Interactive Learning Environments 2015, Vol. 23, No. 1, 55–78

Production and consumption of university Linked Data

Fouad Zablith[a]*, Miriam Fernandez[b] and Matthew Rowe[c]

[a]Olayan School of Business, American University of Beirut (AUB), P.O. Box 11-0236, Riad El Solh, Beirut 1107 2020, Lebanon; [b]Knowledge Media Institute (KMi), The Open University, Walton Hall, Milton Keynes MK7 6AA, UK; [c]School of Computing and Communications, Infolab 21, Lancaster University, Lancaster, LA1 4WA, UK

Linked Data increases the value of an organisation's data over the web by introducing explicit and machine processable links at the data level. We have adopted this new stream of data representation to produce and expose existing data within The Open University (OU) as Linked Data. We present in this paper our approach for producing the data, based on well-defined workflows at the organisation as well as the technical levels. We also discuss the data already available to consume, and show the potential improvements that Linked Data brings by presenting three applications: (1) the OU Expert Search system for finding experts at the OU based on a specified topic of interest, (2) the Social Study application to identify potential courses for students based on their Facebook profile information, and (3) the Linked OpenLearn application that helps students identify related media and courses to OpenLearn units at the OU. Before concluding the paper, we show the potential benefits and an approach towards interlinking data beyond The Open University with other universities, using a common categorisation scheme.

Keywords: Linked Data; semantic web; open learn; education; expert search; social networks

1. Introduction

Educational institutions, as well as other organisations, often maintain their public data disconnected in different systems, making it hard to discover, obtain and integrate. For example, in the case of the Open University, a developer aiming to create an application managing information about courses and educational material will need to request access to three different systems: student services, the library and the media repository. Therefore, he/she will have to deal with different system administrators and data access procedures in order to obtain information that is already public. This makes the internal development of new applications a high cost process, limiting the ways in which universities announce and expose their courses and educational material, hence decreasing students' satisfaction and awareness. As students are the core consumers of university services, this limitation may result in an indirect loss for the organisation.

Additionally, when performing their daily learning and investigation activities, students and academics have to go through the tedious task of browsing different

*Corresponding author. Email: fz13@aub.edu.lb

information resources. This is due to the fact that, although being publicly available, the information is generally not integrated across the websites of different departments. This problem is even more evident when extending it to the general educational sector. Different educational institutions produce yearly large amounts of educational material (videos, slides, documents, etc.). However, when students and educational practitioners have to perform learning and investigation tasks, they generally: (i) not find the best available resources for the topic they aim to investigate or, (ii) spend large amounts of time browsing the websites of different institutions in order to collect and extract key information about the topic. This not only constitutes a productivity loss (in terms of the time required to locate the relevant information) but a reusability loss as well: when suitable and available material that exists is not found, it will need to be recreated.

In this context, we believe that integrating information within university departments and across educational institutions is a key requirement towards enhancing the productivity and supporting the development of the educational sector. In the last few years, we have witnessed the development of Linked Data (LD), a set of principles and technologies that allow putting raw data online in a standard, web enabled representation. Following these principles, every piece of data becomes web addressable, and therefore, data across different places, stores and systems become linkable. The explosion of the Linked Open Data (LOD) movement in the last few years has produced a large number of interconnected datasets containing information about a large variety of topics, including geography, music and research publications among others (Bizer, Heath, & Berners-Lee, 2009). The movement is receiving worldwide support from public and private sectors like the UK[1] and US[2] governments, international media outlets, such as the BBC (Kobilarov et al., 2009) and the New York Times (Bizer, 2009), and companies with a social base like Facebook.[3] Such organisations are supporting the movement either by releasing large datasets of information or by generating applications that exploit it to connect data across different locations.

Despite its relevance and the support received in the last few years, very few pieces of work have either released or exploited LOD in the context of education. A relevant example is the DBLP Bibliography Server Berlin[4], which provides bibliographic information about scientific papers. However, this is changing at a fast pace, as we are witnessing an increase in the number of universities generating LOD. Our view is that education is principally one of the main sectors where the application of the LOD technologies can provoke a higher impact. Like other domains, educational information is heterogeneous at both the inter- and intraorganisation level. For instance, it is common for equivalent courses, or equivalent educational material offered by disparate universities to be described using different data schemas or adhere to differing data structures. We believe that the application of LOD technologies, within and across educational institutions, can explicitly generate the necessary structure and connections among educational resources, providing better support for users and developers to locate, search, and compare content across information sources. An example of the value of this integration is supplying students with the possibility of comparing specific qualifications offered by different universities in terms of courses required, pricing and availability.

In this paper, we describe the creation of Linked Data at the Open University (OU) as part of the Lucero project[5], spanning multiple departments and administrative bodies. We detail the process via which Linked Data is generated and incrementally updated, thereby providing a blue print for other educational organisations to open up their data and enable its sharing and reuse. To demonstrate the utility of such exposed information, we present three example applications, each of which exploit the OU's Linked Data to provide useful education applications to end-users: (1) the OU expert search system, a tool focused on finding the best experts on a certain topic within the OU staff; (2) Social Study, a tool to recommend OU courses to users by exploiting their Facebook profile information; and (3) Linked OpenLearn, an application that enables exploring linked courses, podcasts and tags to OpenLearn units. We also describe the extension of this work to the interorganisation problem and present an approach with the aim to tackle this through the common categorisation of educational videos from disparate educational organisations.

The rest of the paper is organised as follows: Section 2 presents the state of the art in the areas of LOD within the educational context. Section 3 presents the work that has been done within the Lucero project to expose OU data as part of the LOD cloud. Section 4 presents example prototype applications that consume the OU's LOD for Expert Search, Social Study and Linked OpenLearn respectively. Section 5 discusses initial work towards interlinking educational material across institutions, going beyond the Open University data. In Section 6, we highlight the conclusions that we have drawn from this work.

2. Related work

While LOD is being embraced in various sectors, we are currently witnessing a substantial increase in universities adopting the Linked Data initiative. For example, the University of Sheffield's Department of Computer Science[6] provides a Linked Data service describing research groups, staff and publications, all semantically linked together (Rowe, 2010). In addition, more universities are making their Linked Data portals open to the public. This includes the University of Southampton[7], Lincoln[8], Mu¨ nster[9], Bristol[10], Edinburgh (starting with their buildings data[11]) and Oxford University[12]. The University of Southampton is also collecting the description and access points of the available LD datasets generated from different UK academic institutions, and making them available at: http://data-acuk.ecs.soton.ac.uk.

Other works related to the educational context are releasing library catalogues as Linked Data[13]. This includes for example OpenLibrary[14], LIBRIS[15], the American Library of Congress[16], and the German National Library of Economics (Neubert, 2009). In addition to library catalogues, publication materials are made available as Linked Data through for example RKBexplorer[17], DBLP[18], and the Semantic Web Dogfood Server (Mo¨ ller, Heath, Handschuh, & Domingue, 2007). Exposing the data coming from library catalogues and publications repositories will directly benefit the educational sector in easily locating and processing information, to serve the consumer needs (e.g. students, researchers or course producers).

3. The Open University Linked Data

The Open University is the first university in the United Kingdom to expose and publish its organisational information in LOD[19]. This is accomplished as part of the Lucero project (Linking University Content for Education and Research Online[20]), where the data extraction, transformation and maintenance are performed. This enables having multiple hybrid datasets accessible in an open way through the online access point: http://data.open.ac.uk.

The main purpose of releasing all this data as part of the LOD cloud is that members of the public, students, researchers and organisations will be able to easily search, extract and, more importantly, reuse the OU's information and data.

In this section, we provide the details of the steps needed for the production side of Linked Data at the Open University, as part of the Lucero project. We distinguish two phases involved in the production of Linked Data. The first phase is the data preparation, where the interaction occurs between various members of the university to identify, understand and appropriately model existing data. The second phase is the workflow used to process and expose the identified data as Linked Data, guiding the platform and system implementation.

