OLOUD – An Ontology for Linked Open University Data

Barnabás Szász, Rita Fleiner and András Micsik

Faculty of Informatics, University of Debrecen, Kassai út 26., 4028 Debrecen, Hungary
Neumann Faculty of Informatics, Óbuda University, Bécsi út 96/b., 1034 Budapest, Hungary
Institute for Computer Science and Control, Hungarian Academy of Sciences, Lágymányosi u. 11., 1111 Budapest, Hungary

Abstract. The Linked Open Data pursuit has achieved remarkable progress in Europe as well, and studies have shown that it has a positive impact on the quality of education at university level too. Publishing information about university or college course, their corresponding places and related events, such as exams in Linked Data format allows the event information to be aggregated, filtered and delivered to potential participants: students and lecturers via multiple channels and devices. In this paper a new ontology is described, an Ontology for Linked Open University Data (OLOUD), which supports the development and publishing of Linked Open University Datasets and the applications built on the top of these datasets. The domain of the OLOUD ontology consists of the open data one would publish within the university. OLOUD provides a high level model covering multiple education related use cases catalyzing linked data production and consumption within the domain. OLOUD contains classes and properties to describe Organizations, People, their Roles and Publications, Subjects, Courses and other Events together with their temporal and spatial description.

Keywords: Linked Open Data, Linked Open University Data, OWL

1. Introduction

1.1. Motivation

The objectives of our ontology is to support the development and publishing of Linked Open University Datasets and the applications built on the top of these Open Datasets. The domain of our OLOUD ontology consists of the open data one would publish within the university.

The main motivation of developing a new ontology born during the first trial implementation: In this domain 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 are incomplete, deficient and not extensible. Some of our needs were not addressed at all by any of these existing ontologies. For example the possibility of precise semantic representation of temporal and location data (including course schedule, event data and indoor location data) is missing from the already published university related ontologies. This recognition gave the purpose of the new ontology development and also this paper. The aim of this paper is to describe the process, our challenges and the solutions in developing an Ontology for Linked Open University Data, which is well designed, comprehensive and extensible.

1.2. Methodology

During the development of the ontology the guidelines described in [5] were preserved, aiming to design a 4 star vocabulary. The following rules were applied in order to restrict the potential interpretations of the defined classes and properties towards their intended meaning [5]:

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

Our university ontology was developed based on the Uschold and King methodology, which consists of the following steps [12], [13]:

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

The rest of the paper is organized according to the above methodology steps. In section 2 the objectives of the ontology is outlined, in section 3 the ontology specification is described, in section 4 the process of the integration with other ontologies is showed, in section 5 the details of the ontology development are presented, in section 6 the ontology documentation and evaluation are discussed. Finally in section 7 we discuss related work and conclude in section 8.

2. Objectives

The primary users of the services based on our ontology are university students and lecturers but the user base can include anyone with any touch points to the university, for example attendees of events organized by the university or applicants to the university. The first use of the ontology is publishing Linked Open Data from the Óbuda University.

The scope of the ontology is primarily to support publishing of Linked Open University Datasets, secondarily the applications built on top of this data. Applications could include mobile apps for supplying students with course schedule information, including exact times and places for courses and navigation to these places, or providing access to course materials and other curriculum. For lecturers applications built on top of OLOUD would provide overview of their educational activities, data about their research and collaboration opportunities. The different educational units could use these applications to publish data about their research and teaching activities.

OWL 2 DL was chosen as the formal language for OLOUD. Objectives include to reuse existing ontologies as much as possible.

The following is the result of our research on potential open data sources and use cases:

- personal data (employees, researchers): contact information, courses, personal interest, research activities and publications, external collaboration, etc.
- organizational units: leaders, contact information, organizational hierarchy, projects, etc.
- courses: topics, credits, lecturers, curriculum, dependencies and other conditions, etc.
- study programs: objectives, conditions, information about enrollment and degree, courses, etc.
- documents: regulations, curriculum materials, publications, metadata of documents (title, authors, publish date), etc.
- research projects: topics, participants, collaborations, etc.
- statistics: about students, lecturers, budget, expenses, duration of education, employment rates, etc.
- spatial descriptions: buildings, rooms, location of service providers, navigation and transportation support, etc.
- temporal descriptions: events calendar, course schedule, conferences, other events, etc.

