Linked Data for Open and Distance Learning

(Second edition)

Prepared for the Commonwealth of Learning
By Mathieu d’Aquin
July 2014
A REPORT ON:

Linked Data for Open and Distance Learning

Second Edition
July 2014

Prepared for the Commonwealth of Learning by:

Mathieu d’Aquin
Research Fellow Knowledge Media Institute
The Open University, UK

© Commonwealth of Learning
CC-BY-SA

The report is made available under a Creative Commons Attribution-ShareAlike 3.0 Licence (international): http://creativecommons.org/ licences/by-sa/3.0.

For the avoidance of doubt, by applying this licence the Commonwealth of Learning does not waive any privileges or immunities from claims that it may be entitled to assert, nor do the Commonwealth of Learning submit to the jurisdiction, courts, legal processes or laws of any jurisdiction.
# Table of Contents

Table of Contents .......................................................................................................................................... 4

Foreword ....................................................................................................................................................... 5

Executive Summary ...................................................................................................................................... 6

Introduction ................................................................................................................................................... 7

What is linked data........................................................................................................................................ 9

Linked data applications in open and distance learning .............................................................................. 12

The state of linked data in (open) education ............................................................................................... 19

Adopting linked data: Technologies and tools ............................................................................................ 21

Future: Open and distance learning smarter and more connected .............................................................. 24

References................................................................................................................................................... 25
Foreword

Nearly half a century ago a radical movement to open up higher education began when Britain’s new Open University declared that it would be “open to people, to places, to methods and to ideas”. Since then the ideal of openness in public life has made great strides and successive waves of technology have facilitated the adoption of the values of transparency, collaboration and sharing.

Three manifestations of openness are particularly important for education. Open Source Software has increased the quality and diversity of computer applications; Open Access research literature speeds the creation of new knowledge by ensuring that academics build on the work of others; Open Educational Resources (OER), which can be re-used and adapted by anyone, anywhere, have the potential to transform education worldwide with a global pool of excellent learning materials becoming the common wealth of humankind.

A movement that began in idealism has now acquired practical and economic importance. Governments are coming to appreciate the cost savings and quality improvements that OER bring to the educational materials used in their schools and colleges.

But there is a problem. History shows abundantly that unless teachers find a new educational technology easy to use they will not adopt it. Teachers want to teach, not to spend hours searching in the chaos of the Web for material that is appealing, relevant to their needs and educationally credible.

Fortunately some of the world’s most able experts in artificial intelligence, of which Mathieu d’Aquin is one, are now taking up the challenge of creating tools that allow educators to find the needle they want in the haystack of options. The mechanism they are using is the Semantic Web, which may still seem like a recondite and abstruse idea to most Internet users.

The key is the concept of Linked Data, a hot area of development based on the idea that the mechanisms used to share and interlink documents on the Web can be applied to share and interlink data and metadata about these documents, as well as the concepts and entities they relate to. One can thus speak about the Web of Linked Data in the way one talks about the current Web of documents and data.

As the author explains in this report, every “data object” (representing, for example, a person, a place or a topic) is identified by a Web address and characterised using Web links. These can connect to representations of other data objects, identified by other Web addresses. Thus the Web is viewed as a giant data graph that openly draws from any contributing source.

This is a complex topic, engaging some of the world’s best minds. Here Dr. d’Aquin describes the tools, technologies and processes to publish and use Linked Data in a concrete way focusing on learning and teaching applications. As these approaches mature and pass into general use they will give a tremendous boost to the use adaptation of OER.

I am particularly pleased that this work is being done in the Knowledge Media Institute (KMI) at the Open University. The KMI was created in the 1990s to ensure that the UKOU would show the same leadership in integrating the Internet into teaching and learning that it had demonstrated so brilliantly with an earlier generation of technology.

Sir John Daniel, Commonwealth of Learning
June 2012
Executive Summary

Implementing the full vision of Open and Distance Learning (ODL) raises immense technological challenges. The goal is to move from current localised, restricted and locked proprietary content towards the open discovery, use and combination of resources independent from their geographic and institutional origins. It is therefore natural that existing initiatives have taken the web as a base platform, to publish and share open educational resources in the form of online documents. Beyond this first step however, new technological barriers to the full realisation of the ODL vision appear: How to discover these open resources? How to connect resources located in different repositories? How to relate these resources to the context, interest, cultural and technical environment of the learner?

In this report, we introduce a recent development in the area of Web technologies which has the potential to revolutionise the area of ODL: Linked Data. Linked data relies on the simple idea that the mechanisms used nowadays to share and interlink documents on the web can be applied to share and interlink data and metadata about these documents, as well as the concepts and entities they relate to. On the Web of Linked Data, every “data object” (representing for example a person, a place or a topic) is identified by a web address, and characterised using web links that can connect to representations of other data objects, identified by other Web addresses, thus using the web as a giant data graph that openly draws from any contributing source.

