The OU Linked Open Data: Production and Consumption

Fouad Zablith, Miriam Fernandez and Matthew Rowe

Knowledge Media Institute (KMi), The Open University
Walton Hall, Milton Keynes, MK7 6AA, United Kingdom
{f.zablith, m.fernandez, m.c.rowe}@open.ac.uk

Abstract. The aim of this paper is to introduce the current efforts toward the release and exploitation of The Open University's (OU) Linked Open Data (LOD). We introduce the work that has been done within the LUCERO project in order to select, extract and structure subsets of information contained within the OU data sources and migrate and expose this information as part of the LOD cloud. To show the potential of such exposure we also introduce three different prototypes that exploit this new educational resource: (1) the OU expert search system, a tool focused on finding the best experts for a certain topic within the OU staff; (2) the Buddy Study system, a tool that relies on Facebook information to identify common interest among friends and recommend potential courses within the OU that 'buddies' can study together, and; (3) Linked OpenLearn, an application that enables exploring linked courses, Podcasts and tags to OpenLearn units. Its aim is to enhance the browsing experience for students, by detecting relevant educational resources on the fly while reading an OpenLearn unit.

Keywords: Linked Open Data, education, expert search, social networks.

1 Introduction

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. [2]

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 [5] or the New York Times [1], and companies with a social base like Facebook.[3] Such organisations are supporting the movement either by releasing

[1] http://data.gov.uk
[2] http://www.data.gov/semantic/index
[3] http://developers.facebook.com/docs/opengraph

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. One of these few examples is the DBLP Bibliography Server Berlin,[4] which provides bibliographic information about scientific papers. However, education is principally one of the main sectors where the application of the LOD technologies can provoke a higher impact.

When performing learning and investigation tasks, students and academics have to go through the tedious and laborious task of browsing different information resources, analysing them, extracting their key concepts and mentally linking data across resources to generate their own conceptual schema about the topic. Educational resources are generally duplicated and dispersed among different systems and databases, and the key concepts within these resources as well as their inter and intra connections are not explicitly shown to users. 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 to users in their learning and investigation tasks.

In this context, the paper presents the work that has been done within The Open University (OU) towards the release and exploitation of several educational and institutional resources as part of the LOD cloud. First, we introduce the work that has been done within the LUCERO project to select, extract and structure subsets of OU information as LOD. Second, we present the potential of this data exposure and interlinking by presenting three different prototypes: (1) the OU expert search system, a tool focused on finding the best experts for a certain topic within the OU staff; (2) the Buddy Study system, a tool focused on exploiting Facebook information to identify common interests among friends and recommend potential courses within the OU that 'buddies' can study together, and; (3) Linked Open Learn, an application that enables exploring linked courses, Podcasts and tags to OpenLearn units.

The rest of the paper is organised as follows: Section 2 presents the state of the art in the areas of LOD within the education context. Section 3 presents the work that has been done within the LUCERO project to expose OU data as part of the LOD cloud. Sections 4, 5 and 6 present example prototype applications that consume the OU's LOD for Expert Search, Buddy Study and Linked OpenLearn respectively. Section 7 describes the conclusions that we have drawn from this work, and section 8 presents our plans for future work.

2 Related Work

While LOD is being embraced in various sectors as mentioned in the previous section, we are currently witnessing a substantial increase in universities adopting

[4] http://www4.wiwiss.fu-berlin.de/dblp/

the Linked Data initiative. For example, the University of Sheffield's Department of Computer Science[5] provides a Linked Data service describing research groups, staff and publications, all semantically linked together[6]. Similarly the University of Southampton has recently announced the release of their LOD portal (http://data.southampton.ac.uk), where more data will become available in the near future. Furthermore, the University of Manchester's library catalogue records can now be accessed in RDF format[6]. In addition, other universities are currently working on transforming and linking their data: University of Bristol,[7] Edinburgh (e.g., the university's buildings information is now generated in LOD[8]), and Oxford[9]. Furthermore the University of Muenster announced a funded project, LODUM, the aim of which is to release the university's research information as Linked Data. This includes information related to people, projects, publications, prizes and patents.[10]

With the increase of the adoption of LOD publishing standards, the exchange of data will be much easier, not only within one university, but also across the LOD ready ones. This enables, for example, the comparison of specific qualifications offered by different universities in terms of courses required, pricing and availability.