Our expectations of the ontology are the following:

- describe all the significant units of a university,
- list all the employees and lecturers of the university with their roles,
- describe all subjects and courses based on these subjects,
- describe significant areas of the university buildings and a route among these,
- define and maintain the course schedule with the least effort possible,
- describe events related to the university with basic information, like name, place and time of the event.
3. Ontology specification

The specification of an ontology includes the class definitions, the entities of the classes as well as their properties and the relationships between classes (which includes the parent-child relations between two classes and entities of these two classes).

The ontology comes with a root level class: OLOUDThing, which is the parent class of all other OLOUD classes. The following are the main classes in the ontology, which is shown in Figure 1:

1. *Organization* that contains the units of the university including the university itself, the different faculties, departments and other operational units. These units have name, phone number, web page, executive, etc.
2. *Publication*: such entities are defined by the author(s), title, creation and/or publication date, language and topic, etc.
3. *Person*: such entities have names, personal web pages, phone numbers, organizational unit, role, etc.
4. *Role*: such entities have name, start and end dates, organizational unit and holder.
5. *Subject*: such entities are featured by their name, code, credit number and the person responsible for it.
6. The entities of the *Course* class are the actual instances of subjects having spatial and temporal descriptions, identification number and instructor. In order to describe entities of the Subject and Course classes properly some auxiliary classes for course term, course type, university majors, academic degree, language, specialization and training type were introduced.
7. *Event* is a high level class, its instances are representing arbitrary time/space regions describing a wide range of context: e.g.; conferences and concerts, exams and ceremonies, etc.
8. *Location* class is used to represent all the necessary entities describing indoor locations. It contains subclasses describing buildings with internal structure, e.g. floors, rooms, hallways and vertical passages and Point of Interests. The overview of the related classes is shown in Figure 2.
9. *Instant*, *Interval* and *TemporalSeq* classes form the Owl Time and TimeAggregates ontologies are used to represent single or recurring instants and intervals in the temporal space. They are used for temporal description of entities belonging to *Course*, *Role* and *Event* classes.

The OLOUD ontology consists of two modules: OLOUD-BASE\(^1\) and OLOUD-LOC\(^2\). The former describes all the university related concepts (classes 1-7, from above) and uses the prefix oloud. The latter provides the indoor location description and navigation features with the help of the Location class and uses the prefix loc.

4. Integration with other ontologies

In the process of creating the schema for a Linked Open Dataset it is advisable to reuse as much as possible of the available ontologies or vocabularies [5], but there are quite different vocabulary reuse strategies [9]. The two basic forms are (1) reusing classes and properties from existing vocabularies directly, and (2) establishing links on schema-level. The second case means defining new classes as either subclasses (rdfs:subClassOf) or equivalent classes (owl:equivalentClass) and properties as subproperties (rdfs:subPropertyOf) or equivalent properties (owl:equivalentProperty) of the classes and properties of the reused ontology. This way instances of these classes can use the properties of the parent classes.

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. There are no established recommendations in the literature about how to select vocabularies for reuse [9].

In case of OLOUD our strategy was the following. First the necessary concepts (classes and properties) were identified and then the list of vocabularies was chosen from the best known and the lowest number of OWL 2 DL compliant ones that could serve to express the defined concepts. In case of the classes there were always new OLOUD terms introduced with links to the members of the previously identified list (if possible). In case of the properties only the most well-known vocabulary terms were reused directly (like FOAF, Time, Event, Dublin Core), in other cases links were established on schema level (if they were found).

The following ontologies and RDF schemas were chosen to be reused and extended in our model.