Core to this overview of linked data for ODL is the description of applications that have been enabled by the adoption of the technologies and principles underlying linked data. These applications show highly innovative ways in which issues such as the access to and discovery of resources can be made simpler through relying on web standard to obtain and connect data across the web. They also demonstrate how the current state of adoption of linked data by educational institution and distance learning providers, although still preliminary, is already enabling more open and flexible use of the information, and creating new opportunities for application developers to create tools supporting both learners and teachers. Indeed, through analysing the content of the LinkedUp catalogue of web data for education, we can see how the linked data approach enables an ecosystem of data where each individual provider contributes to a common, open and global network, rather than limiting themselves to a silo of information.

Of course, this promising, early states requires some effort from a wider range of organisation to join and adopt linked data. We therefore also describe how this idea is being realised and how it can be adopted. We describe the tools, technologies and processes to publish and use linked data in a concrete way, focusing in particular on learning and teaching applications, and conclude on the need for the area to grow with the use of linked data, and the potential this has to enable a smart “web of educational data”.

Introduction

With the growing use of web technologies, education is currently experiencing massive changes, and increasingly moving towards open and distance learning. MOOCs are the most obvious illustration of this points. Many new platforms have appeared in the last few years, with, although still preliminary, the majority of US and EU universities being involved in one way or another in MOOCs and millions of students from all over the world studying a large variety of subjects through online platforms (see for example Universities UK, 2013). In parallel, open educational resources are becoming more and more mature, providing base material for student and teachers everywhere, often through the web (Hoosen, 2012); and of course, traditional universities are following the trend, providing more convenience and easier access to qualifications for their students, through offering online versions of their qualifications, mixing web-based teaching with more traditional classroom-based teaching. That is coming in addition (and in the continuation) of the model implemented by institutions such as The Open University\(^1\), that were created to give access to higher education and provide university degrees to greater numbers through open and distance learning.

There is however a downside to this rapid development, and the emergence of so many initiatives, from so many different places. Indeed, as often the case when a technology-backed concept is being democratised, this multiplication of platforms and providers of open and distance learning means that (potential) learners are quickly moving from having limited and restricted access to education, to being submerged with options and opportunities. While this is a good thing in principles, with different platforms are being built in isolation, as technological silos, it becomes a challenge to know and assess what is available, match it to one’s own need and, ultimately, make the best of what is available, to the learner’s own benefit. In other words, if we think about the technological side of education, we are now moving from learners being confronted to the neatly controlled and closed information systems of educational institutions, to having to deal with a distributed, emerging network of dedicated, accessible systems, offering much more openness and opportunities, but being also much harder to navigate.

In this report, we investigate how recent linked data technologies can help alleviate this issue, through creating a global, interoperable graph of educational information on the web. The principles behind linked data are relatively simple. It is about making the publication, connection, reuse and access to data as easy as they are for documents on the web. The web indeed provides a global platform where documents can be published, read and linked to each other, without the need of a complex control mechanism, and with the aim to make them easily accessible by anybody, wherever they are, and whatever system they use. Linked data is a set of technologies that make it possible to reuse the same principles to facilitate the publication and reuse of data, providing data objects with URIs (web addresses) and connecting them in a global information graph through web links.

Since the first version of this report (d’Aquín, 2012a), showing early applications and examples of the use of linked data technologies for open and distance learning, many new applications and examples have emerged. More platforms and institutions have also embraced linked data, providing open information about their resources that can connect to others in a way that forms a richer, more useful network for learners around the world. In other words, the “web of linked educational data” is emerging and increasingly becoming a global resource for learners and teachers. We provide here an overview of these developments, of the basic concepts they use in employing linked data, of the applications that can be achieved as a result, and of the way organisations and initiatives, by simply publishing their information

\(^1\) http://www.open.ac.uk
on the web following the linked data principles, can contribute to the growth of open, \textit{linked} education on the web.
What is Linked Data

The foundation of the Web is that it is a network of documents connected by hyperlinks (see Figure 1a). Each document is identified by a Web address, a URI (Berners-Lee et al, 1998), and might represent a document which content is encoded using a standard, universally readable format (most commonly HTML\(^2\)). Following this, the simplest way to describe Linked Data is that it is about using these same principles of the Web architecture not only for documents, but also for data.

![Figure 1](image)

*Figure 1:* The web as a network of interconnected documents (a); Linked data as a network of interconnected data entities (b).

The foundation of Linked Data is therefore that data objects on the Web are identified, similarly to documents, by URIs. The representation of the data—i.e. the information associated with a data object—is then represented by Web links, which can themselves be characterised by URIs. This makes it possible to represent information in such a way that it is materialised as a graph, where nodes are URIs or literal data values (strings, numbers) and the edges are links between them (see Figure 1b).

