Programme can be completed. A Curriculum is valid for a given time interval, meaning that a student can be assigned to it only if his enrollment time falls into this period of time. For each Subject there is an **Organizational Unit** responsible for it. **Courses** are advertised based on a Subject, have temporal (**Course Time**) and spatial (**Location**) attributes and one or more assigned **Teacher**(s). Additionally, learning materials (**Learning Resource**) can be assigned to specific courses and subjects.

It is important to understand the difference between Subjects and Courses, which constitutes the basic units of the teaching process. Course is the elementary unit of the educational process, where students, teachers, location and date are assigned. Subject is a higher level component of the training process, it represents a part of knowledge students need to acquire.

As seen in Section 3, we had to extend and connect existing ontologies for our domain model. The result is the Ontology for Linked Open University Data (OLOUD) presented in the rest of the section. Figure 2 depicts the main classes as ovals and the most significant attributes as arrows in OLOUD. The prefixes for classes and properties clearly show what has been reused and where we had to fill missing links. We selected to reuse the following ontologies: AIISO for the description of university faculties, courses, subjects and roles; FOAF for the description of persons; OWL Time for temporal definitions; Event for describing events and Dublin Core for generic properties such as language.

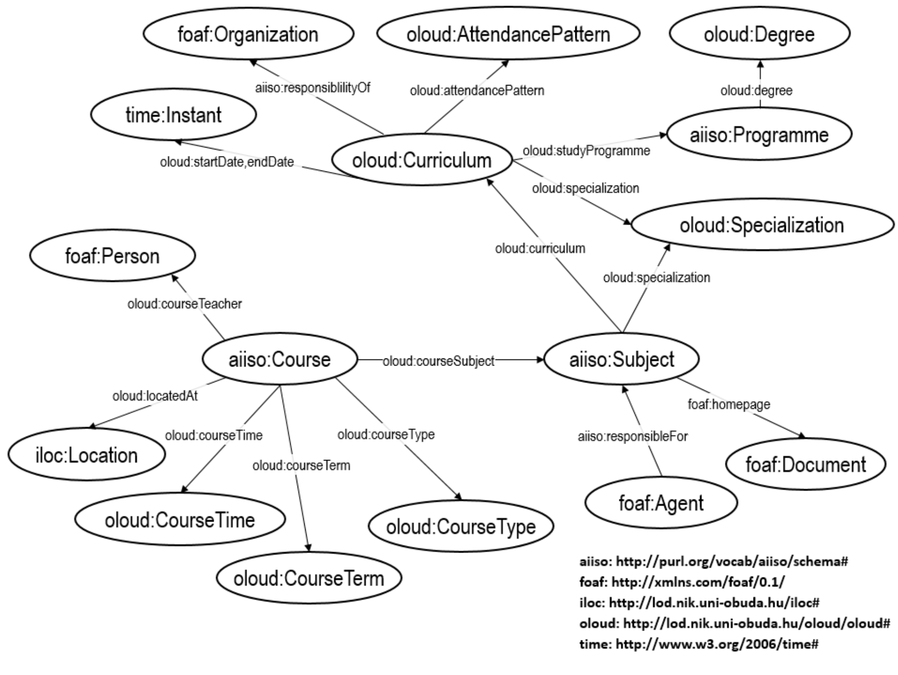

Fig. 2. Main classes and properties in OLOUD

Curriculum, Subject and *Course* are the most important classes in OLOUD. In the following the description of the classes are given with their direct connections. Curricula contain the list of subjects with their dependencies, since each subject can have various prerequisite subjects. The faculty of the university responsible for the Curriculum is given by the *aiiso:responsibilityOf* property. S*tartDate* and *endDate* properties determine the period, when a Curriculum is valid. The possible Specializations, the Degree, the Attendance pattern and the Study programme of the study are given by the corresponding *specialization, degree, attendancePattern and studyProgramme* properties. The language of the studies is given with the *dcterms:language* property.

