

OLOUD - An Ontology for Linked Open University Data

Rita Fleiner¹, Barnabás Szász², András Micsik³

¹Department of Applied Informatics, John von Neumann Faculty of Informatics, Óbuda University, Bécsi út 96/b, 1034 Budapest, Hungary, fleiner.rita@nik.uni-obuda.hu

²Faculty of Informatics, University of Debrecen, Kassai út 26, 4028 Debrecen, Hungary

³Institute for Computer Science and Control, Hungarian Academy of Sciences, Lágymányosi u. 11, 1111 Budapest, Hungary, andras.micsik@sztaki.mta.hu

Abstract: The Ontology for Linked Open University Data (OLOUD) is a practical approach to model course information at a typical Hungarian university. OLOUD aims to integrate data from several sources and provide personal timetables, navigation and other types of help for students and lecturers. The modeled domains include curricula, subjects, courses, semesters and personnel, but also buildings and events. Although there are several ontologies for the mentioned domains, selecting a set of ontologies fitting our use case was not an easy task. We summarize problems we met such as missing links, inconsistencies as well as many overlaps between ontologies. Finally, 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.

Keywords: Linked Open Data; Linked Open University Data; Ontology; OWL

1 Introduction

One aspect of the Smart University or Smart Campus concept is to improve the teaching and learning environment using modern data fusion and data consumption techniques. A campus has a large number of people with a substantial set of common information needs [1]. Therefore, it is of great importance in this area to establish a common data model which enables the interconnection of fragmented data from heterogeneous data sources.

In this paper, we focus on university course information as a special segment of the open data in the higher education domain: The aim is to facilitate the implementation of Smart Universities by defining a common data model for

course information. We chose the ontological representation as the most modern description method for the problem domain. The ontology can be used as a data model to influence or integrate traditional SQL databases, as an RDF schema in triple stores, and in reasoners or rule systems to enhance collected data. The original objective was to develop a generic data model for university ‘course related’ data. During the 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 of the educational model. We found that especially the meanings of course, subject and study programme are quite different in currently available educational models in Europe. Therefore, we decided to base our work on the Hungarian concepts in this field. During the ontology development, we relied on the existing concept definitions and data structures used by university information systems in Hungary such as the Neptun student information system¹ or the Moodle e-learning platform².

The aim of this paper is to describe the process of developing the Ontology for Linked Open University Data³ and to show how it was used to generate Linked Data for courses at the Óbuda University. In order to achieve our goal, we defined the following partial objectives:

- to review the existing ontologies in this field and evaluate their possible usage,
- to reveal the use cases and objectives of the new ontology and define the basic concepts in the domain,
- to present the ontology development process,
- to describe the generation of university ‘course related’ data as LOD according to the ontology,
- to demonstrate the possible use of the above ontology and data by presenting SPARQL queries according to the questions in the use cases.

In Section 2 potential use cases are explored for the possible uses of our ontology. In Section 3 we explore existing work related to our goals, ontologies that can be used to model university ‘course related’ open data and identify gaps in existing ontologies. Section 4 describes our new ontology and how it re-uses other existing ontologies. Section 5 is about the evaluation and dataset generation. Finally, we conclude in Section 6.

¹<https://neptun.uni-obuda.hu/>

²<https://elearning.uni-obuda.hu/>

³<http://lod.nik.uni-obuda.hu/oloud/>

2 Use Cases, Objectives and Concepts

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 successfully navigate through the potential obstacles. Foreign students unfamiliar with local specifics might find it even more challenging. With our data model, we want to support the generation and the management of integrated university ‘course related’ data and the appearance of future mobile and web applications that are built on the use of this data. In the following, the major concepts and tasks are defined that we aim to support with our Linked Data approach.

2.1 Concepts in Hungarian Higher Education

Figure 1 summarizes the major concepts for university students and teachers in Hungary concerning university courses, subjects, curricula and study programme. Though the Bologna Process ensures a certain level of compatibility for education systems in the EU, this does not reach deeper constructs of the educational model. We found that the meaning of these concepts varies in currently available models.