\(^{1}\)http://lod.nik.uni-obuda.hu/oloud-base\(^{2}\)http://lod.nik.uni-obuda.hu/oloud-indoor#
The AIISO\(^3\), AIISO-Roles\(^4\), Participation\(^5\) and TEACH\(^6\) Ontologies provide a schema to describe academic concepts, like lectures and lecturers, internal structure of the institution, subjects, courses, academic roles and participation in these roles.

FOAF (Friend of a Friend)\(^7\) provides a basic vocabulary to describe personal entities and their attributes.

To describe events the Event Ontology\(^8\) was reused, which is using the W3C OWL-Time Ontology\(^9\) to describe the temporal attribute of an event. To define recurring events, the Temporal Aggregates Ontology provides the respective constructs.\(^[7]\)

The root level class (loc:Location) in the Indoor Navigation module is a subclass of the geo:Spatial-Thing class reusing the W3C Basic Geo Vocabulary\(^10\) in order to describe the real-world location of our entities in a standard way. The vCard Ontology\(^11\) is used to describe addresses of buildings.

The purpose of Dublin Core\(^12\) is to be a common vocabulary describing documents and their attributes, it is used to describe the OLOUD Ontology itself.

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 them. 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 inaccuracy and mistakes in them.

The concepts in the TEACH ontology are very close to our needs, but they lack some important features that were essential for our purposes. For example in order to describe university courses the concept of Subject is necessary, which doesn’t exist in TEACH. Another problem was that one would expect an owl:DataTypeProperty based on the example data, 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\(^13\) was imported (that is published together with the OLOUD ontology).

Problems also emerged 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 Protégé decided wrongly to use owl:ObjectProperty instead of owl:DataTypeProperty.

Overall, the aim was to create a comprehensive ontology wrapping existing modelling efforts into a single model. This is in accordance with the three star recommendation in \(^5\), and thus our users need to learn a single ontology only, and we also keep the local control over our ontology.

5. Ontology Development

5.1. URI model

Different namespaces were used for ontology concepts and for instances. The T-Box (terminology) of the Ontology is identified with the following URI schema: http://lod.nik.uni-obuda.hu/oloud-base#{class or property name} and the A-Box (instance data) is identified with http://lod.nik.uni-obuda.hu/data#{instance_ID} URIs.

5.2. Main OLOUD module

The main OLOUD module provides classes and properties to describe the organizational structure of a university, the subjects and courses of its training, university related events, persons and their publications. Figure 1 shows the main classes (as ovals) and the most significant object properties (as arrows) between the classes. Classes defined in this module are written with oloud prefix, classes and properties needed from other ontologies are written with their prefix and marked with dashed line. The main classes are Person, Publication, Course, Subject, Organization, Event and Role, but to remain consistent some auxiliary classes for describing course term, course type, university majors, academic degree, language, specialization and training type were also introduced.

\(^3\) http://purl.org/vocab/aiiso/schema-20080925.rdf#  
\(^4\) http://purl.org/vocab/aiiso-roles/schema#  
\(^5\) http://purl.org/vocab/participation/schema#  
\(^6\) http://linkedscience.org/teach/ns#  
\(^7\) http://xmlns.com/foaf/0.1/  
\(^8\) http://motools.sourceforge.net/event/event.html  
\(^9\) http://www.w3.org/TR/owl-time/  
\(^10\) http://www.w3.org/2003/01/geo#vocabulary  
\(^11\) http://www.w3.org/TR/vcard-rdf/  
\(^12\) http://purl.org/dc/terms/  
\(^13\) http://lod.nik.uni-obuda.hu/teach-fixed.owl
5.3. Indoor Location/Navigation Module