3.1. Preparing the data

Working with the data at the OU was not a one person job. It involved various people with different areas of expertise, ranging from marketing and copyrights experts, to technical and communications experts. Figure 1 illustrates the process that we followed to prepare the data.

3.1.1. Initial meeting with the data owner

The data owners are one of the key persons in the process. They are the ones who have been working for years with their data, andare the best to understand the value behind it. The people involved in this meeting are the Lucero core team (a mix of technical and information management people), and the data owner. During this meeting, we

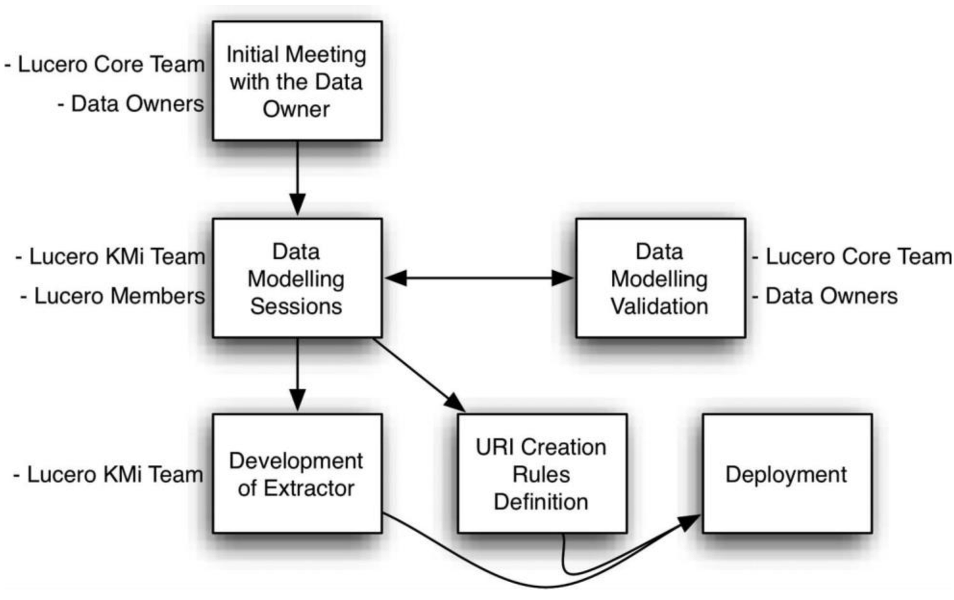

Figure 1. Data preparation process.

identify the potential data that is available and get a sample to have an initial idea of the structure and content involved. Another key factor discussed during such a meeting is the copyright that applies to the data, or the metadata, we are going to process. At this level, we also identify potential links to external data sources, for example links to Ordnance

Survey[21] to get further information of buildings' postcodes in the UK. We also identify with the data owners the users and use-cases, which would help to get an idea of the audience and potential applications for the data.

3.1.2. Data modelling sessions

After getting the data sample from the data owner, the Lucero team have multiple data modelling sessions. During the sessions, appropriate ontologies to reuse for modelling the data are identified. In cases where more than one ontology can be used (e.g. Courseware[22], AIISO[23], XCRI[24], and MLO[25] provide different vocabularies to represent courses information), we applied all the available ontologies to increase interoperability with external datasets. The ontologies are then mapped to the data. In the case where no existing ontologies cover parts of the data (e.g. OU specific podcast iTunesU links), the team had to come up with the appropriate terminology to represent such uncovered data parts.

3.1.3. Data modelling validation

In order to assess the generated models for the data, the Lucero core team held further meetings with the data owners to validate the models. The feedback from such meetings is used to refine the initial models. Depending on the complexity of the data, some cases required several meetings before getting the model to fit the purpose. For example, generating the model for course offerings involved more effort than others. This is due to the university's requirement that the correct pricing should be given to students depending on their region, a sensitive issue.

3.1.4. Development of extractor

Once the data model is approved, the KMi team of the Lucero project handle the technical part of developing the extractors for the data. As we discuss next, the extractors are developed depending on the way the data is exposed, and applying the appropriate model following the type of data being extracted.

3.1.5. URI creation rules definition

Another task performed by the KMi team is defining and coding the Uniform Resource Identifier (URI) rules that have been identified during the data modelling sessions. This step is important for ensuring the creation of links at the data levels, following the Linked Data principles[26] (Heath & Bizer, 2011). Further details about this process will be discussed in the next section, part of the Lucero workflow.

3.1.6. Deployment

The final step of the process is to deploy the extractors and URI rules on the server providing the Linked Data platform. This is also a technical task involving the KMi team. At this level, deployment should be performed without interrupting the access to existing data on the server. Another requirement is that the deployment should take into account the continuous update occurring at the original data sources. This technical matter is also part of the workflow that we discuss next.

3.2. Technical workflow for producing the data

We discuss in this part the steps involved in the creation of Linked Data at a technical level. To achieve this, the main requirement is to have a set of tools that generate Resource Description Framework (RDF) data from existing data sources, load such RDF into a triple store, and make it accessible through a web access point. Figure 2 illustrates the Lucero workflow, used to collect, maintain and expose Linked Data within the OU. It is component based, and designed to be efficient in terms of time, flexibility and reusability.

3.2.1. Step 1: Collect

The first step in the process is data collection. As in the case of most big organisations, the Open University owns and maintains various systems, which produce different datasets, often exposed in different formats. For example, courses information at the university are already accessible in XML format, while podcasts were exposed from the podcast web platform[27] through RSS feeds. The scheduler component takes care of initiating the extraction/update process at specific time

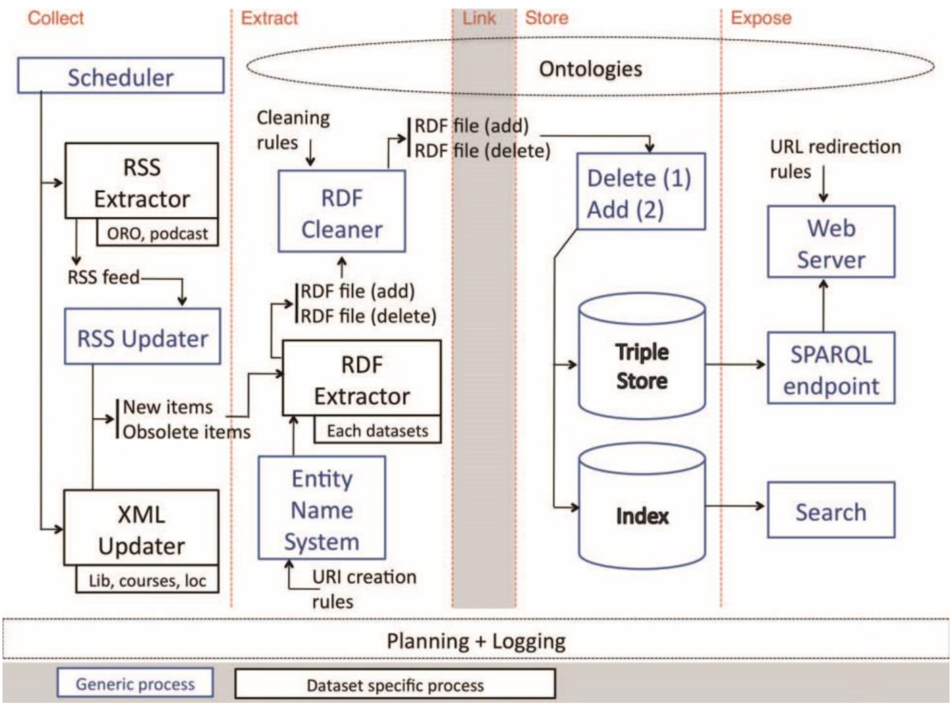

Figure 2. The Lucero workflow.

intervals. This update process is responsible for checking what was added, modified, or removed from the dataset, and accordingly triggers the appropriate action in the following steps. Having such a process in place is important in the OU scenario where the data sources are continuously changing. The workflow is designed in a way to access the data, without having to ask the data owners to change or modify their current system implementation. We made the effort to reuse what is already publicly available, or if this is not the case, we asked the data owners to produce data dumps on a regular basis (e.g. the university buildings data, or the library catalogue information). The datasets' characteristics played a major role in the implementation and setup of the components. For example, when the data sources are available in XML format, the XML updater will handle the process of identifying new XML items and pass them to the next step, where the RDF data is generated.