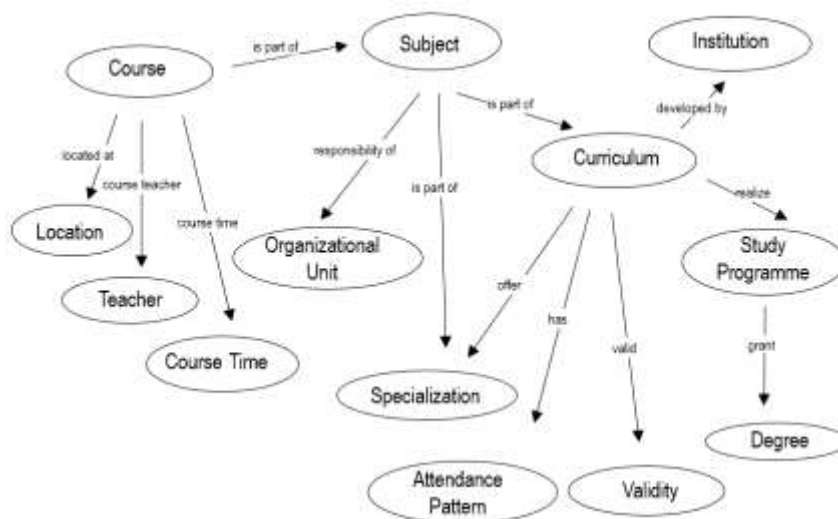


Figure 1
Use case concepts

Table 1 demonstrates how different ontologies use different labels for more or less similar concepts. Therefore, we give a short summary of how these terms are interpreted in Hungarian education. After enrolling to the university each student is assigned to a **Curriculum** (in Hungarian “tanterv”), which is a set of **Subjects** (in Hungarian “tantárgy”) and their relations (i.e. dependencies among the subjects). Curriculum might specify **Specializations** (in Hungarian “specializáció”), which are sets of compulsory and optional subjects. Each Curriculum has a specific **Attendance Pattern** (in Hungarian “munkarend”, e.g. full-time, part-time, correspondence), and a result as a specific **Degree** (in Hungarian “fokozat”, e.g. BSc, MSc, BA, MA, PhD). A Curriculum is in many-to-one relationship with a **Study Programme** (in Hungarian “szak”, e.g. Computer Science Engineer), offered by the university. The curriculum is the specification how the Study Programme can be completed. A Study Programme determines the qualification that a student will get after the successful completion of his/her studies. A Study Programme must be accredited by an external body. A Curriculum is valid for a given time interval, meaning that a student can be assigned to it only if his/her enrollment time falls into this period. For each Subject there is an **Organizational Unit** responsible for it. **Courses** (in Hungarian “kurzus”) are advertised based on a Subject, have temporal (**Course Time**) and spatial (**Location**) attributes and one or more assigned **Teacher(s)**.

Table 1
Possible mapping of core terms

AIISO	Teach	XCRI-CAP	Bowlogna	MLO-Adv
Programme	StudyProgram	Course	Study Track	Learning Opportunity Specification
Subject			Subject	Learning Opportunity Specification
Course	Course		Teaching Unit	Learning Opportunity Instance

In the case of Hungarian universities, it is crucial to understand the difference between Subjects and Courses. Course is the elementary unit of the educational process, where date, location, teacher and students are assigned. Course is the framework, which has a specific training type (like lecture, practice or seminar) and has some requirements that students must complete. Courses are organized in individual sessions in a weekly or a custom cadence within a semester, and such a course event is called the session of the course. Subject is a higher-level component of the training process, it is the unit of the curriculum with a specified training content, and fulfillment is rewarded with a number of credits. It may contain more courses, which all must be completed for the completion of a subject.

2.2 Use Cases

In the following we list some of the tasks we aim to support with our Linked Data approach. Courses are organized into a series of lectures, lab exercises or seminars, 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 would benefit from an integrated view containing course description (title, identifier, abstract, and dependencies), course time and location. A personal information service may provide students with on-demand information about their daily schedule, navigation to the next lecture, overlaps of classes, etc.

There is also a need for ‘long term’ planning. 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 only in every second semester. 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 the beginning of each semester. 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/her 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.

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 usage. One needs at least an overall list of reservations by the reserving person, location and date.

According to the above use cases a set of competency questions was defined as the requirement of the OLOUD ontology.

- What are the attributes of a given curriculum, like start of validity, end of validity, study programme, degree, language, attendance pattern?
- What type of specializations exist in each curriculum?
- What are the compulsory subjects of a given curriculum?
- What are the compulsory subjects of a given specialization?
- What are the attributes of a given subject, like responsible teacher, credit number, degree, language, attendance pattern?
- What type of courses consists of a given subject?
- What are the courses for a given subject in a specific semester?
- What are the prerequisite subjects of a given subject?

- What are the attributes of a given course, like the teacher, the course type, the requirement type and hours?
- On which day a given course is, what time does it start and how long it is?
- What is the location of a given course?
- What courses are at the same time partially or completely overlapping?
- What is the course schedule (with course identifier, time and lecturer) for a specific lecture hall or lab?
- What is the navigation route between two course locations?
- What are the dates of the individual sessions of a given course?
- What event is going to be at a specific location at a specific time?

In the following we use the above formulated questions as competency questions for the evaluation of the ontology.

3 Related Work and Ontologies

In this section, existing work is discussed related to linked data in the educational field. Linked Universities⁴ and Linked Education⁵ are two European initiatives created to enable education with the power of Linked Data. Linked Universities is an alliance of European universities engaged in exposing their public data as linked data. It promotes a set of vocabularies describing ‘academic related’ entities. LinkedEducation.org is an open platform aimed at further promoting the use of Linked Data for educational purposes.

The Open University in the UK was the first university that created a linked data platform to expose information from its departments. The evolution process of the Open University Linked Open Data platform is described in [2], [3]. This process started as a research experiment and evolved to a data hub for the open content of the university. The platform is now the key information service at the Open University, with several applications and websites exploiting linked data through data.open.ac.uk and establishing connections with other educational institutions and information providers. In the publications, the authors describe the main milestones and tasks accomplished to achieve this state. The Open University datasets can be classified into the following six groups: open educational resources, scientific production, social media, organizational data, research project output and publication metadata.