3 The Open University Linked Open Data

The Open University is the first UK University to expose and publish its organizational information in LOD.[11] This is accomplished as part of the LUCERO project (Linking University Content for Education and Research Online)[12], 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.

3.1 Creating the OU LOD

Detailed information about the process of LOD generation within the OU is available at the LUCERO project website.[12] We briefly discuss in this section

[5] http://data.dcs.shef.ac.uk
[6] http://prism.talis.com/manchester-ac
[7] https://mmb.ilrt.bris.ac.uk/display/ldw2011/University+of+Bristol+data
[8] http://ldfocus.blogs.edina.ac.uk/2011/03/03/university-buildings-as-linked-data-with-scraperwiki
[9] http://data.ox.ac.uk
[10] http://www.lodum.de
[11] http://www3.open.ac.uk/media/fullstory.aspx?id=20073
[12] http://lucero-project.info

the steps involved in the creation of Linked Data. To achieve that, the main requirement is to have a set of tools that generate RDF data from existing data sources, load such RDF into a triple store, and make it accessible through a web access point.

Given the fact that the OU's data repositories are scattered across many departments, using different platforms, and subject to constant update, a well-defined overflow needs to be put in place. The initial workflow is depicted in Figure 1, and is designed to be efficient in terms of time, flexibility and reusability. The workflow is component based, and 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 entities and pass them to the RDF extractor, where the RDF data is generated, and ready to be added to (or removed from) the triple store. Finally the data is exposed to the web, and can be queried through a SPARQL endpoint.[13]

The scheduler component takes care of initiating the extraction/update process at specific time intervals. This update process is responsible for checking what was added, modified, or removed from the dataset, and accordingly applies to the triple store the appropriate action. Having such a process in place is important in the OU scenario where the data sources are continuously changing. Another point worth mentioning is the linking process that links entities coming from different OU datasets (e.g., courses mentioned in Podcast data and library records), in addition to linking external entities (e.g., course offerings in a GeoNames defined location[14]). To achieve interlinking OU entities, independently from which dataset the extraction is done, we rely on an Entity Named System, which generates a unique URI (e.g., based on a course code) depending on the specified entity (this idea was inspired from the Okkam project[15]) . 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 [3].

3.2 The Data

Data about the OU courses, Podcasts and academic publications is already available to be queried and explored, and the team is now working to bring together educational and research content from the university's campus information, OpenLearn (already available for testing purposes) and library material. More concretely, data.open.ac.uk offers a simple browsing mechanism, and a SPARQL endpoint to access the following data:

[13] http://data.open.ac.uk/query
[14] http://www.geonames.org
[15] http://www.okkam.org

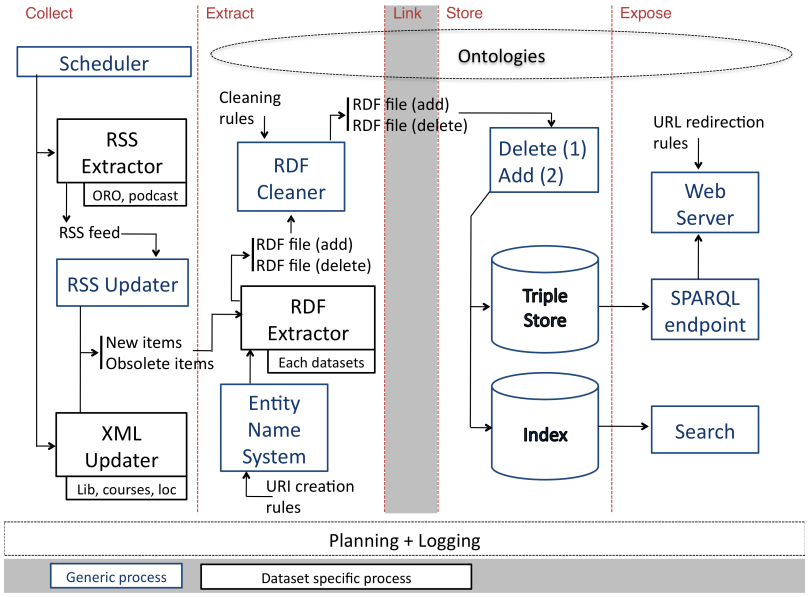

Fig. 1. The LUCERO Workflow

– The Open Research Online (ORO) system[16], which contains information about academic publications of OU research. For that, the Bibliographic Ontology (bibo)[17] is mainly used to model the data.
– The OU Podcasts,[18] which contain Podcast material related to courses and research interests. A variety of ontologies are used to model this data, including the W3C Media Ontology,[19] in addition to a specialised SKOS[20] representation of the iTunesU topic categories.
– A subset of the courses from the Study at the OU website,[21] which provides courses information and registration details for students. We model this data by relying on the Courseware,[22] AIISO[23] and GoodRelations ontologies [4], in addition to extensions that reflect OU specific information (e.g., course assessment types).