3.2.2. Step 2: Extract and link

Once the data items are identified in the collection step, they are passed to the RDF Extractor. An RDF extractor is developed for each type of data, to apply the appropriate models that have been agreed upon. To achieve interlinking the data entities, independently from which dataset the extraction is done, we rely on the Entity Named System component, which generates a unique URI depending on the specified item (this idea was inspired from the Okkam project[28]). The URI creation rules apply a unique combination following this pattern: ‘‘URI Prefix’’ þ ‘‘unique reference’’. For example, when the OU course on Natural and Artificial Intelligence is identified, code named ‘‘m366’’, the URI generation rules specify that the pattern in this case produces the URI ‘‘http:// data.open.ac.uk/course/m366’’, unique across all the university datasets. In this case, the URI prefix is ‘‘http://data.open.ac.uk/course’’ and the unique reference is ‘‘m366’’. Such unique URIs enable a seamless integration and extraction of linked entities within common objects that exist in the triple store and beyond, one of the core Linked Data requirements (Heath & Bizer, 2011). In the cases, where resources do not have a unique URI (e.g. a person referenced differently in various datasets[29]), we created different URIs for the same entity, and linked them through an owl:sameAs relation. In addition to links that are internal to the university's datasets, links to external datasets are specified as well at the extractors' level. Further details about which external datasets the OU dataset links to are presented in more details in the following section in the form of ontologies used. Currently, the linking process is hardcoded in the extractor. For example, links from the local library material records subject categories, to the Library of Congress Subject Headings (LCSH). URIs (extracted from its SPARQL endpoint[30]) are derived from a set of matchers predefined in the code. The matchers are built based on the analysis of the data fields available (e.g. the sequence of the subjects in the OU records, vs. the subjects in the LCSH records). If a match is found between the library subject record and the external LCSH URI, a link is created[31]. In other words, we reuse the existing subject URI available in the LCSH dataset in the OU records. The ultimate solution, left as part of the future work, would be to implement an automated technique for links discovery and creation. The extractor then passes the generated RDF files to the RDF Cleaner component. This component is in charge of deciding whether the extracted RDF file should be added or removed from the triple store.

3.2.3. Step 3: Store

The RDF cleaner passes the RDF files to the triple store, with an associated flag that indicates if the RDF should be deleted or added to the store. This mechanism is based on the analysis of the content of the generated files, compared to the last file that has been stored in the triple store. The comparison is based on the hash value of the file content. If the component detects any changes, the old file will be marked for removal, and the new one is subsequently added. In case the hash value is still the same, no action will be required. The values of the hash are stored in a separate database, along with the identifiers of the RDF files. This feature is important to lift the load from the triple store. Hence, instead of adding everything generated to the store, which will handle replacing the same entities that already exist (a default feature of the triple store used[32]), we run this process externally without touching the store. This will leave our engine focussed on serving and adding only the relevant data graphs. Part of the future implementation plan is to index our data, and provide a search mechanism on top of the index. The triple store in our workflow sits in the background, to expose the Linked Data in the following step.

3.2.4. Step 4: Expose

The final step of the workflow is to expose the produced Linked Data on the web, thereby making it accessible for consumption. We make the data accessible through a SPARQL endpoint[33], handled by the web server and the triple store. This is one of the most common ways to extract and consume Linked Data, making the query results directly processable and therefore usable within applications. The data can also be explored through the web browser. To achieve this, the web server handles the redirection rules to serve the right page to users depending on the page requested. For example, viewing the course m366 through the ‘‘http://data.open.ac.uk/page/ course/m366.html’’ URL will display the user-friendly page where users can directly click on the links and browse the data. While getting the resource RDF in RDF/ XML format can be done also through the browser by going to the resource page e.g. http://data.open.ac.uk/resource/course/l204. The redirection rules also handle redirecting the course to the original ‘‘Study at the OU’’ course pages.

The current implementation source code is open for reuse and can be accessed online through the project code repository[34].

3.3. The data available

Various datasets have been transformed and exposed as Linked Data at the Open University. This includes student services, media, research and estates-related data among others. We discuss in this section the available data currently accessible through data.open.ac.uk:

- Courses information: This dataset provides information about available courses including, among others: title, description, level, credits involved, location availability, presentation details, assessment type, related material (e.g. books, podcasts, YouTube videos) and subject. The sources used to generate this dataset are the XML documents of existing ‘‘Study at the OU’’[35] courses.
- Publications data: This dataset is comprised of publications by researchers at the Open University. A publication entity has the following information: a

title, type (e.g. conference or journal article), creators (i.e. authors or editors), publication date, where it has been presented, etc. The sources of this dataset are the RSS feeds supplied by the Open Research Online System[36].

. Library catalogue: This dataset contains material maintained by the library of the Open University. The data around such material contain information about: the type (e.g. book or media collection), description, date, publisher, links to courses, etc. Linked data has been generated in this case from the MARC records (Byrne, 1997) dumped daily from the university's library system[37].

. Podcasts: This set contains information about the OU podcasts. Each podcast has the following main properties: title, depiction, description, download link, collection part, duration, publication date, subject, genre, related course, etc. The data were generated based on the RSS feeds published through the OU podcast system[38].

. OpenLearn material: OpenLearn provides learning material freely available online. This dataset exposes the following main properties for each OpenLearn unit: title, description, related course, unit URL, subject, etc. The source used to create this dataset is the RSS feed of the OpenLearn system[39].

. Arts events: This dataset includes arts events from the Open Arts Archive project[40], with the collaboration of several galleries and museums across the UK. The main events properties are: event type, description, label, time, involved agents, subject, topics among others. The data are extracted from the Open Arts Archive website RSS feed.

. Reading experiences data: This data contains the records of reading experiences collected as part of the Reading Experiences project[41] at the OU. An experience is modelled with the following list of properties: reader and text involved, date of the experience, location, type, label, and the link to the document that describes the experience. The data are produced from a dump of the reading experiences project database.

. YouTube channel: The Open University maintains a YouTube channel[42], exposed through this dataset. YouTube videos contain the following main properties: title, description, related tags, download link, duration and thumbnails info. The data is generated by using the available YouTube API[43].

. University buildings: The Open University buildings information is maintained by the estates department, and is also made available as Linked Data. A building is modelled with the following main properties: name, postcode, depiction and entities located inside (e.g. floors, facilities, etc). The data is built starting from a CSV file exported from the estates internal systems.

. Department news: This dataset contains news articles from departments at the Open University. Currently, only the Knowledge Media Institute planet news website[44] is included. The model follows the existing RDF data of the news article, with the following main properties: title, creator, news date, link to the story and the description. The RDF data has been added directly from the news site.

. University staff: Staff information is also available as Linked Data. The main properties of a person entity are: name, homepage, projects involved, phone number, depiction, topics of interest, authored publications, among others. The KMi staff information is produced directly from the server feeding the

KMi website data. As part of the future work, we plan to gather the information from the other departments through the main Open University website.

We report in Table 1 the list of ontologies used for modelling our datasets. As previously mentioned, there are cases where we use more than one ontology to model the same entity.

Table 1. List of ontologies used for modelling the datasets.