The main difference with Óbuda University is the lack of navigation and timetable data at the Open University. There are 125 classes and 785 properties from 57 public vocabularies to describe the data at the Open University. Their main effort

⁴<http://linkeduniversities.org/>

⁵<http://linkededucation.org>

in the modelling was to reuse the most matching terms from existing vocabularies directly instead of being restricted to the semantics of only a few widely-used ontologies. The large number of the used vocabularies, the redundancy in the data and in the used properties are the consequences of their approach.

The general process for building linked open university data and a use case at Tsinghua University are described in [4]. 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 [5]. 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 state of linked data for education is studied in [6]. They collect existing datasets explicitly related to the education field, extract key information, and analyze them. The goal is to better understand what is already available to application developers in this area, what common practices are being used and how the considered datasets connect with each other through common content and vocabulary reuse. They found 144 different vocabularies used in ‘education related’ datasets. The most popular vocabularies are not specific to education, but are used to represent general concepts and relations, such as resource metadata (Dublin Core [7]), people (FOAF [8]), topics (SKOS [9]), time (W3C Time Ontology [10]) and bibliography (BIBO [11]). More education specific vocabularies are also widely used, such as the Academic Institution Internal Structure Ontology (AIISO [12]), or the Model of Learning Opportunities (MLO [13]).

We found a very useful review of vocabularies and ontologies for modelling course information in higher education [14]. Parts of the following section were influenced by this work.

3.1 Ontologies for Education

The scope of this section is to review existing ontologies that are available to describe various aspects of educational courses and to evaluate whether they can be used to model ‘course related’ information in the Hungarian higher educational landscape. For simplicity, in the following we refer to ontologies, vocabularies and lighter schema constructs as ontology uniformly. Table 2 provides an overview about ontologies related to our use case and their description targets. There are big differences in the interpretation and in the elaboration of the used terms in the various ontologies.

Table 2
Coverage of relevant ontologies

Ontology	Course	Subject	Curricula, Study Prog.	Speciali- zation	Degree	Teacher	Organization	Time	Location
FOAF						✓	✓		
Vcard						✓	✓		
Event	✓							✓	
W3 Time								✓	
iLoc									✓
Aiiso +	✓	✓	✓	✓		✓	✓		
Teach	✓		✓	✓	✓	✓		✓	✓
XCRI			✓		✓		✓		
Course- Ware	✓								
VIVO	✓			✓		✓		✓	
MLO, ECIM	✓	✓			✓		✓	✓	
Bowlogna	✓		✓	✓	✓	✓	✓		

AIISO (Academic Institution Internal Structure Ontology) provides classes and properties to describe the structure of an academic institution. It is designed to be used in conjunction with the Participation ontology [15], which stands for describing the roles that people play within groups. Participation has only one class, but any domain can extend it by creating subclasses for their own roles within their areas of expertise. AIISO Roles [16] is an example for such extension; it describes roles that people play in an academic institution. The AIISO ontology proved to be useful in our work because it distinguishes at class level the Course and the Subject concepts. These classes are subclasses of KnowledgeGrouping. There is a note in AIISO that this class became deprecated. Probably it was a plan of the authors, but there is not any information on how and when this would be done. AIISO offers only a few properties to describe courses (i.e. code, description, teaches and responsibility) and is mainly used to connect subjects and courses with the organizational structure of the university.

In our work we experienced problems in reusing some properties from the AIISO ontology because they were defined as `rdf:Property` and in the automatic conversion to OWL the development tool Protégé⁶ decided wrongly to use `owl:ObjectProperty` instead of `owl:DatatypeProperty`. Another difficulty we faced was that although there exist classes like Programme and Module, their precise meaning is not defined, so their usage is ambiguous. Besides all this AIISO was chosen to form the basis of our data model because the structure of the concepts in this ontology fit into the Hungarian system the best.

⁶ <http://protege.stanford.edu/>

TEACH [17] is a lightweight vocabulary providing detailed properties to describe a course, but it does not model the provider of the course. The concepts in TEACH lack some important features that are essential for our purposes. For example, the concept ‘Subject’ is necessary to describe university courses, which does not exist in TEACH. Another problem was that one would expect an owl:DataTypeProperty based on the example data for TEACH, but the ontology itself declared the specific property (e.g. teach:courseDescription, teach:ects) as owl:ObjectProperty. These shortcomings were fixed, and the corrected version of the TEACH ontology (<http://lod.nik.uni-obuda.hu/teach-fixed.owl>) was used in the first phase of our work. In TEACH one can find further problems though. For example, the properties hasAssignment, hasAssignmentMaterial, and hasCourseMaterial have appeared in the index of terms of TEACH. However they are not defined in the vocabulary specification. The classes Student and Lecture are defined but there is no property available in the ontology to relate them to a Course. A potential disadvantage of the TEACH ontology is that it is not linked into other ontologies and some of the definitions are missing or do not have domains or ranges specified. Because of the above problems of this vocabulary, we decided not to use it.

XCRI-CAP [18] is the abbreviation for eXchanging Course Related Information, Course Advertising Profile. The term *course* in the UK is equivalent with the term *study programme* in Hungary, thus XRI-CAP does not contain the description about the course and subject in our terminology. XCRI-CAP is the UK standard for describing study programme marketing information. XCRI represents a lot of data about the provider and the programme and it also differentiates between a programme and the particular presentation of it. XCRI-CAP is in XML format and does not exist in RDF.