  Furthermore, there are other sources of data that are currently being processed. This includes for example the OU list of provided publications, the

[16] http://oro.open.ac.uk
[17] http://bibliontology.com/specification
[18] http://podcast.open.ac.uk
[19] http://www.w3.org/TR/mediaont-10
[20] http://www.w3.org/2004/02/skos
[21] http://www3.open.ac.uk/study
[22] http://courseware.rkbexplorer.com/ontologies/courseware
[23] http://vocab.org/aiiso/schema

library catalogue, and public information about locations on the OU campus (e.g., buildings) and university staff.

4 The OU Expert Search

Expert search can be defined as the task of identifying people who have relevant expertise in a topic of interest. This task 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 [7], 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 [8]. Automating expert finding from such documents provides an efficient and sustainable approach to expertise discovery.

OU researchers, students and lecturers constantly produce a plethora of documents, including for example conference articles, journal papers, thesis, 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.[24]

4.1 Consumed Data

This application is based on two main sources of information: (a) LOD from the Open Research Online system, 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 or 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.

[24] The OU Expert Search is accessible to OU staff at: http://kmi-web15.open.ac.uk:8080/ExpertSearchClient

To exploit this information the system performs two main steps. Firstly when the system receives the user's query, i.e., the area of expertise where a set of experts need to be found (e.g., "semantic search"), the system uses the title and abstract of the publications to find the top-n documents related to that area of expertise. At the moment n has been empirically set to 10.

Secondly, once the top-n documents have been selected, the authors of these documents are extracted and ranked according to five different criteria: (a) original score of their publications, (b) number of publications, (c) type of publications, (d) date of the publications and, (e) other authors of the publication.

The initial score of the publications is obtained by matching the user's keyword query against the title and the abstract of the OU publications. Publications that provide a better match within their title and abstract against the keywords of the query are ranked higher. This matching is performed and computed using the Lucene[25] text search engine. Regarding the number of publications, authors with a higher number of publications (among the top-n previously retrieved) are ranked higher. Regarding the type of publication, theses are ranked first, then books, then journal papers, and finally conference articles. The rationality behind this is that an author writing a thesis or a book holds a higher level of expertise than an author who has only written conference papers. Regarding the date of the publication, we consider the 'freshness' of the publications and continuity of an author's publications within the same area. More recent publications are ranked higher than older ones, and authors publishing in consecutive years about a certain topic are also ranked higher than authors that have sporadic publications about the topic. Regarding other authors, experts sharing a publication with fewer colleagues are ranked higher. The rationality behind this is that the total knowledge of a publication should be divided among the expertise brought into it, i.e., the number of authors. Additionally we also consider the order of authors in the publication. Main authors are considered to have a higher level of expertise and are therefore ranked higher.

To perform the first step (i.e., retrieving the top-n documents related to the user's query) we could have used the SPARQL endpoint and, at run-time, searched for those keywords within the title and abstract properties of the publications. However, to speed the search process up, and to enhance the query-document matching process, we have decided to pre-process and index the title and abstract information of the publications using the popular Lucene search engine. In this way, the fuzzy and spelling check query processing and ranking capabilities of the Lucene search engine are exploited to optimise the initial document search process.

To perform the second step, once the top-n documents have been selected, the rest of the properties of the document (authors, type, and date) are obtained at run-time using the SPARQL endpoint.

Finally, once the set of authors have been ranked, we look for them in the OU staff directory (using the information about their first name and last name). If the author is included in the directory, the system provides related information about

[25] http://lucene.apache.org/java/docs/index.html

the job title, department within the OU, e-mail address and phone number. By exploiting the OU staff directory we are able to identify which experts are members of the OU and which of them are external collaborators, or old members not further working for the institution.