Subjects are described by their name, code, credit number, person and organization responsible for it: *foaf:name, aiiso:code, subjectCredit, aiiso:responsibilityOf*. The connection between a Subject and its Courses is given by the *courseSubject* attribute. The prerequisite conditions between Subject entities are set by the *subjectRequires* property. In case the subject has its own webpage, it can be provided using *foaf:homepage*.

The individuals of the Course class are the actual instances of subjects running in a given semester (*courseTerm*) having spatial, temporal and type descriptions, identification number, name and instructor: *locatedAt, courseTime, courseTerm, courseType, aiiso:code, foaf:name* and *courseTeacher*.

New classes are introduced for the full description of the Curriculum, Subject and Course individuals including *StudyProgramme, Degree, AttendancePattern, Specialization, CourseTerm* and *CourseType*.

Entities describing indoor locations for Course and Event individuals are represented by the *Location* class from the iLOC[3] ontology. Courses and events can be assigned to Rooms, and Rooms are connected via a network of POIs (Points of Interest), which can be doors, hallway connections, etc. The offices of lecturers can also be included in the description of campus buildings.

OWL Time [12] and the TimeAggregates Ontology [13] are used to express temporal descriptions of courses. Our objective was to enable SPARQL queries according to date, time and duration and to define course time as recurring events (e.g. labs on every Tuesday from 9 am till 10.30 am in the 2016 Fall semester). The triples representing such information need the introduction of several additional individuals, hence the management of such information is time consuming and error prone. The generation of course time data was implemented according to an automatized process described in [14] and complete examples can be found on the web[4].

5 Results

The development tool – Protégé – was used for validation purposes as it revealed problems in the imported ontologies and in our ontology as well. The OOPS! Ontology Pitfall Scanner was also used to check our OWL [15], and the found small problems were fixed.

---

[3] http://lod.nik.uni-obuda.hu/iloc/iloc.html

[4] http://lod.nik.uni-obuda.hu/unfolding/example.html

Linked data triplets were created based on public data[5] at the Óbuda University. The location data was created manually based on building layout diagrams of the university, while the subject and course data were converted using custom PHP scripts from relational database dumps extracted from the electronic administration system (Neptun) of the Óbuda University. The entities of the Person class and their personal data were scraped from the personal webpages. The university event descriptions were generated with a crawler from the OU webpage. The dataset was also extended with links to the GeoNames geographic dataset. At this moment, the database contains about 1000 entities with more than 6000 triples.

As an example, we provide a sample Course and Subject description:

```
:AB0_LA_01_E_2014-15-1 rdf:type aiiso:Course,
aiiso:code "AB0_LA_01_E"@hu ;
foaf:name "Adatbázisok"@hu ;
oloud:courseSubject :NAIAB0SAED ;
oloud:locatedAt :PC_Labor_220 ;
oloud:courseTeacher :Dominika_Fleiner ;
oloud:courseTime
odata:CourseTime;courseTerm=2014Fall;hour=17;minute=55;durationHour=1;durati
onMinute=35;dayofweek=1 ;
oloud:courseType :Lab ;
oloud:courseTerm :2014Fall .

:NAIAB0SAED rdf:type aiiso:Subject,
oloud:subjectCredit "4.00"^^xsd:Integer;
foaf:homepage "http://users.nik.uni-obuda.hu/to/tantargy/adatbazisok-0" ;
aiiso:code "NAIAB0SAED" ;
foaf:name "Adatbázisok"@hu ;
aiiso:responsibilityOf :AII ;
oloud:studyProgramme :ComputerEngineer ;
aiiso:responsibilityOf :Dominika_Fleiner ;
oloud:attendancePattern :FullTime ;
oloud:specialization :ObligatoryPart ;
oloud:degree :bsc;
oloud:curriculum :MérnökInformatikusBSc_2012 .
```

Figure 3 provides a visualization of a part of the linked data graph. Browsing the dataset with a LOD tool such as LODmilla [16] provides access to useful integrated data for students. Furthermore, an in-door way finding service for students is also available, providing a simple itinerary to the place of the given lecture. It is available as a mobile app, and backed by an open SPARQL endpoint.