The ReSIST Courseware Ontology [19] is a simple ontology with only four classes and many properties like title, teacher, credits, prerequisites, assessment method, etc. It was developed within the ReSIST project between 2006 and 2009. It is an early ontology without any usage at present. The trouble with this ontology is that it is closely related to the Aktors ontology, which is no longer defined anywhere online.

The Metadata for Learning Opportunities (MLO) Advertising ontology is similar in a way to XCRI-CAP, because its purpose is to standardize the specifications for describing and exchanging information about learning opportunities. It can be considered the European equivalent of the British Standard XCRI-CAP for advertising learning opportunities. MLO-Adv contains the following four classes:

- Learning Opportunity Object: an abstract resource used within the context of education or training. It has the following three subclasses:
- Learning Opportunity Provider: a person or organization that offers the learning opportunities

- Learning Opportunity Specification: description of a learning opportunity, consisting of information that will be consistent across multiple instances of the learning opportunity.
- Learning Opportunity Instance: single occurrence of a learning opportunity, it might have a particular date or location.

MLO includes some properties from Dublin Core Elements such as contributor, date, description, identifier, subject, title, and type. ECIM [20] is an extension of MLO, which provides a common format for representing credits awarded for completion of a learning opportunity. XCRI, MLO and ECIM ontologies are similar in that they differentiate between a course specification and a course instance or course offering. The specification contains information about a course or a study programme that remains consistent from one presentation to the next, whereas the instance defines those aspects that vary between presentations for example location or start date. This has the advantage that there will be a smaller amount of data that needs to be updated between years and offerings.

The goal of the VIVO ontology [21] is to represent academic research communities, and thus it enables the discovery of researcher interests, activities, and accomplishments. In a later phase of our work VIVO may be useful to represent research groups within the university, including researchers' grants and external roles. Currently, its focus is quite different from the focus of OLOUD, for example VIVO has its own Course class, but its main properties are credits and prerequisites.

The Bowlogna ontology [22] describes terms used by the Bologna process. It can represent the departments, the teaching units together with information about their ECTS credits and teaching language. It can also be used to store students' examinations, their results and degrees. Although it aimed at providing a standard schema for European universities, in our modeling work we did not find any usage of it.

The main goal of this study was to reveal the usability of the above ontologies in our use case. The following general consequences were drawn:

- It is crucial to understand the meaning of the main concepts and the relationships among them in case of each ontology. Unfortunately, in most cases these ontologies use essential concepts without defining their meaning (i.e. what do concepts like course, subject, module, study programme exactly mean and how do they relate with each other?).
- A specific ontology is usable only if the definitions of the main concepts fit into the use case. Furthermore, in the decision process it is important to see what kind of implementation is used in case of a certain property or relationship (e.g. defining temporal information of a course can be achieved in various ways, but does the actual one satisfy our requirements?).
- The correctness of the formal description of the ontology is important. If it contains shortcomings and mistakes, its reuse is cumbersome.

- Currently there is no other ontology suitable for the use case scenarios described in the paper. Existing ontologies miss properties and thus cannot provide a full description of teaching activities. Furthermore, existing ontologies contradict each other in the naming and semantics of subject, course, curriculum, etc.

Basically, we had to find a set of ontologies filling all capability columns with a minimal number of overlaps. The selection criteria were also determined by several rational considerations, like availability, maintenance, usage and modularity. Before the final selection was made, we had to harmonize the term usage and adapt terms to the Hungarian system if it was possible at all. The final decision was to base our work on the AIISO ontology, because the structure of the concepts Course-Subject-Programme in this ontology fit into the Hungarian system the best.

4 Ontology Development

The main motivation of developing a new ontology born during the first trial implementation: In the domain of university courses there are existing ontologies and our initial approach was to build a linked open university dataset using these existing ones. It turned out that the existing ontologies in this field do not cover fully our requirements and some of them contain mistakes. Facts and connections related to the course-subject-curriculum concepts cannot be fully described by the existing ontologies, only very partially. This recognition confirmed the purpose of the new ontology development.

4.1 Methodology

The five-star model of good Linked Data vocabulary use [23] acted as a guideline during our work. Our aim was to design a 4-star vocabulary. The following rules were applied to restrict the potential interpretations of the defined classes and properties towards their intended meaning:

- Dereferenceable human readable information should exist about the ontology (e.g. a web page documenting it).
- The ontology should be described by a formal language, like OWL.
- The ontology should be linked to other ontologies.
- The ontology should contain metadata about itself (e.g. authors, modification date, used ontology language, status of the ontology terms, license information, etc.).

OLOUD was developed based on the Uschold and King methodology, which consists of the following steps [24], [25]:

1. Identify the objectives of the ontology development and the intended usage (see Section 2.2); determine the necessary formalization level (see Section 4.3).
2. Specify the ontology by outlining the domain. This includes the identification and the clear textual definition of key concepts and relations (see Section 2.1). Furthermore, setting up identifiers for concepts and relations is necessary.
3. Formalize the terms defined in the specification using a formal language (see Section 4.3).
4. Integrate with existing ontologies. During specification and formalization, it is an important step to research third party ontologies for potential reuse and inclusion (see Section 4.2).
5. Evaluate the fruition of the objectives and the completeness of the ontology based on a predefined (generic and ontology specific) criteria (see Section 5.2).
6. Specify the documentation principles, which should be aligned to type and objective of the ontology (see Section 4.3).