This module provides the indoor location description and navigation features. It defines classes to describe buildings with internal structure: floors, rooms, hallways and vertical passages and Point of Interests. The POIs have significant role in the navigation, as these are the elements of a classical indoor route. Figure 2 depicts the classes (as ovals) and properties (as arrows). Dashed lines mark external classes or properties, dotted arrows mark subclass relationships. The module supports two levels of free-text type indoor navigation: The first option is based on the structure of the building (and based on the loc:isPartOf property) and provides simple instructions like „Enter the main building. Go to the 2nd floor. Look for Room 212.“ The second option builds on the top of the structural description, introducing special points in the space: „Point of Interest (POI)“. The route one can choose to navigate through is a network of POIs within the building. One POI instance is connected to one or multiple other POI by the loc:connectsPOI property.
As event locations are Rooms, the loc:belongsToRoom and loc:hasPOI properties are responsible to connect the rooms to the POI network by the nearby or including POIs. The second approach is more aligned to instructions one would expect asking someone familiar with the building: „Enter the main hall. Pass the coffee machine. Go to the elevator. Go to the 2nd floor. Pass the restrooms. Look for Room 212.” While the second approach allows more accurate navigation, there is the extra cost of modelling the POI network compared to the first approach. 

To navigate with linked data based on the location/navigation module, SPARQL 1.1 with OWL inference (for symmetric properties) is suitable. The following is an example SPARQL query to select the shortest (least intermediate steps) route between two rooms in a building (with maximum length of 3 for the sake of terseness):

```
PREFIX loc: <http://lod.nik.uni-obuda.hu/oloud-indoor#>
PREFIX geo: <http://www.w3.org/2003/01/geo/wgs84_pos#>

  BIND (<room1> AS ?start ).
  BIND (<room2> AS ?end).
  OPTIONAL {?start rdfs:label ?sl.}
  OPTIONAL {?p1 rdfs:label ?l1.}
  OPTIONAL {?p2 rdfs:label ?l2.}
  OPTIONAL {?p3 rdfs:label ?l3.}
  OPTIONAL {?end rdfs:label ?el.}
  ?p1 loc:belongsToRoom ?start.
  ?p3 loc:belongsToRoom ?end.
} ORDER BY ?distance LIMIT 1
```

6. Ontology evaluation, documentation

6.1. Triplification – the Óbuda University use case

In order to evaluate and validate the ontology LOD triplets were created based on public data at the Óbuda University (OU). Triplets of the different classes were derived from different sources, and the technique – the actual method the triplets were created – depended on

![Fig. 2. The structure of the Indoor Location/Navigation module](http://lod.nik.uni-obuda.hu/oloud-indoor#)
the actual source. The location data was created manually based on building layout diagrams of the Óbuda University, while the subject and course data was automatically converted with PHP scripts from relational database dumps extracted from the electronic administration system of the Óbuda University. The entities of the Person class and their personal data was scraped from the personal webpages. The university events were generated with a crawler from the OU webpage. The triplets were loaded and tested by Protégé, and also by a Fuseki test server at the OU.

The dataset was extended with links to the GeoNames14 geographic dataset for linking cities.

As already mentioned, 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. For the instance_ID of the Person class the name of the person, for the faculties and departments their abbreviation were chosen. The uniqueness of the instance_ID among these instances has to be ensured.

6.2. 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 its properties within Protégé. Misconceptional issues were discovered in several cases by that Protégé couldn’t create a property for the intended purpose. The root cause of these issues was mostly flaws in the imported third party ontologies, one frequent issue was mistyped properties in the third party ontologies as mentioned before. 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 third phase of the ontology evaluation it was inspected whether the original objectives and expectations were met, if one can create statements fulfilling the criteria: with OLOUD one can model generic information about university organizations (e.g. departments), persons, their publications and roles within an organization, subjects being taught and the specific courses with details about their recurring time and location details, as well as generic events.

This included additional individual creation and testing the complex use cases with different SPARQL queries, such as:

- Timetable generation (courses with time and location information) for students and lecturers
- Chain of POI based navigation to a specific event’s location
- Simple navigation based on internal structure of buildings

The OOPS! Ontology Pitfall Scanner was also used to check our OWL 

The OLOUD Ontology was implemented in a way that it is self-documenting. 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 [14].

The Ontology is licensed under the terms of Creative Commons 3.015

7. Related work

Linked Universities16 and Linked Education17 are two European initiatives 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. It promotes a set of vocabularies18 describing academic related entities. LinkedEducation.org is an open platform aimed at further promoting the use of Linked Data for educational purposes.