Without the structure and conceptual information provided by the OU LOD, the implementation of the previously described ranking criteria, as well as the interlinking of data with the OU staff directory, would have required a huge data pre-processing effort. The OU LOD provides the information with a fine-grained structure that facilitates the design of ranking criteria based on multiple concepts, as well as the interlinking of information with other repositories.

4.2 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 particularity the title and the abstract of the publications) are periodically indexed using Lucene to speed-up and enhance the search process by means of the exploitation of its fuzzy and spell checker query processing, and ranking capabilities. The rest of the properties (authors, date, and type of publications) are accessed at run time, once the top-n publications have been selected.

The front end is a thin client implemented as a web application using only HTML, CSS and Javascript (jQuery).[26] The client doesn't 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.

4.3 Example and Screenshots

In this section, we provide an example of how to use the OU expert search system. As shown in Figure 2, the system receives as a keyword query input "*semantic search*", with the topic for which the user aims to find an expert. As a result, the system provides a list of authors ("*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 top of the title of the publication, the summary is also visualised (see the example in Figure 2 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 information in the open.ac.uk domain.

[26] http://www.jquery.com

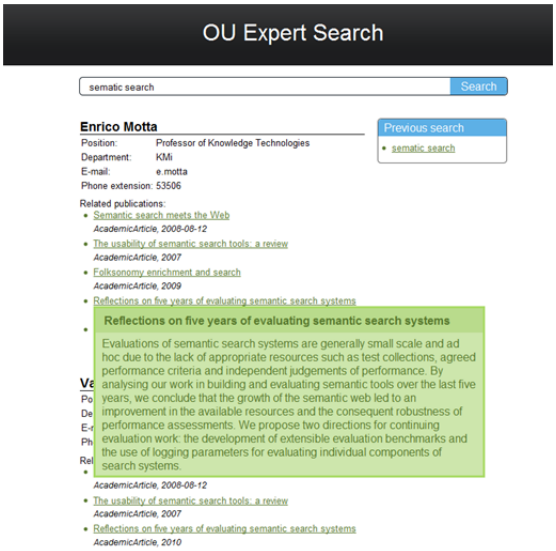

Fig. 2. The OU Expert Search system

5 Buddy Study

The Open University is a well-established institution in the United Kingdom, offering distance-learning courses covering a plethora of subject areas. A key factor in enabling learning and understanding of course materials is support for students, provided in the form of an on-hand tutor for each studied module, where interactions with the tutor are facilitated via the Web and/or email exchanges. An alternative method of support could be provided through peers, in a similar manner to a classroom environment, where working together and explanations of problems from disparate viewpoints enhances understanding.

Based on this thesis, Buddy Study[27] combines the popular social networking platform Facebook with the OU Linked Data service, the goal being to suggest learning partners – so called 'Study Buddies' – from a person's social network on the site together with possible courses that could be pursued together.

5.1 Consumed Data

Buddy Study combines information extracted from Facebook with Linked Data offered by The Open University, where the former contains 'wall posts' – messages posted publicly on a person's profile page – and comments on such wall posts, while the latter contains structured, machine-readable information describing courses offered by The Open University.

[27] http://www.matthew-rowe.com/BuddyStudy

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 most recent n wall posts and the comments on those posts – n can be varied, thereby affecting the later recommendations. Given the extracted content, cleaning is then performed by removing all the stop words, thus reducing the wall posts and comments to their basic terms.

A bag of words model is compiled for each person in the user's social network as follows: for each wall post or comment posted by a given person all the terms are placed in the bag, maintaining duplicates and therefore frequencies. This model maintains information of the association between a user and his/her social network members in the form of shared terms. A bag of words model is then compiled for each OU course in a similar manner: first we query the SPARQL endpoint of the OU's Linked Data asking for the title and description for each course. For the returned information, stop words are removed and the title and description – containing the remaining terms – are then used to build the bag of words model for the course.

The goal of Buddy Study is to recommend study partners to support course learning. Therefore we compare the bag of words model of each person with the bag of words model of each course, recording the frequency and terms that overlap. The user's social network members are then ranked based on the number of overlapping terms – the intuition being that the greater the number of common terms with courses, the greater the likelihood of a course being correlated with the user. Variance of n will therefore affect this ranking, given that the inclusion of a greater number of posts will increase the number of possible study partners, while smaller values for n will yield more recently interacted with social network members. Variance of this parameter is provided in the application.