4.2 Integration with other Ontologies

In the process of creating a schema for a Linked Open Dataset it is advisable to reuse as much as possible of the available ontologies or vocabularies. There are quite different vocabulary reuse strategies [26]. The two basic forms are (1) reusing classes and properties from existing vocabularies directly, and (2) establishing links at schema-level. The second case means defining new classes as either sub-classes or equivalent classes and properties as sub-properties or equivalent properties of the classes and properties of the reused ontology. The reuse strategies can be influenced by various factors, like reuse only one (or a few) domain specific vocabulary to provide a clear data structure, or reuse only popular vocabularies to make the data easier to be consumed.

In case of the development of the OLOUD ontology our strategy was the following. First, the necessary concepts (classes and properties) were identified. Then, re-usable vocabularies which could serve to express the defined concepts were chosen according to criteria such as wide usage, OWL 2 compatibility, and regular maintenance. The study for vocabulary selection was detailed in Section 3, and it resulted in the choice to use AIISO as the basis and then use other ontologies to fill in the gaps. Ontologies that are not specific to education are used to represent general concepts and relations, such as resource metadata (Dublin Core), people (FOAF), time (W3C Time and Temporal Aggregates Ontology [27]), events (Event [28]), address (vCARD [29]) and indoor location (iLOC [30]). In case of necessary classes and properties missing from the previous ontology list, new OLOUD terms were introduced. If it was possible the new terms were linked on schema level to the above ontologies with the `rdfs:subPropertyOf` or `rdfs:subClassOf` properties.

Integration work posed the problem of fragmentation. In several cases an ontology was needed only for a single property (e.g. address). FOAF and vCard are similar ontologies, but each lacks some important properties, and thus both had to be used to fill in the holes. In the integration process, it was revealed that too many ontologies were needed to express desired goals, and some ontologies were hard to reuse because of the inaccuracy and mistakes in them.

4.3 Ontology Description

Figure 2 represents the overview of the new ontology: the main classes, the highlighted object properties connecting them and the essence of the class hierarchy as well.

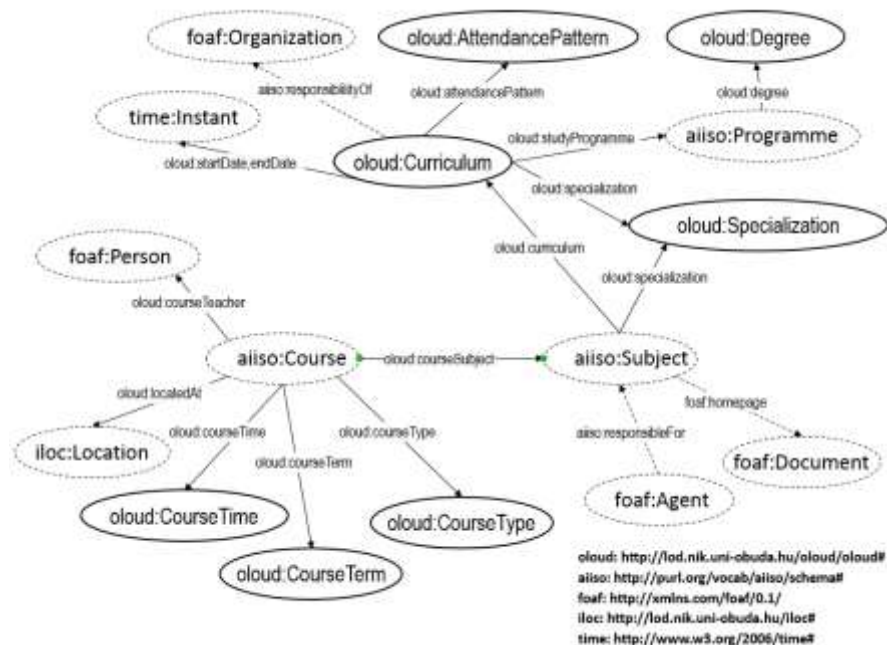


Figure 2

Overview of the main classes and properties in OLOUD

Classes defined in OLOUD are written with *cloud* prefix, classes and properties needed from other ontologies are written with their prefix and marked with a dashed line. The most important classes in OLOUD are *Curriculum*, *Subject*, *Course* and *Programme*. Curriculum class is defined as a subclass of the aiiso:KnowledgeGrouping class, Subject, Course and Programme classes are used directly from AIISO. In the following, these classes are described with their direct connections.

Curricula at Hungarian universities contain the list of subjects with their dependencies (i.e. each subject can have various prerequisite subjects). A specific Curriculum entity is connected to the Subject entities with the *curriculum* property. The faculty or department of the university responsible for the Curriculum is determined with the *aiiso:responsibilityOf* property. Each Curriculum has a period of validity determined by the *startDate* and *endDate* properties. The possible Specializations in each Curriculum are determined with the *specialization* property. The Attendance pattern, the Study programme and the Degree of the study are set by the *attendancePattern*, *studyProgramme* and *degree* properties. The language of the studies according to the Curriculum is given with the *dcterms:language* property.