In [2] the evolution process of the Open University Linked Open Data platform is described, how it

14 http://www.geonames.org/
15 http://creativecommons.org/licenses/by/3.0/
16 http://linkeduniversities.org
17 http://linkededucation.org
evolved from a research experiment to a data hub for the open content of the university. Their datasets can be classified in the following six groups: Open Educational Resources, Scientific production, Social media, Organizational data, Research project output, Metadata. There are common areas in the linked open datasets of the two universities, but for example navigation and timetable data are missing at the Open University. To describe the data at the Open University 125 classes, 785 properties from 57 public vocabularies are used. Their main effort in the modelling was to reuse terms from existing vocabularies directly. The large number of the used vocabularies, the redundancy in the data and in the used properties are the consequences of their approach: to privilege the reuse of existing terms and the will to choose the best possible terms instead of being restricted to the semantics of only a few widely used ontologies. Our aim with the OLOUD was to provide a frame ontology for the description of the university open data that can be easily expanded if necessary.

Besides correcting the mistakes in the already existing ontologies in this field, OLOUD improves the facilities of the TEACH and AIISO ontologies by connecting the lecture rooms with indoor navigation capabilities. TEACH is rather an RDF schema than an ontology. Available examples use AIISO, TEACH and iCal ontologies to describe courses. AIISO is mainly used to connect courses with the organizational structure of the university. Lectures can either be represented as a set of dedicated calendar events or by providing the frequency, day and time of events. In contrast of this, with OLOUD one can provide detailed description in an ontological way.

There were several attempts to provide indoor navigation ontologies. OntoNav [11] is a semantic indoor navigation system and an ontological framework of handling routing requests. Geodint [6] uses standard shortest path algorithm in a derived graph model for navigation. ONALIN [4] provides routing for individuals with various needs and preferences; it takes the ADA (American Disability Act) standards, among other requirements, into consideration. However, none of these approaches are suitable or available for re-use currently. Therefore, a simplified version of ONALIN has been re-created as the location module of the OLOUD ontology. SPARQL 1.1 features can be used, especially path properties in navigation for finding path between two points, which provides the advantage to have a single data model, and thus path finding can be combined with additional semantic queries.

Correndo, et al. [1] address the issue of representing time entities (i.e. Instants and Intervals) as Linked Data, and how to exploit topological temporal relationships in order to increase the connectivity degree within Linked Data sets. They present an approach to describe temporal entities as reusable URIs that can be adopted by data publishers as a temporal context to their information resources. The approach identifies a set of discrete temporal entities as relevant for a certain domain (e.g. financial years for the public sector) while a RESTful API is provided to users to dynamically create their own temporal entities. Once a dynamic temporal URI is resolved, information is provided to situate such URI in reference to the domain relevant entities. The URI resolution employs simple topological temporal reasoning in order to exploit the qualitative relationships between entities. [10] provides a framework to generate RDF description to entities based on information encoded in their URI. These instances (such as Instant in OWL Time) can be generated based on a template defined by SPARQL Construct queries. This approach allows one being able to resolve URIs, generate and store instance details in a triple store allowing queries and reason on it.

8. Conclusion

The OLOUD ontology provides classes and properties to describe public data generally available in the university domain. As its design is completely generalized, we hope that it will be useful for other universities as well. We have already contacted other Hungarian universities and promoted the use of the OLOUD ontology for publishing their Linked Data.

This ontology provides the basis of several ongoing student projects, which either generate new datasets for the university or implement new services on top of the model and data provided by the OLOUD effort. In the future our plan is to gradually introduce the ontology for more and more datasets about Obuda University.

Universities may have some specific type of open data that they wish to publish. In the future the OLOUD ontology can be extended with some arising features. For example the ability to list all the ongoing and past researches conducted by and at the university,
including the results, the funders and all the participants or the ability to define the external relationships, the contact persons and external collaborations might become justified expectations to extend the OLOUD model.

References