We tested the complex use cases with implementing different SPARQL queries answering the questions in Section 2. For example, one can ask about the courses for a given subject in a specific semester:

```
PREFIX oloud:http://lod.nik.uni-obuda.hu/oloud/oloud#
SELECT DISTINCT ?course WHERE {
?course a <http://purl.org/vocab/aiiso/schema#Course>.
?course oloud:courseSubject <http://lod.nik.uni-obuda.hu/data/NAIAB0SAND>.
?course oloud:courseTerm <http://lod.nik.uni-obuda.hu/data/2014Fall>. }
```

[5] http://lod.nik.uni-obuda.hu/marmotta/

What is the course schedule (with course identifier, time and lecturer) for a specific lecture hall or lab?

```
PREFIX oloud: <http://lod.nik.uni-obuda.hu/oloud/oloud#>
PREFIX ta: <http://ontology.ihmc.us/temporalAggregates.owl#>
PREFIX time: <http://www.w3.org/2006/time#>
PREFIX iloc: <http://lod.nik.uni-obuda.hu/iloc/iloc#>

SELECT DISTINCT ?course ?room ?day ?beginhour ?beginminute ?teacher WHERE {
  ?room a iloc:Room.
  ?course oloud:locatedAt ?room;
          oloud:courseTeacher ?teacher;
          oloud:courseTime ?ct .
  ?ct ta:hasTemporalAggregateDescription ?tad .
  ?tad ta:hasithTemporalUnit ?day ;
       ta:hasStart ?start .
  ?start time:hasDurationDescription ?dd ;
         time:hasBeginning ?begin .
  ?dd time:hours ?durationhour ;
      time:minutes ?durationminute .
  ?begin time:inDateTime ?begindatetime .
  ?begindatetime time:hour ?beginhour ;
                 time:minute ?beginminute .
}
```

Further sample SPARQL queries and more information about the dataset and the ontology can be found on the ontology homepage[6].

On our LOD server we currently serve the dataset using Marmotta[7], but we also experiment with Virtuoso. In the future we wish to proceed with various application developments using the generated LOD dataset. For example, a ‘curriculum assistant’ mobile application helping students to select their courses at the start of the semester might be useful.