Dataset	Ontologies
Courses information	Courseware: http://courseware.rkbexplorer.com/ontologies/courseware AIISO: http://vocab.org/aiiso/schema XCRI: http://svn.cetis.ac.uk/xcri/trunk/bindings/rdf/xcrirdfs.xml MLO: http://svn.cetis.ac.uk/xcri/trunk/bindings/rdf/mlordfs.xml Dublin Core: http://dublincore.org/2010/10/11/dcterms.rdf GoodRelations: http://www.heppnetz.de/ontologies/goodrelations/v1 GeoNames: http://www.geonames.org/ontology
Publications	Dublin Core: http://dublincore.org/2010/10/11/dcterms.rdf Bibliographic Ontology (Bibo): http://bibotools.googlecode.com/svn/bibo-ontology/tags/1.3/bibo.xml.owl
Library catalogue	DBpedia: http://dbpedia.org MulDiCat:http://iflastandards.info/ns/muldicat Nice Tag: http://ns.inria.fr/nicetag/2010/09/09/voc.html Dublin Core: http://dublincore.org/2010/10/11/dcterms.rdf Media Ontology: http://www.w3.org/TR/2010/WD-mediaont-1020100608/ Courseware: http://courseware.rkbexplorer.com/ontologies/courseware Media Vocabulary: http://payswarm.com/vocabs/media
Podcasts	DBpedia: http://dbpedia.org Nice Tag: http://ns.inria.fr/nicetag/2010/09/09/voc.html Dublin Core: http://dublincore.org/2010/10/11/dcterms.rdf Media Ontology: http://www.w3.org/TR/2010/WD-mediaont-1020100608/ Media Vocabulary: http://payswarm.com/vocabs/media Digitalbazar Media Ontology: http://digitalbazaar.com/media
OpenLearn	DBpedia: http://dbpedia.org Nice Tag: http://ns.inria.fr/nicetag/2010/09/09/voc.html Dublin Core: http://dublincore.org/2010/10/11/dcterms.rdf Media Ontology: http://www.w3.org/TR/2010/WD-mediaont-1020100608/ Creative Commons: http://creativecommons.org/ns
Arts events	DBpedia: http://dbpedia.org Linked Events: http://linkedevents.org/ontology/ Dublin Core: http://dublincore.org/2010/10/11/dcterms.rdf FOAF: http://xmlns.com/foaf/spec/

Reading experiences	DBpedia: http://dbpedia.org Linked Events: http://linkedevents.org/ontology/

(continued)

Table 1. (Continued).

Dataset	Ontologies
	Dublin Core: http://dublincore.org/2010/10/11/dcterms.rdf FOAF: http://xmlns.com/foaf/spec/ CiTO: http://purl.org/spar/cito/
YouTube channel	DBpedia: http://dbpedia.org Nice Tag: http://ns.inria.fr/nicetag/2010/09/09/voc.html Dublin Core: http://dublincore.org/2010/10/11/dcterms.rdf Media Ontology: http://www.w3.org/TR/2010/WD-mediaont10-20100608/ Media Vocabulary: http://payswarm.com/vocabs/media FOAF: http://xmlns.com/foaf/spec/
University buildings	vCards Ontology: http://www.w3.org/2006/vcard/ns Postcode Ontology: http://data.ordnancesurvey.co.uk/ontology/ postcode/ Dublin Core: http://dublincore.org/2010/10/11/dcterms.rdf GeoNames: http://www.geonames.org/ontology Buildings and Rooms Vocabulary: http://vocab.deri.ie/rooms University of Southampton Ontology: http://rdf.ecs.soton.ac.uk/ontology/ecs FOAF: http://xmlns.com/foaf/spec/
Department news	Dublin Core: http://dublincore.org/2010/10/11/dcterms.rdf SIOC Ontology: http://rdfs.org/sioc/ns
University staff	FOAF: http://xmlns.com/foaf/spec/ Contact Ontology: http://www.w3.org/2000/10/swap/pim/contact Dublin Core: http://dublincore.org/2010/10/11/dcterms.rdf

We highlight in Figure 3 the various datasets that have been so far transformed into Linked Data. In terms of the amount of data currently available in the triple store, we give a glimpse in Table 2 of the number of instances of the major entities in the OU Linked Data sets.

4. Linked data-based applications

Thus far we have explained how Linked Data pertaining to the Open University, crossing a range of departments and topics, has been produced automatically. To demonstrate the utility of this data within an educational context, we now describe three example applications, each of which consume the provided Linked Data in a unique manner: an Expert Search application, a Social Study application and a

Linked OpenLearn application.

4.1. The OU expert search application

Our first application is the OU Expert Search system. This application uses existing Linked Data provided by the Open University to recommend experts about a given topic of interest. The expect search task can be defined as the process of identifying

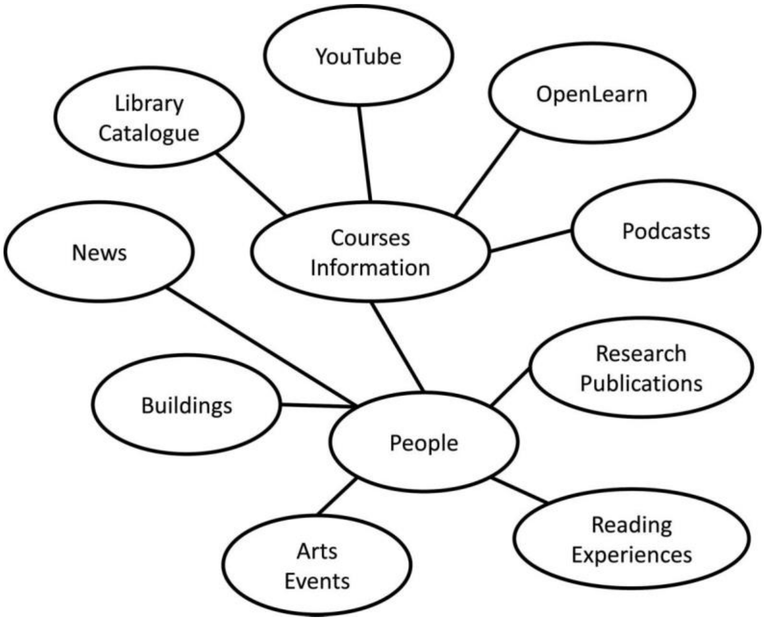

Figure 3. The available linked datasets.

Table 2. Number of instances of the major entities available at the time of writing.

Entity type	Number of instances
Courses	580
Research publications	16,000
Video podcasts	2,200
Audio podcasts	1,500
Open educational resources	640
YouTube videos	900
University buildings	100
Library catalogue books	12,000

people who have relevant expertise in a topic of interest, and is key for every enterprise, but especially for universities, where interdisciplinary collaborations among research areas is considered a high success factor. Typical user scenarios in which expert search is needed within the university context include: (a) finding colleagues from whom to learn, or with whom to discuss ideas about a particular subject; (b) assembling a consortium with the necessary range of skills for a project proposal, and; (c) finding the most adequate reviewers to establish a program committee.

As discussed by Yimam-Seid and Kobsa (2003), developing and manually updating an expert system database is time consuming and hard to maintain. However, valuable information can be identified from documents generated within an organisation (Zhu, Huang, Song, & Ruger, 2010). Automating expert finding from such documents provides an efficient and sustainable approach to expertise discovery.

Open University researchers, students and lecturers constantly produce a plethora of documents, including for example conference articles, journal papers, theses, books, reports and project proposals. As part of the Lucero project, these documents have been pre-processed and made accessible as LOD. The purpose of this application is therefore to exploit such information so that OU students and researchers can find the most appropriate experts starting from a topic of interest[45].

4.1.1. Consumed data

This application is based on two main information sources: (a) LOD from the Open Research Online system (publications data), and (b) additional information extracted from the OU staff directory. The first information source is exploited in order to extract the most suitable experts about a certain topic. The second information source complements the previous recommended set of experts by providing their corresponding contact information within the OU. Note that sometimes, ex-OU members and external collaborators of OU researchers may appear in the ranking of recommended experts. However, for those individuals, no contact information is provided, indicating that those experts are not part of the OU staff.