The application is not finished yet; we still need to recommend possible courses that could be studied with each possible study buddy. This is performed in a similar fashion, by comparing the bag of words model of the social network member with the model of each course, counting the frequencies of overlapping terms for each course, and then ranking accordingly. Due to space restrictions, and to avoid information overload, we only show the top-10 courses. For each social network user, and for each course that is suggested, Buddy Study displays the common terms, thereby providing the reasons for the course suggestion.

If for a moment we assume a scenario where Linked Data is not provided by the OU, then the function of Buddy Study could, in theory continue, by consuming information provided in an alternative form. However, this application forms the prototype upon which for future work – explained in greater detail within the conclusions of this paper – is to be based. Such advancements will utilise concepts for study partner recommendation rather than merely terms, the reasoning behind this extension is to alleviate the noisy form that terms take. By leveraging concepts from collections of terms, recommendations would be generated that are more accurate and better suited to the user in question. Without Linked Data, this is not possible.

5.2 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)[28]. Authentication is provided via Facebook Connect,[29] enabling access to Facebook information via the Graph API. The ARC2 framework[30] is implemented to query the remote SPARQL endpoint containing The Open University's Linked Data, and parse the returned information accordingly.

5.3 Example and Screenshots

To ground the use of Buddy Study, Figure 3 shows an example screenshot from the application when recommending study partners for Matthew Rowe – one of the authors of this paper. At this rank position in the results, the possible study mate is shown together with the courses that could be studied together. The courses are hyperlinked to their resource within the OU Linked Open Data service, and in the proceeding brackets the terms that correlate with the courses are shown. In this instance the top-ranked course is identified by the common terms 'API' and 'Info'.

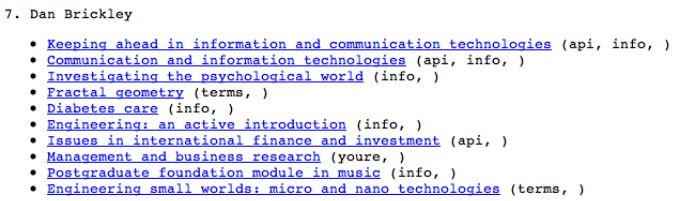

Fig. 3. Buddy Study showing the 7th ranked social network member

6 Linked OpenLearn

The Open University offers a set of free learning material through the OpenLearn website.[31] Such material cover various topics ranging from Arts[32], to Sciences and Engineering.[33] In addition to that, 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

[28] https://github.com/facebook/php-sdk
[29] http://developers.facebook.com/docs/authentication
[30] http://arc.semsol.org
[31] http://openlearn.open.ac.uk
[32] OpenLearn unit example in Arts: http://data.open.ac.uk/page/openlearn/a216_1
[33] A list of units and topics is available at: http://openlearn.open.ac.uk/course

not always explicitly available, making it hard for students to easily exploit all the available resources. For example, while there exists a link 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[34] 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 easily find a linked Podcast, and play it directly without having to go through the Podcast website.

6.1 Consumed Data

Linked OpenLearn relies on The Open University's Linked Data to achieve what was previously considered very costly to do. Within large organizations, it's 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 goals 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 Linked Data SPARQL endpoint (http://data.open.ac.uk/query), 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 application. Finally the list of related tags are 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. The application is still a prototype, and there is surely room for further data to extract. For example, once the library catalogue is made available, a much richer interface can be explored by students with related books, recordings, computer files, etc.

[34] http://fouad.zablith.org/apps/openlearnlinkeddata

6.2 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,[35] 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.

6.3 Example and Screenshot

To install the application, the user has to drag the applications' bookmarklet[36] 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 4 illustrates one arts related OpenLearn unit, 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.