[6] http://lod.nik.uni-obuda.hu/oloud/

[7] http://marmotta.apache.org/

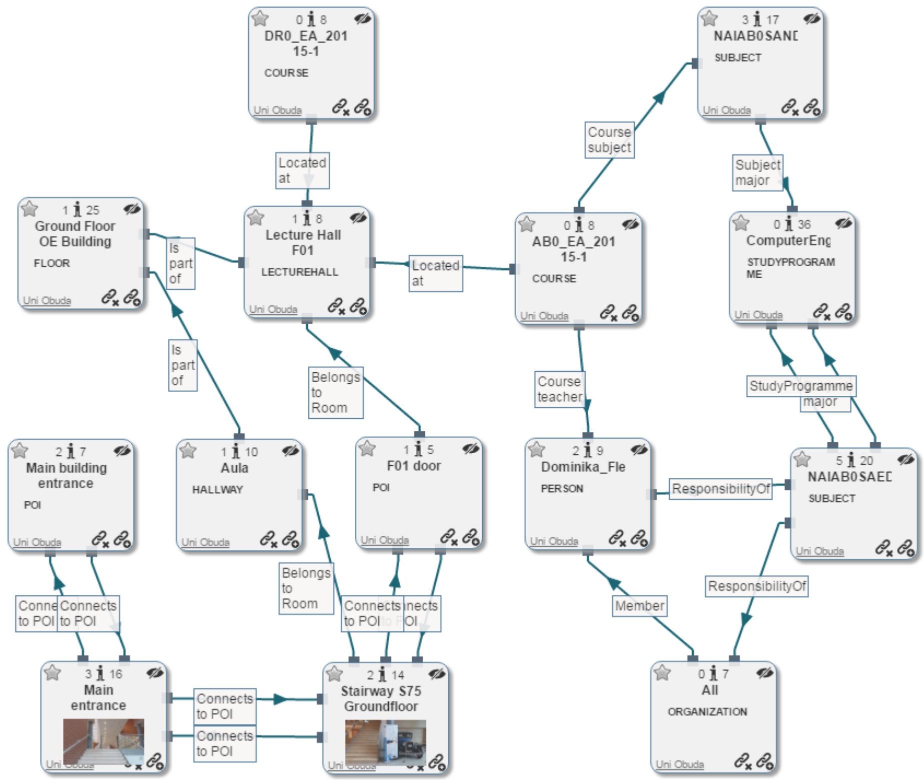

Fig. 3. Extract of the dataset in the Lodmilla browser

6 Conclusions

The starting point of our work was to implement useful, "smart" services for university students based on linked data. We realized that there are many ontologies or vocabularies for the domain, yet none of them is suitable for our purpose. The biggest problems we found were the missing discriminations between subjects and courses and curricula and study programs. We created the OLOUD ontology which amalgamates selected ontologies and facilitates the full description of course-related information. Hence, we think our work helps to better clarify the role of concepts frequently used in this domain.

The OLOUD ontology provides the basis of several ongoing student projects, which either integrate new datasets for the university or implement new services on top of the OLOUD dataset. In the future our plan is to collect all knowledge as linked data that is practical for the daily life at Óbuda University.

References

1. Rohs, M., & Bohn, J.: Entry points into a smart campus environment-overview of the ETHOC system. In 23rd International Conference on Distributed Computing Systems, Workshops Proceedings, pp. 260-266, IEEE (2003)
2. Daga, E., d'Aquin, M., Adamou, A., Brown, S.: The Open University Linked Data - data.open.ac.uk. Semantic Web Journal, 7(2) pp. 183–191. (2015)
3. Ma, Y., Xu, B., Bai, Y., Li, Z.: Building linked open university data: Tsinghua university open data as a showcase. In: The Semantic Web, pp. 385-393., Springer Berlin Heidelberg (2011)
4. The Lucero project: So, what's in linked datasets for education (2012) http://lucero-project.info/lb/2012/04/so-whats-in-linked-datasets-for-education/
5. d'Aquin, M., Adamou, A., Dietze, S.: Assessing the educational linked data landscape. In: Proceedings of the 5th Annual ACM Web Science Conference, pp. 43-46. ACM (2013)
6. Barker, P.: A short project on linking course data (2015) http://blogs.pjjk.net/phil/a-short-project-on-linking-course-data/
7. Styles, R., Shabir, N.: Academic institution internal structure ontology (aiiso). (2008) http://vocab.org/aiiso/schema
8. Styles, R., Wallace, C., Moeller, K.: Participation ontology. (2008) http://vocab.org/participation/schema
9. Kauppinen, T., Trame, J., Westermann, A.: Teaching Core Vocabulary Specification (TEACH ontology). (2012) http://linkedscience.org/teach/ns/
10. Stubbs, M., Wilson, S.: eXchanging Course-Related Information: a UK service-oriented approach. (2006) http://hdl.handle.net/1820/837
11. EN 15982:2011 Metadata for Learning Opportunities (MLO)-Advertising.
12. OWL Time Ontology. https://www.w3.org/TR/owl-time/
13. Pan, F.: Temporal aggregates for Web services on the semantic Web. IEEE International Conference on Web Services (ICWS'05), pp. 831-832. IEEE (2005)
14. Szasz, B., Fleiner, R., Micsik, A.: Linked data enrichment with self-unfolding URIs. In 2016 IEEE 14th International Symposium on Applied Machine Intelligence and Informatics (SAMI) pp. 305-309. IEEE (2016)
15. Poveda-Villalón, M., Suárez-Figueroa, M. C., & Gómez- Pérez, A.: Validating ontologies with OOPS!. In Knowledge Engineering and Knowledge Management, pp. 267-281. Springer Berlin Heidelberg (2012)
16. Micsik, A., Turbucz, S., Györök, A.: LODmilla: a Linked Data Browser for All. In: Proceedings of the Posters and Demos Track of 10th International Conference on Semantic Systems - SEMANTiCS2014, pp. 31-34. CEUR Workshop Proceedings (2014)