As previously mentioned, the information provided by Open Research Online contains data that describe publications originating from OU researchers. In particular, among the properties provided for each publication, this system exploits the following ones: (a) the title, (b) the abstract, (c) the date, (d) the authors and, (e) the type of publication, i.e. conference paper, book, thesis, journal paper, etc. Below we present an example of a SPARQL query to extract the set of books from the Open Research Online System:

```
PREFIX bibo:5http://purl.org/ontology/bibo/4
PREFIX dc:5http://purl.org/dc/terms/4
PREFIX foaf:5http://xmlns.com/foaf/0.1/4
SELECT distinct ?book ?date ?title ?abstract
WHERE {?book a bibo:Book. OPTIONAL {?x dc:date ?date} OPTIONAL {?x dc:title
?title} OPTIONAL {?x bibo:abstract ?abstract} }OFFSET 0 LIMIT 10
```

In this case, Linked Data provides the information with a fine-grained structure that facilitates the design of ranking criteria based on multiple features (Zablith, Fernandez, & Rowe, 2011), as well as the interlinking of information with other repositories.

4.1.2. System implementation

The system is based on lightweight client server architecture. The back end (or server side) is implemented as a Java Servlet, and accesses the OU LOD information by means of HTTP requests to the SPARQL endpoint. Some of the properties provided by the LOD information (more particularly the title and the abstract of the publications) are periodically indexed using Lucene to speed-up and enhance the search process. The rest of the properties (authors, date and type of publications) are accessed at run time.

The front-end is a thin client implemented as a web application using HTML, CSS and Javascript (jQuery[46]). The client does not handle any processing of the data, it only takes care of the visualisation of the search results and the search input. It communicates with the back-end by means of an HTTP request that passes as a parameter the user's query and retrieves the ranking of authors and their corresponding associated information by means of a JSON object.

4.1.3. Example and screenshot

In this section, we provide an example of how to use the OU expert search system. As shown in Figure 4, the system receives as input a keyword query indicating the topic for which the user aims to find an expert (e.g. ‘‘semantic search’’). As a result, the system provides a list of authors (e.g. ‘‘Enrico Motta’’, ‘‘Vanessa Lopez’’, etc), who are considered to be the top OU experts in the topic. For each expert, if available, the system provides the contact details (department, e-mail, phone extension) and the top publications about the topic. For each publication, the system shows its title, the type of document, and its date. If the user passes the cursor on the

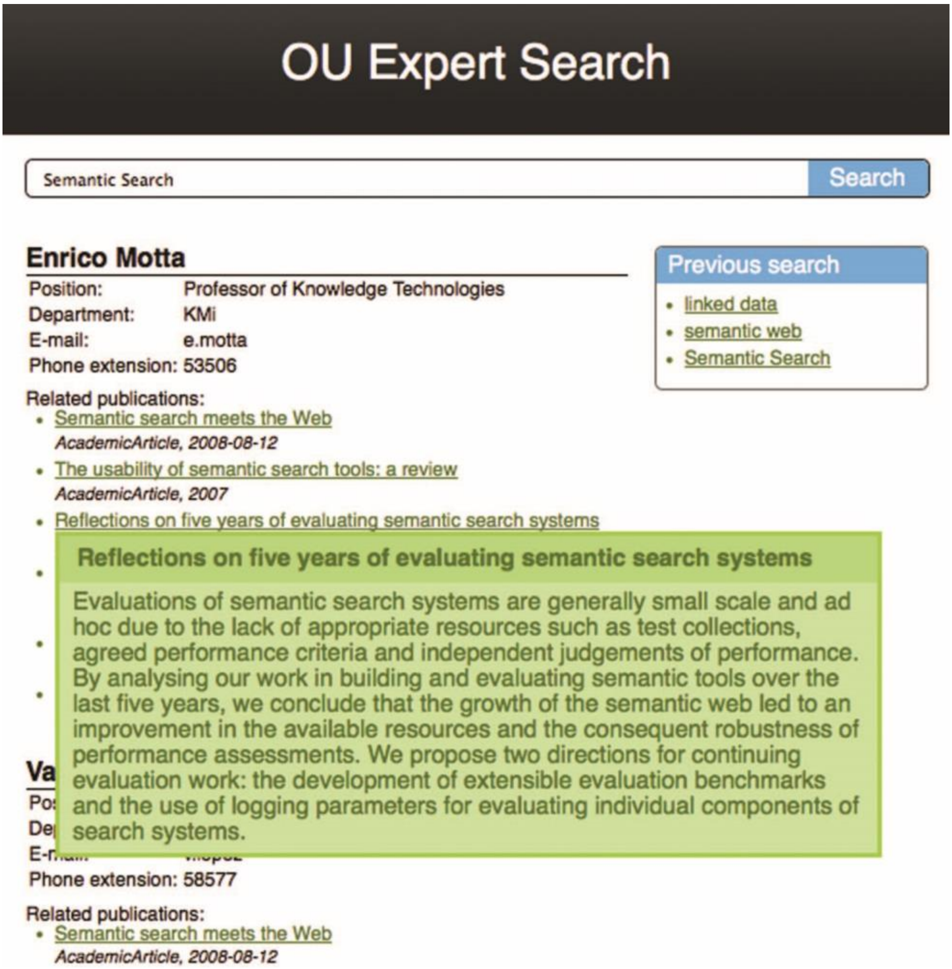

Figure 4. The OU expert search system.

top of the title of the publication, the summary is also visualised (see the example in Figure 4 for the publication ‘‘Reflections of five years of evaluating semantic search systems’’). In addition, the title of the publication also constitutes a link to its source page in data.open.ac.uk.

4.2. Social study application

Our second application is the Social Study application that recommends potential study partners and courses that could be studied together with those people. The Open University offers distance-learning courses covering myriad subject areas, however a common problem when deciding on which course to study is choosing a course that is relevant and close to an individual’s interests. One solution to this problem is to take advantage of existing profile information to bootstrap the decision process, in doing so leveraging information describing a person’s interests upon which possible course for studying could be pursued.

Based on this thesis, Social Study[47] combines the popular social networking platform Facebook with the OU Linked Data service, the goal being to suggest Open University courses that share common themes with a user's interests.

4.2.1. Consumed data

Social Study combines information extracted from Facebook with Linked Data offered by The Open University, where the former contains the profile information describing a given user – i.e. his/her interests, activities and ''likes'' – while the latter contains structured, machine-readable information describing courses offered by The Open University.

Combining the two information sources, in the form of a ''mashup'', is performed using the following approach. First the user logs into the application – using Facebook Connect – and grants access to their information. The application then extracts the concepts that the user has expressed an interest in on his/her profile. In Facebook, such interests can be expressed through one of three means: interests – where the user explicitly states that they are interested in a given subject or topic; activities – where the user describes his/her hobbies and pastimes, and; likes – where the user identifies web pages that he/she is interested in that are then shared with the individual's social network.

This collection of concepts extracted from each of these interest facets provides the profile of the given user. To suggest courses from this collection, the OU SPARQL endpoint is queried for all courses on offer, returning the title and description of each course. This information is then compared with the profile of the user as follows: each of the concepts that make up the user's interest profile – in the form of ngrams – are compared against the description and title of each course, and the frequency of concept matches is recorded for each course.

The goal of Social Study is to recommend relevant courses to the user given their interests, therefore the greater the number of concept matches, the greater the likelihood that the course is suited to the user. The courses are then ranked based on the number of overlapping concepts, allowing the user to see the most relevant courses at the top of the list, together with the list of concepts that informed the decision for the rank position of the course.

If for a moment we assume a scenario where Linked Data is not provided by the OU, then the function of Social Study could, in theory continue, by consuming information provided in an alternative form – given that the query component for the course information could be replaced by another process. However, the presence of Linked Data made the effort required to access and process courses information minimal. This work was an evolution of previous work (Zablith et al., 2011) that attempted to utilise the terms found in wall posts on Facebook in order to inform study partners about relevant courses. In evolving such work, our intuition is that the user interest profile that is presented on such a platform can be bootstrapped to suggest courses, thereby avoiding the time-consuming task of manual profile population – from which course suggestions would then be derived.

4.2.2. System implementation

The application is live and available online at the previously cited URL. It is built using PHP, and uses the Facebook PHP Software Development Kit (SDK[48]). Authentication is provided via Facebook Connect[49], enabling access to Facebook information via the Graph API. The ARC2 framework[50] is implemented to query the remote SPARQL endpoint containing The Open University's Linked Data, and parse the returned information accordingly.