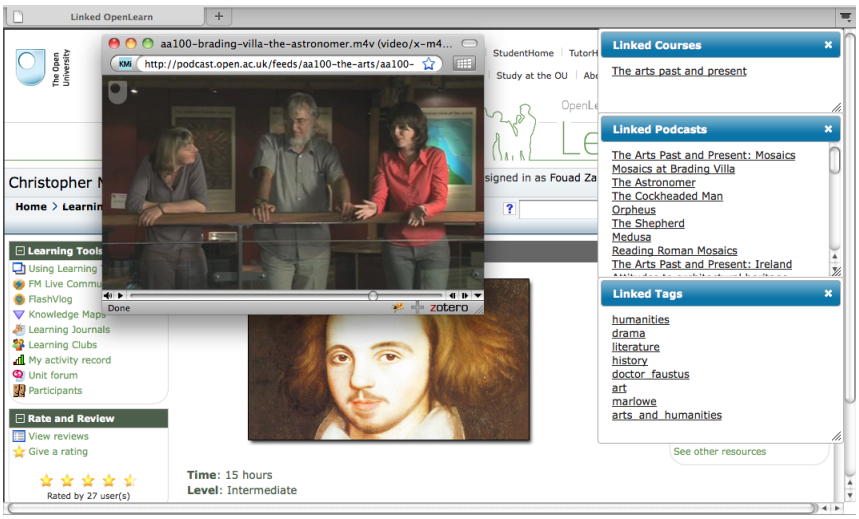

Fig. 4. Linked OpenLearn Screenshot

[35] http://www.jqueryui.com

[36] The bookmarklet is available at: http://fouad.zablith.org/apps/openlearnlinkeddata, and has been tested in Firefox, Safari and Google Chrome

7 Conclusions

In this section we report on our experiences when generating and exploiting LOD within the context of an educational institution. Regarding our experience 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 can 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 assure 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. However, 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.

Regarding our experiences 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 have open opportunities for the development of novel and interesting applications like the three presented in this paper. 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 in a single site, all the information they are looking for.

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 that can be exploited within the education community, 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. We hope that, in the near future, more researchers and developers will embrace LOD approach, by creating new applications and learning from previous experiences to expose more and more educational data in a way that is directly linkable and reusable.

8 Future Work

The application of Linked Data within the OU has opened multiple research paths. Regarding the production of Linked Data, in addition to transforming the library records to LOD, the LUCERO team is currently working on connecting the OU's Reading Experience Database (RED)[37] to the Web of Data. Such database aims to provide access and information about reading experiences around the world. It helps the readership for books issued in new editions for new audiences in different countries to be tracked. Its publication as LOD is an interesting example about how the integration of Linked Data technology can open new investigation paths to different research areas, in this case humanities.

Regarding the consumption of LOD, we envision, on the one hand, to enhance the three previously mentioned applications and, on the other hand to generate new applications as soon as more information is available and interconnected. As example of the former, for the Buddy Study application we plan to extend the current approach for identifying common terms between social network members and courses to instead utilise common concepts. At present the use of online messages results in the inclusion of abbreviated and slang terms, resulting in recommendations that are generated from noise. By instead using concepts, we believe that the suggested courses would be more accurate and suitable for studying. As an example of the latter, we aim to generate a search application over the RED database, able to display search results on an interactive map and link them not just to relevant records within the RED database, but also with relevant objects of the LOD cloud.

References

1. C. Bizer. The emerging web of linked data. *IEEE Int. Systems*, pages 87–92, 2009.
2. C. Bizer, T. Heath, and T. Berners-Lee. Linked data-the story so far. *Int. J. Semantic Web Inf. Syst.*, 5(3):1–22, 2009.
3. T. Heath and C. Bizer. *Linked Data: Evolving the Web into a Global Data Space*. 2011.
4. M. Hepp. GoodRelations: an ontology for describing products and services offers on the web. *Knowledge Engineering: Practice and Patterns*, pages 329–346, 2008.
5. G. Kobilarov, T. Scott, Y. Raimond, S. Oliver, C. Sizemore, M. Smethurst, C. Bizer, and R. Lee. Media meets semantic webhow the bbc uses dbpedia and linked data to make connections. pages 723–737, 2009.
6. M. Rowe. Data.dcs: Converting legacy data into linked data. In *Linked Data on the Web Workshop, WWW2010*, 2010.
7. D. Yimam-Seid and A. Kobsa. Expert-finding systems for organizations: Problem and domain analysis and the DEMOIR approach. *Journal of Organizational Computing and Electronic Commerce*, 13(1):1–24, 2003.
8. J. Zhu, X. Huang, D. Song, and S. Rüger. Integrating multiple document features in language models for expert finding. *Knowledge and Information Systems*, 23(1):29–54, 2010.

[37] http://www.open.ac.uk/Arts/reading