Subjects are featured by their name, code, number of credits, person and organization responsible for it: *foaf:name*, *aiiso:code*, *subjectCredit*, *aiiso:responsibilityOf*. A Subject entity belongs to a specific Curriculum entity. The connection between a Subject and its Courses is set by the *courseSubject* property defined in OLOUD, since AIISO does not provide any property to connect these concepts. The prerequisite conditions between Subject entities are set by the *subjectRequires* property.

The entities of the Course class are the actual instances of subjects having spatial, temporal and type descriptions, identification number, name and instructor: *locatedAt*, *courseTime*, *courseTerm*, *courseType*, *aiiso:code*, *foaf:name* and *courseTeacher*. To describe entities of the Curriculum, Subject and Course classes properly some auxiliary classes were introduced: *StudyProgramme*, *Degree*, *AttendancePattern*, *Specialization*, *CourseTerm*, *CourseType*. The *aiiso:Programme* class is used to represent the Study Programme concept and *Specialization* is defined as a subclass of the *aiiso:Module*.

Location class from the iLOC ontology is used to represent all the necessary entities describing indoor locations for Course and Event entities. 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.

Entities providing temporal description of courses in OLOUD are based on OWL Time and Temporal Aggregates Ontologies. Our objective was to enable SPARQL queries according to date, time and duration and to define course time as recurring events. These objectives can be satisfied with the above ontologies. We suggest using a separate ontology module for the ‘time related’ concepts. In this module subclasses are defined for classes in OWL Time and Temporal Aggregates Ontologies, facilitating the generation of entities describing recurring events.

The OLOUD ontology consists of two modules: OLOUD-BASE [31] and OLOUD-TIME [32]. The former describes all the ‘university related’ concepts, uses the prefix *oloud* and namespace *http://lod.nik.uni-obuda.hu/oloud/oloud#*. The latter provides the necessary classes and properties to describe course time

data as recurring events, uses the prefix *otime* and namespace *http://lod.nik.uni-obuda.hu/oloud/otime#*. In OLOUD-BASE there are 7 classes, 16 object properties, 5 data properties and 14 individuals, while in OLOUD-TIME 6 classes are defined at this moment.

OWL 2 RL was chosen as the formal language for OLOUD. The advantage of this ontology approach is that new classifications can be inferred by rules and class restrictions, such as subjects announced for the current semester, subjects meeting the prerequisite criteria in case of a specific student or course announcements having various properties. The OLOUD Ontology was implemented in a self-documenting way. Based on the request MIME type it can be downloaded in different formats including the human consumable HTML output, which is automatically generated from the following properties: *rdfs:label*, *rdfs:comment*, *rdfs:domain*, *rdfs:range*. The ontology description was implemented as metadata best practice described in [33], by adding the recommended metadata instances and addressing the outlined policies. The Ontology is licensed under the terms of Creative Commons 3.0⁷

5 Ontology Evaluation and Dataset Generation

5.1 Evaluation

The role of the evaluation is to verify the fulfillment of the initial goals and the completeness of the ontology based on the predefined criteria. The evaluation was regularly carried on during the process of ontology development. The development tool – Protégé – was leveraged for validation purposes. The first phase of the evaluation was the definition of a class and the corresponding properties within Protégé. Inconsistency was discovered in several cases as Protégé was not able to create a property for the intended purpose. The root cause of these issues was mostly flaws in the imported third party ontologies. The continuous evaluation also included immediate trials of the new concepts, creating individuals and properties for these new individuals with the help of Protégé. During the ontology evaluation, it was inspected whether the original objectives and expectations were met. We tested the complex use cases with implementing different SPARQL queries answering the questions in section 2 [34]. The OOPS! Ontology Pitfall Scanner was also used to check our OWL [35]. The minor issues the scanner found were fixed.

⁷ <http://creativecommons.org/licenses/by/3.0/>

5.2 Triplification – the Óbuda University Use Case

To evaluate and validate the ontology LOD triples were created based on public data at the Óbuda University. The triples were also tested with Protégé for consistency. On our LOD server⁸ we currently serve the dataset using Marmotta⁹. The resulted linked data is organized into six graphs¹⁰ according to the type of the entities (i.e. subjects, courses, events, persons, location and others). At this moment, the database contains about 1000 entities with more than 6000 triples. Example data can be found in [34]. Triples of the different classes were derived from different sources, and the technique – the actual method the triples were created by – depended on the actual source. The location data was created manually based on building layout diagrams of Óbuda University, while the subject and course data was automatically generated with PHP scripts from relational database dumps extracted from the electronic administration system of Óbuda University. The university event descriptions were generated by scraping data from the university webpage.

Expressing ‘time related’ data of recurring events of course instances was not an easy task. There are multiple ways to model temporal information, but probably the most used ontology for this purpose is the OWL Time Ontology. It provides basic constructs to define and describe points and intervals bounded with a start and endpoint in the temporal space. OWL Time provides two approaches to describe a point of time: either using the `xsd:datetime` datatype or using the `DateTimeDescription` class. While the first one offers an easy way to define a point of time by a well-structured string, it lacks some of the features the `DateTimeDescription` class provides. On the other hand, manually modeling and maintaining `DateTimeDescription` entities are error prone and tiring because these require at least 7-8 triples in a format that is reusable in a semantic sense.