4.2.3. Example and screenshot

To ground the use of Social Study, Figure 5 shows an example screenshot from the application when recommending courses for Matthew Rowe – one of the authors of this paper. The screenshot displays to the end user the order of courses together with the common interest concepts that their interest profile shares with those courses. The top-ranked course ‘‘The technology of music’’ matches the interest concepts music and techno that the user has specified in their profile. The greater the number of shared concepts with the course is, the greater the likelihood that the user will be interested in the course.

Figure 5. Social study showing the top ranked courses together with the interest concepts.

4.3. Linked OpenLearn application

Our third application is the Linked OpenLearn application, which ties together existing OpenLearn material with podcasts and courses also from The Open University, thereby using Linked Data to cross department boundaries and link content together. The Open University offers a set of free learning material through the OpenLearn website[51]. Such material covers various topics ranging from Arts[52], to Sciences and Engineering[53]. Additionally, the OU has other learning resources published in the form of podcasts, along with courses offered at specific presentations during the year. While all these resources are accessible online, connections are not always explicitly available, making it hard for students to find and exploit all the available resources. For example, while there exist links between specific podcasts and related courses, such links do not exist between OpenLearn units and podcasts. This leaves it to the user to infer and find the appropriate and relevant material to the topic of interest.

Linked OpenLearn[54] is an application that enables exploring linked courses, podcasts and tags to OpenLearn units. It aims to facilitate the browsing experience for students, who can identify on the spot relevant material without leaving the OpenLearn page. With this in place, students are able, for example, to find a linked podcast, and play it directly without having to go through the podcast website.

4.3.1. Consumed data

Linked OpenLearn relies on The Open University's Linked Data to achieve what was previously considered very costly to do. Within large organisations, it is very common to have systems developed by different departments, creating a set of disconnected data silos. This was the case of podcasts and OpenLearn units at the OU. While courses were initially linked to both podcasts and OpenLearn in their original repositories, it was practically hard to generate the links between podcasts and OpenLearn material. However, with the deployment of Linked Data, such links are made possible through the use of coherent and common URIs of represented entities.

To achieve our goal of generating relevant learning material, we make use of the courses, podcasts, and OpenLearn datasets in data.open.ac.uk. As a first step, while the user is browsing an OpenLearn unit, the system identifies the unique reference number of the unit from the URL. Then this unique number is used in the query passed to the OU SPARQL endpoint, to generate the list of related courses including their titles and links to the ''Study at the OU'' pages.

In the second step, another query is sent to retrieve the list of podcasts related to the courses fetched above. At this level, we get the podcasts' titles, as well as their corresponding downloadable media material (e.g. video or audio files), which enable users to play the content directly within the OpenLearn unit page. Finally, the list of related tags is fetched, along with an embedded query that generates the set of related OpenLearn units, displayed in a separate window. The user at this level has the option to explore a new unit, and the corresponding related entities will be updated accordingly. Below is an example of the SPARQL query used to extract the title and course codes linked to the value of variable ''$name'', which is the code of the OpenLearn unit the user is browsing, derived from the URL.

```
PREFIX ol:5http://data.open.ac.uk/openlearn4
PREFIX olont:5http://data.open.ac.uk/openlearn/ontology/4
PREFIX aiiso:5http://purl.org/vocab/aiiso/schema#4
PREFIX dc:5http://purl.org/dc/terms/4

SELECT distinct ?title ?code
WHERE ol:$name olont:relatesToCourse ?course . ?course dc:title ?title .
?course aiiso:code ?code}
```

The application is still a prototype, and there is surely room for improvement. For example, now that the data from the library catalogue has been made available, a much richer interface can be developed to support students towards the exploration of related books, recordings, computer files, etc.

4.3.2. System implementation

We implemented the Linked OpenLearn application in PHP, and used the ARC2 library to query the OU Linked Data endpoint. To visualise the data on top of the web page, we relied on the jQuery User Interface library[55], and used the dialog windows for displaying the parsed SPARQL results. The application is operational at present, and is launched through a Javascript bookmarklet, which detects the OpenLearn unit that the user is currently browsing, and opens it in a new iFrame, along with the linked entities visualised in the jQuery boxes.

4.3.3. Example and screenshot

To install the application, the user has to drag the applications' bookmarklet[56] to the browser's toolbar. Then, whenever viewing an OpenLearn unit, the user clicks on the bookmarklet to have the related entities displayed on top of the unit page. Figure 6

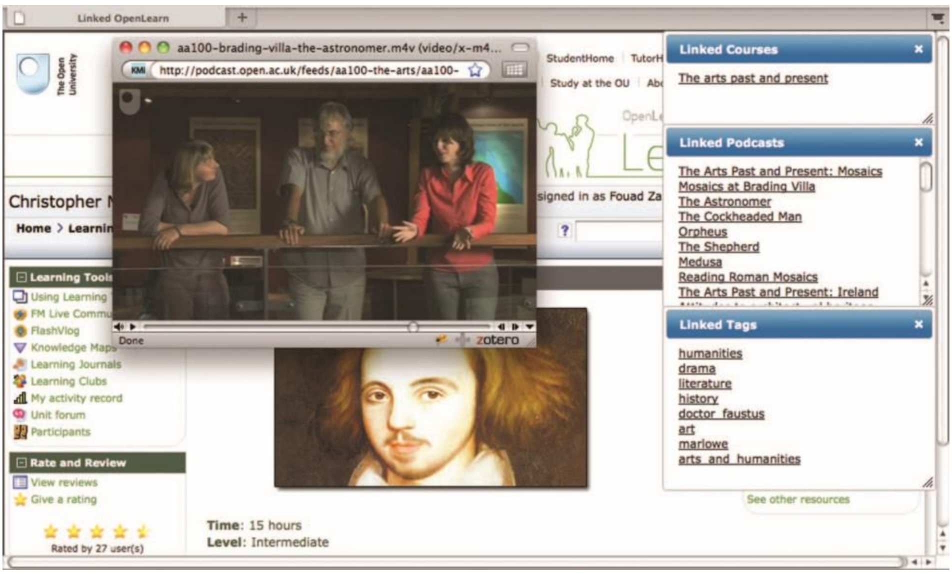

Figure 6. Linked OpenLearn screenshot.

illustrates an arts-related OpenLearn unit (referenced earlier), with the connected entities displayed on the right, and a running podcast selected from the ''Linked podcasts'' window. The user has the option to click on the related course to go directly to the course described in the ''Study at the OU'' webpage, or click on linked tags to see the list of other related OpenLearn units, which can be browsed within the same window.

5. Beyond the Open University data

As we have seen in the previous sections, the production of LD within the OU has allowed the integration of information across different departments, the enrichment of information with external data sources, and the development of novel applications. This is mainly reflected in Table 1, where the reuse of external ontologies and resources such as: DBpedia, Library of Congress Subject Headings, GeoNames, etc. All these elements present a clear example of the impact and benefits that LD technology can bring to the higher education sector. But our view is that the potential of LD in education and research goes well beyond the individual benefit of each institution. The availability of educational resources in an open, reusable form represents an unprecedented base of material for both students and teachers. It also offers new opportunities to address major issues in education, such as the access to high quality material from places where a strong university infrastructure is not available, as well as factoring the cost of content production across universities. Our view is that full potential of LD in the high education sector can only be achieved through community-wide efforts providing cross-university data that can be aggregated, integrated and compared. In this direction, several educational institutions have already joined the LD movement (see Section 2).

However, while the educational content is made available openly to be reused in different contexts from the ones for which it was produced, a major challenge appears in the aggregation, discovery and integration of various sources of educational material. When being produced and published in different contexts, educational resources can be vastly heterogeneous, especially in terms of their descriptive metadata and the classification schemes used to categorise them.

As part of this line of research, and as a first attempt to interlink educational information across universities through the use of LD principles and technologies, we have been working on the generation of a dataset of video material[57]. This dataset has been generated by selecting, extracting, structuring and interlinking information of more than 14,000 video lectures produced by 27 different educational institutions, including the OU. The technical details can be found in Fernandez, d’Aquin, & Motta (2011). In this section, we aim to briefly describe our experience integrating, under a common browsable and searchable space, educational material from various sources.