Temporal Aggregates ontology was used to express temporal information of courses as recurring events (e.g. lectures on every Monday from 8 am until 9.30 am in the 2015 Fall semester). The precise implementation of such information as Linked Data needs the introduction of several additional entities, hence the management of such information is time consuming and error prone. In case the given LOD dataset contains lot of temporal information, manually publishing all the necessary triples would be cumbersome. We used self-unfolding URI scheme for the time entity and an attached template to auto generate the required triples based on the information in the URI. The generation of course time data was implemented as an automatized process described in [36] using SPARQL Construct queries that can be executed in a scheduled manner.

⁸ <http://lod.nik.uni-obuda.hu/marmotta/>

⁹ <http://marmotta.apache.org>

¹⁰ <http://lod.nik.uni-obuda.hu/marmotta/core/admin/contexts.html>

Different namespaces were used for ontology concepts and for the generated data instances. The T-Box (terminology) of the Ontology is identified with the following URI schema: `http://lod.nik.uni-obuda.hu/oloud/oloud#{class or property name}` and the A-Box (instance data) is identified with `http://lod.nik.uni-obuda.hu/data/{instance_ID}` URIs. The structure of the instance_ID for the different classes were chosen according to the nature of the specific class. For example, the subject code (used by the university to refer to the subject) was a proper choice for the instance_ID of the Subject class, because it is unique among the different subjects. For the instances of the Course class the course code (used by the university) had to be complemented with the semester code, because the course code is unique only within a single semester.

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 too many ontologies or vocabularies for the domain, and none of them is suitable for our purpose. We created the OLOUD ontology, which amalgamates selected ontologies and fills the missing links between existing concepts.

During the ontology development, we relied on the existing concept definitions and data structures used in the Hungarian landscape such as the Neptun student information system or the Moodle e-learning platform. This was necessary because of the specialties of the Hungarian educational system (probably all national systems have smaller or bigger differences from others), and also because all previous ontologies for education were specific to some goals, and neither of them aimed at a holistic description of the domain.

In Hungary, the structure of the training programs in higher education is unified, the meanings of basic terms like courses, subjects, specializations, curricula are treated uniformly. This is not derived from acts of legislation, but from everyday practice, which is characterized by the overwhelming use of a specific electronic student administration system at universities in Hungary (called Neptun). Thus, most Hungarian universities follow the same data model (i.e. the ones standing behind Neptun) in their workflow.

Although there is a uniformly used electronic student administration system in Hungarian universities, the need for an ontology for university open data still exists. The reason is complex: (1) open university data exist in more sources (not only in Neptun), (2) the availability of public data from Neptun is cumbersome and it is far from the requirements of a 4-star dataset, (3) the reuse and the integration of open data from several data sources is difficult.

There is a need for a common understanding of the basic terms of the educational process. This can help foreign students to find the way in their university studies, and the interoperability between universities in different countries. 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 and transform to linked data all knowledge that is practical for the daily life at Óbuda University. Furthermore, 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. In the future, the OLOUD ontology can be extended with new features. For example, the need to list information about the ongoing and past research of the university might induce the extension of the OLOUD model.

References

- [1] Rohs, M., & Bohn, J. (2003, May) Entry Points into a Smart Campus Environment-Overview of the ETHOC System. In Distributed Computing Systems Workshops, 2003. Proceedings. 23rd International Conference on (pp. 260-266) IEEE
- [2] Daga, E., d’Aquin, M., Adamou, A., & Brown, S. (2015) The Open University Linked Data-data. *open. ac. uk. Semantic Web Journal*. <http://www.semantic-web-journal.net/content/open-university-linked-data-dataopenacuk>
- [3] d’Aquin, M., Brown, S. (2012) Linked Data at the Open University: From Technical Challenges to Organizational Innovation <http://www.slideshare.net/mdaquin/linked-data-at-the-open-university-from-technical-challenges-to-organizational-innovation>
- [4] Ma, Y., Xu, B., Bai, Y., & Li, Z. (2011, December) Building Linked Open University Data: Tsinghua University Open Data as a Showcase. In Joint International Semantic Technology Conference (pp. 385-393) Springer Berlin Heidelberg
- [5] 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/>
- [6] d’Aquin, M., Adamou, A., & Dietze, S. (2013, May) Assessing the Educational Linked Data Landscape. In Proceedings of the 5th Annual ACM Web Science Conference (pp. 43-46) ACM
- [7] DCMI Metadata Terms (2012) <http://purl.org/dc/terms/>
- [8] Brickley, D., & Miller, L. (2014) FOAF Vocabulary Specification 0.99. Namespace Document 14 January 2014-Paddington Edition
- [9] Miles, A., & Bechhofer, S. (2009) SKOS Simple Knowledge Organization System Reference. W3C recommendation, 18, W3C
- [10] Hobbs, J. R., & Pan, F. (2006) Time Ontology in OWL. W3C working draft, 27, 133

- [11] D’Arcus, B., & Giasson, F. (2009) Bibliographic Ontology Specification. URL: <http://bibliontology.com/specification>
- [12] Styles, R., & Shabir, N. (2008) Academic Institution Internal Structure Ontology (aiiso) <http://vocab.org/aiiso/schema>
- [13] WS-LT, C. E. N. (2008) Metadata for Learning Opportunities (MLO)–Advertising. In Workshop Agreement. CEN. <http://www.estandard.no/files/CWA15903-00-2008-Dec.pdf>
- [14] Barker, P. (2015) A Short Project on Linking Course Data <http://blogs.pjjk.net/phil/a-short-project-on-linking-course-data/>
- [15] Styles, R., Wallace, C., & Moeller, K. (2008) Participation Ontology <http://vocab.org/participation/schema>
- [16] Styles, R., Wallace, C. (2008) Academic Institution Internal Structure Ontology Roles (AIISO Roles) <http://vocab.org/aiiso-roles/schema>
- [17] Kauppinen, T., Trame, J., & Westermann, A. (2012) Teaching Core Vocabulary Specification (TEACH ontology) <http://linkedscience.org/teach/ns/>
- [18] Stubbs, M. (2006). XCRI| eXchanging Course-Related Information. Techn. rep., Manchester Metropolitan University, Aytoun Street, Manchester, M1 3GH
- [19] Rodriguez, B., Millard, I. (2006) ReSIST Courseware Ontology <http://courseware.rkbexplorer.com/ontologies/courseware>
- [20] Educational Credit Information Model (ECIM) (2010) <ftp://ftp.cen.eu/CEN/Sectors/TCandWorkshops/Workshops/CWA16077.pdf>
- [21] Börner, K., Conlon, M., Corson-Rikert, J., & Ding, Y. (2012) VIVO: A Semantic Approach to Scholarly Networking and Discovery. *Synthesis Lectures on the Semantic Web: Theory and Technology*, 7(1) 1-178
- [22] Demartini, G., Enchev, I., Gapany, J., & Cudré-Mauroux, P. (2013) The Bowlogna Ontology: Fostering Open Curricula and Agile Knowledge Bases for Europe's Higher Education Landscape. *Semantic Web*, 4(1) 53-63
- [23] Janowicz, K., Hitzler, P., Adams, B., Kolas, D., & Vardeman II, C. (2014) Five Stars of Linked Data Vocabulary Use. *Semantic Web*, 5(3) 173-176. <http://www.semantic-web-journal.net/content/five-stars-linked-data-vocabulary-use>
- [24] Uschold, M., and M. King. *Towards a Methodology for Building Ontologies*. IJCAI’95 Workshop on Basic Ontological Issues in Knowledge Sharing. Diss. Ed. D. Skuce, 1995

-
- [25] Uschold, M. Building Ontologies: Towards a Unified Methodology. Technical Report University of Edinburgh Artificial Intelligence Applications Intitute AIAI TR (1996)
- [26] Schaible, J., Gottron, T., & Scherp, A. (2014) Survey on Common Strategies of Vocabulary Reuse in Linked Open Data Modeling. In *The Semantic Web: Trends and Challenges* (pp. 457-472) Springer International Publishing
- [27] Pan, F. (2005) Temporal Aggregates for Web Services on the Semantic Web. *IEEE International Conference on Web Services (ICWS'05)* DOI=<http://dx.doi.org/10.1109/ICWS.2005.118>
- [28] Raimond, Y., & Abdallah, S. (2007) The Event Ontology. Technical report, 2007, <http://motools.sourceforge.net/event>
- [29] Iannella, R., & McKinney, J. (2013) vCard Ontology for Describing People and Organisations, <http://www.w3.org/TR/vcard-rdf/>
- [30] Szasz, B., Fleiner, R., Micsik, A. & Simon-Nagy, G. (2016) iLOC – An Indoor Ontology. <http://lod.nik.uni-obuda.hu/iloc/iloc-20160409.owl>
- [31] Szasz, B., Fleiner, R., Micsik, A. (2016) OLOUD: Ontology for Linked Open University Data. <http://lod.nik.uni-obuda.hu/oloud/oloud-20160609.owl>
- [32] Szasz, B., Fleiner, R., Micsik, A. (2016) Ontology Extension for Temporal Data. <http://lod.nik.uni-obuda.hu/oloud/oloud-time-20160609.owl>
- [33] Vandebussche, Pierre-Yves, and Bernard Vatant. "Metadata recommendations for linked open data vocabularies." Version 1 (2011): 2011-12
- [34] Szasz, B., Fleiner, R., Micsik, A. (2016) OLOUD – Ontology for Linked Open University Data. <http://lod.nik.uni-obuda.hu/oloud/index.html>
- [35] Poveda-Villalón, M., Suárez-Figueroa, M. C., & Gómez- Pérez, A. (2012) Validating Ontologies with OOPS!. In *Knowledge Engineering and Knowledge Management* (pp. 267-281). Springer Berlin Heidelberg
- [36] Szasz, B., Fleiner, R. and Micsik, A. (2016) Linked Data Enrichment with Self-Unfolding URIs. In *2016 IEEE 14th International Symposium on Applied Machine Intelligence and Informatics (SAMI 2016)* pp. 305-309