5.1. Choosing the common content categorisation scheme

There are almost as many categorisation schemes as there are open educational repositories. For example, Videolectures.net uses its own categorisation scheme that contains 23 main root elements including: Architecture, Arts, Biology, Business, Chemistry or Computer Science among others. On the other hand, OU Podcasts are classified under three different categorisation systems: OU specific subject headings, iTunes categories[58] and iTunesU categories[59].

It is therefore the case that, even if aggregated in a common metadata representation, resources originating from different sources would be categorised in topics identified differently, and still reside in different categorisation silos. Therefore, it is important to establish common categorisation schemes under which, educational content can be classified. When integrating video material, we searched for a categorisation scheme meeting the following requirements: to be general (i.e. aiming to cover all subjects in ‘‘the universe of information''), fully public, and available in RDF. We identified four different classification schemes fulfilling these requirements: the Open Directory Project topic hierarchy (ODP)[60], DBPedia categories[61], Library of Congress Subject Headings [62] and the International Press Telecommunications Council News Codes[63] (Fernandez et al., 2011). However, we finally chose the Open Directory Project topic hierarchy because of its maturity and the availability of classification tools to exploit it. We believe that community efforts should be performed to establish standardised schemas to classify educational data. For this purpose, we created a portal for the recommendation of such vocabularies for university data: linkeduniversities.org. This portal was created as a result of a meeting with the members of the LODUM project[64] and it currently counts with the support of seven different educational organisations.[65]

It is important to emphasise that Linked Data technologies provide tools to express, share and exploit semantic mapping or merging of concepts across vocabularies, e.g. represented using SKOS[66], as well as using properties from standardised ontology languages such as owl:sameAs, owl:equivalentProperty or owl:equivalentClass. This can allow universities and libraries to align their own vocabularies with a ''pivot'' classifications scheme, instead of modifying their own vocabularies. Having a ''pivot'' classification scheme reduces the impact of heterogeneity by providing a single point of search and navigation for content. While vocabularies can be aligned in a semi-(automatic) manner, maintaining these alignments can be a challenging task when the participating vocabularies change without employing robust notification systems and versioning approaches. It is therefore important for each organisation to balance the trade-off between: (i) aligning their own vocabularies with a ''pivot'', or commonly agreed, classification scheme and maintain these alignments and, (ii) provide their data following the agreed vocabularies and classification scheme.

5.2. Automatically classifying distributed content in a common scheme

To provide a homogeneous and navigable representation of open educational material, we had to categorise the content originating from different sources in the selected common categorisation scheme, the ODP. For this purpose, we reused the TextWise[67] categorisation service. Although the performed evaluation showed the efficiency of this service, obtaining 98% of coverage and 89% of accuracy (i.e. correct classifications), 49% of those classifications were not considered specialised (i.e. there was a more refined ODP category that could have been assigned to the content). It is our view that a deep evaluation of current available classification systems like UClassify,[68] AlchemyAPI,[69] DBPediaSpotLight,[70] Open Calais,[71] should be further performed in the context of this task, in order to optimise the generation of homogenous representation of educational data across origins.

As we have pointed out in this section, when interlinking educational video material across institutions, two main challenges have been identified: (i) to agree on a common categorisation scheme under which educational information can be explored and retrieved and (ii) to provide efficient mechanisms to map each institutional-dependent categorisation scheme into the common searchable and browsable space.

6. Conclusions

In this section, we report on our experiences when generating and exploiting Linked Data within the context of an educational institution. Regarding our work on transforming information distributed in several OU repositories and exposing it as LOD, the process complexity was mainly dependent on the datasets in terms of type, structure and cleanliness. Initially, before any data transformation could be done, it was required to decide on the vocabulary to use. This is where the type of data to model plays a major role. With the goal to reuse, as much as possible, already existing ontologies, it was challenging to find the adequate ones for all our data. While some vocabularies are already available, for example to represent courses, it required more effort to model OU specific terminologies (e.g. at the qualifications level). To ensure maximum interoperability, we chose to use multiple terminologies (when available) to represent the same entities. For example, courses are represented as modules from the AIISO ontology, and at the same time as courses from the Courseware ontology. Other factors that affected the transformation of the data are the structure and cleanliness of the data sources. During the transformation process, we faced many cases where duplication, and information not abiding to the imposed data structure, hampered the transformation stage. This initiated the need to generate the data following well-defined patterns and standards, in order to get easily processable data to add to the LOD cloud. This also allowed the team to automatically identify errors and report them to the corresponding data owners for further correction, enhancing the global data quality of the organisation.

Regarding our experience exploiting the data, we have identified three main advantages of relying on the LOD platform within the context of education. Firstly, the exposure of all these material as free Web resources provides open opportunities for the development of novel and interesting applications like the three presented in this paper. It is worth to note that there are further applications that make use of the OU Linked Data presented in separate papers (e.g. d'Aquin, Zablith, & Motta, 2011; Zablith, d'Aquin, Brown, & Green-Hughes, 2011) and on the Lucero project applications page[72]. The second main advantage is the structure provided by the data. This is apparent in the OU Expert Search system, where the different properties of articles are exploited to generate different ranking criteria, which when combined, provide much stronger support when finding the appropriate expertise. Finally, the links generated across the different educational resources have provided a new dimension to the way users can access, browse and use the provided educational resources. A clear example of this is the exploitation of LOD technology within the OpenLearn system, where OpenLearn units are now linked to courses and podcasts, allowing students to easily find through a single Web page relevant material that could be of interest.

We believe that universities need to evolve the way they expose knowledge, share content and engage with learners. We see LOD as an exciting opportunity to support this evolution, especially by interlinking people and educational resources within and across institutions. This interlinking of information will facilitate the learning and investigation process of students and research staff, enhancing the global productivity and satisfaction of the academic community. To achieve this goal, the community needs to share experiences and to agree on a set of collective vocabularies to model, structure and classify educational information, establishing a homogenous and navigable representation of educational material. To be truly effective, educational Linked Data should be the result of community-wide efforts rather than advances at the level of individual research groups.

Notes

1. http://data.gov.uk
2. http://www.data.gov/semantic/index
3. http://developers.facebook.com/docs/opengraph
4. http://www4.wiwiss.fu-berlin.de/dblp/
5. http://lucero-project.info/lb/
6. http://data.dcs.shef.ac.uk
7. http://data.southampton.ac.uk/
8. http://data.lincoln.ac.uk/
9. http://data.uni-muenster.de/
10. https://mmb.ilrt.bris.ac.uk/display/ldw2011/UniversityþofþBristolþdata
11. http://ldfocus.blogs.edina.ac.uk/2011/03/03/university-buildings-as-linked-data-with-scr aperwiki
12. http://data.ox.ac.uk
13. http://www.w3.org/2005/Incubator/lld/
14. http://openlibrary.org/
15. http://librisbloggen.kb.se/2008/12/03/libris-available-as-linked-data/
16. http://id.loc.gov/authorities/about.html
17. http://www.rkbexplorer.com/data/
18. http://dblp.l3s.de
19. http://www3.open.ac.uk/media/fullstory.aspx?id ¼ 20073
20. http://lucero-project.info
21. http://www.ordnancesurvey.co.uk
22. http://courseware.rkbexplorer.com/ontologies/courseware
23. http://vocab.org/aiiso/schema
24. http://svn.cetis.ac.uk/xcri/trunk/bindings/rdf/xcri_rdfs.xml
25. http://svn.cetis.ac.uk/xcri/trunk/bindings/rdf/mlo_rdfs.xml
26. http://www.w3.org/TR/cooluris/
27. http://podcast.open.ac.uk
28. http://www.okkam.org
29. See for example: http://data.open.ac.uk/page/person/bd5f674635cceb3f14e227a0eee8 e693
30. http://api.talis.com/stores/lcsh-info/services/sparql
31. See for example the book: http://data.open.ac.uk/page/library/463853
32. http://www.ontotext.com/owlim
33. http://data.open.ac.uk/query
34. http://code.google.com/p/luceroproject
35. http://www3.open.ac.uk/study
36. http://oro.open.ac.uk
37. http://voyager.open.ac.uk
38. http://podcast.open.ac.uk
39. http://openlearn.open.ac.uk
40. http://www.openartsarchive.org
41. http://www.open.ac.uk/Arts/reading
42. http://www.youtube.com/user/TheOpenUniversity
43. http://code.google.com/apis/youtube/overview.html
44. http://news.kmi.open.ac.uk
45. The OU Expert Search is accessible to OU staff at: http://kmi-web15.open.ac.uk:8080/ ExpertSearchClient. For external users a video is available under

http://smartproducts1.kmi.open.ac.uk/ExpertSearchClient/OUExpertSearch.avi as well as some documentation http://smartproducts1.kmi.open.ac.uk/ExpertSearchClient/index.htm
46. http://www.jquery.com
47. http://www.matthew-rowe.com/SocialStudy
48. https://github.com/facebook/php-sdk
49. http://developers.facebook.com/docs/authentication
50. http://arc.semsol.org
51. http://openlearn.open.ac.uk
52. OpenLearn unit example in Arts: http://data.open.ac.uk/page/openlearn/a216_1
53. A list of units and topics is available at: http://openlearn.open.ac.uk/course
54. http://fouad.zablith.org/apps/linkedopenlearn
55. http://www.jqueryui.com
56. The bookmarklet is available at: http://fouad.zablith.org/apps/linkedopenlearn, and has been tested in Firefox, Safari and Google Chrome
57. http://smartproducts1.kmi.open.ac.uk/web-linkeduniversities/index.htm
58. http://itunes.apple.com/us/genre
59. http://deimos.apple.com/rsrc/doc/iTunesUAdministrationGuide/iTunesUintheiTunesStore/chapter_13_section_3.html
60. http://www.dmoz.org/rdf.html
61. http://dbpedia.org/About
62. http://id.loc.gov/authorities/about.html
63. http://www.iptc.org/site/NewsCodes/
64. http://lodum.de/
65. http://linkeduniversities.org/lu/index.php/members/
66. http://www.w3.org/2004/02/skos/
67. http://textwise.com
68. http://www.uclassify.com/
69. http://www.alchemyapi.com
70. http://dbpedia.org/spotlight
71. http://www.opencalais.com
72. http://data.open.ac.uk/applications/

Notes on contributors

Fouad Zablith is a Visiting Assistant Professor at the Olayan School of Business of the American University of Beirut. His current research focus is on the use of semantic technologies to enhance user experiences in various contexts. He is also giving lectures on business information and decision systems. Dr. Zablith received his PhD from the Knowledge Media Institute of the Open University in the UK. His thesis was on using online ontologies to support ontology evolution. He investigated a pattern-based approach to assess the relevance of ontology changes during evolution. Dr. Zablith was part of several EU and UK funded projects. Within the LUCERO project, he worked on transforming university resources into linked data, and investigated several scenarios on consuming this data to deliver better information discovery and applications. He was also involved in the NeOn project where he applied his PhD work, and in the Open Arts Archive project as a web consultant to deliver a platform to share and expose arts related events across the UK. His work was published in a range of conferences and journals, and he served on the program committee of research events and conferences, and refereed a range of journal articles in his research area.

Dr Miriam Fernandez is currently a research associate at KMi. She received her BS, MSc (best student of promotion) and PhD (summa cum laude European mention) from the Universidad Autonoma de Madrid, Spain, in 2005 and 2009, respectively. Her research is focused on the synergy of Information Retrieval and Semantic Web technologies. She has participated in several European projects (aceMedia, Mesh, X-Media, SmartProducts), published in top-level conference proceedings

(ECIR, SIGIR, ESWC, ISWC, WWW) and journals (TKDE, TCSVT, JWS), and work for one of the main companies in the search engine and Web development market (Google Zurich).

Matthew Rowe is a Lecturer in Social Computing at Lancaster University in the UK. His current work explores user behaviour in social networks, assessing how users develop over time in given contexts and predicting their behaviour. His past work has explored behaviour analysis in order to predict community evolution and attention to content on social web platforms, as part of the EU funded project ROBUST, and his PhD, completed in 2010, explored automated techniques for the disambiguation of identity web references, where such techniques were supported by data leveraged from the Social Web. He has published in a range of international conferences and journals including the Journal of Web Semantics, the Semantic Web Journal and the International and European/Extended Semantic Web Conferences.

References

Bizer, C. (2009). The emerging web of linked data. IEEE Intelligent Systems, 24, 87–92.

Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data – the story so far. International Journal on Semantic Web and Information Systems, 5, 1–22.

Byrne, D.J. (1997). MARC manual: Understanding and using MARC records. Englewood, CO: Libraries Unlimited Incorporated.

d'Aquin, M., Zablith, F., & Motta, E. (2011). wayOU – Linked data-based social location tracking in a large, distributed organisation. In G. Antoniou, M. Grobelnik, E. Simperl, B. Parsia, D. Plexousakis, P. De Leenheer, & J.Z. Pan (Eds.), Proceedings of the 8th extended semantic web conference (Demo Part II) (pp. 461–465). Berlin, Heidelberg: Springer-Verlag.

Fernandez, M., d'Aquin, M., & Motta, E. (2011). Linking data across universities: An integrated video lectures dataset. In L. Aroyo, C. Welty, H. Alani, J. Taylor, & B. Abraham (Eds.), Proceedings of the 10th international conference on the semantic web – Volume Part II, Bonn, Germany (pp. 49–64). Berlin, Heidelberg: Springer-Verlag.

Heath, T., & Bizer, C. (2011). Linked data: Evolving the Web into a global data space (1st ed.). Synthesis Lectures on the Semantic Web: Theory and Technology. San Rafael, CA: Morgan & Claypool.

Kobilarov, G., Scott, T., Raimond, Y., Oliver, S., Sizemore, C., Smethurst, M., ... Lee, R. (2009). Media meets semantic web. How the BBC uses DBpedia and linked data to make connections, In L. Aroyo, P. Traverso, F. Ciravegna, P. Cimiano, T. Heath, E. Hyvnen ... E. Simperl (Eds.), Proceedings of the 6th European semantic web conference on the semantic web: Research and applications (pp. 723–737). Springer-Verlag, Berlin, Heidelberg.

Mo¨ ller, K., Heath, T., Handschuh, S., & Domingue, J. (2007). Recipes for semantic web dog food – the ESWC and ISWC metadata projects. In 6th international semantic web conference and 2nd ASWC, Busan, Korea.

Neubert, J. (2009). Bringing the ‘‘thesaurus for economics’’ on to the web of linked data. In Proceedings of the Linked Data on the web workshop, CEUR workshop proceedings (Vol. 538), Madrid, Spain. Aachen: M. Jeusfeld.

Rowe, M. (2010). Data.dcs: Converting legacy data into Linked Data. In Proceedings of the Linked Data on the Web Workshop, CEUR Workshop Proceedings (Vol. 628), Raleigh, North Carolina. Aachen: M. Jeusfeld.

Yimam-Seid, D., & Kobsa, A. (2003). Expert-finding systems for organizations: Problem and domain analysis and the DEMOIR approach. Journal of Organizational Computing and Electronic Commerce, 13(1), 1–24.

Zablith, F., d'Aquin, M., Brown, S., & Green-Hughes, L. (2011). Consuming linked data within a large educational organization. Workshop: Second International Workshop on Consuming Linked Data (COLD) at International Semantic Web Conference (ISWC), Bonn, Germany.

Zablith, F., Fernandez, M., & Rowe, M. (2011). The OU Linked Open Data: Production and consumption. Paper presented at Linked learning 2011: 1st international workshop on eLearning approaches for the linked data age, 8th extended semantic web conference (ESWC2011), Heraklion, Greece.

Zhu, J., Huang, X., Song, D., & Ruger, S. (2010). Integrating multiple document features in language models for expert finding. Knowledge and Information Systems, 23(1), 29–54.