For example, a university like The Open University publishes information about the courses it offers through its website\(^3\), as well as using linked data\(^4\). It achieves that by assigning to every course a

---

\(^2\) [http://www.w3.org/TR/html401/](http://www.w3.org/TR/html401/)

\(^3\) [http://www.open.ac.uk/courses](http://www.open.ac.uk/courses)

\(^4\) [http://data.open.ac.uk/page/context/course](http://data.open.ac.uk/page/context/course)
A dedicated URI that acts both as an identifier for the course on the Web, and as a way to access structured information about this course. For example, http://data.open.ac.uk/course/aa100 is the URI for the course with code AA100, which is an undergraduate (level 1) course in Arts and Humanities, entitled “The arts past and present”. Through the links between this URI and others, information about this course is being represented regarding the topics and description of the course, where it is available, how it is assessed, what course material and open educational resources relate to it, etc. (see Figure 2). While most of the other data objects it relates to are also identified by URIs within the domain of the Open University, it is important to remark here that it links to other data sources, such as the UK government’s information about the Open University5 or information provided by the Geonames platform6 about the countries in which the course is available. This demonstrates how, from these basic principles, information originating from widely different systems and sources can be seamlessly integrated.

![Figure 2: Extract of linked data representation of the Open University course AA100.](image)

While linked data has, for a few years, been confined in the corresponding academic research community, it is now gaining significant momentum in many different areas (see Figure 3 showing a catalogue of interlinked datasets). Governments (most notably in the US7 and the UK8) are in particular leading open data initiatives where information about aspects such as transport, environment, public spending and education are being made available as Linked Data, to be easily referenced, accessed and reused by the public. Datasets, such as the one of the geonames initiative are also being made available as Linked Data, making it possible for anybody to link to reference information regarding geographical places in the world. One of the most reused source of information on this Web of Data is DBpedia9, a Linked Data version of Wikipedia10 providing easily reusable and query-able general information in almost every domain. DBpedia basically transforms every page in Wikipedia into a linked data entity, analysing the “infobox” on the Wikipedia page to extract structured information about each entity, and their relationships to others. Many cultural organisations such as museums and libraries11 have also realised the

---

6 [http://geonames.org](http://geonames.org)
7 [http://data.gov](http://data.gov)
8 [http://data.gov.uk](http://data.gov.uk)
9 [http://dbpedia.org](http://dbpedia.org)
10 [http://www.wikipedia.org](http://www.wikipedia.org)
11 see the British Library, the BNF and generally, the Europeana project - see [http://www.europeana.eu](http://www.europeana.eu)
potential of the linked data, and started publishing open data, including for example the content of their catalogues and collections, as Linked Data. Finally, Linked Data is more and more used by universities and other education institutions to make their public information and open resources more accessible, discoverable and reusable (this report includes several examples in this area).

Figure 3: The Linked Open Data Cloud - Linked data sets and their connections (see http://lod-cloud.net/).
Linked Data Applications in Open and Distance Learning

The previous section presents a general overview of the principles of linked data, with the idea that it can help sharing and connecting information across the global network of the web. However, to truly understand the benefits of adopting linked data, the most straightforward way is to look at the innovative applications and services that have already been developed using it. We therefore discuss here several illustrative examples, showing the variety of tasks and issues in which linked data can help for open and distance learning, and for education in general. These examples should not however be seen as a definitive answer to the question “what can be done with linked data for open and distance learning?” Indeed, this question is still open, as new ways to exploit the growing information base available as linked data on the web are emerging constantly. Actually, several of the applications described below were submitted to the LinkedUp Challenge\(^\text{12}\) (see d’Aquin et al., 2014b), which goal is to push forward the use of web data for education.

To illustrate the use of linked data for open and distance learning, we start with one of the simplest and most straightforward example: Supporting the publication and exploration of information about available courses and resources. An example of such a use is the “Study At the OU” mobile application from the Open University\(^\text{13}\). “Study at the OU” is the website of the Open University that contains the description of the courses and qualifications that can be obtained from the University\(^\text{14}\). The corresponding mobile application exposes the course catalogue and additional information about the topics covered by the available courses to smart-phones and tablets (see Figure 4). As part of this application, it is possible to select a topic and obtain information both about the courses available on this topic, and about the related resources such as podcasts, Youtube videos and OpenLearn units. This last feature is implemented using the Linked Data platform of the Open University\(^\text{15}\), querying resources that are directly related to the topic being considered, or for resources attached to courses that are related to this topic. In other terms, it transparently delivers relevant links to heterogeneous resources by integrating these resources under a common Linked Data representation (d’Aquin, 2012b).

\(^{12}\) http://linkedup-challenge.org
\(^{13}\) https://market.android.com/details?id=uk.ac.open.saou for Android and http://ax.itunes.apple.com/gb/app/studyatou/id403500460?mt=8#ls=1 for iOS
\(^{14}\) see http://www3.open.ac.uk/study/
\(^{15}\) http://data.open.ac.uk
Figure 4: Screenshots of the “Study at the OU” mobile application.

The advantage of Linked Data here is that the rich structure of the linked data-based, graph description of the resource metadata can support a range of navigation paths, using information that can spread across multiple systems and sources. Another advantage is that, by abstracting from the specificities of systems and formats in which the resources might be handled, it provides a way to homogeneously integrate heterogeneous resources.

While the previous example shows the benefit of linked data in connecting information within an organisation, in this case the Open University, the real power of the approach is when such information, especially about available resources, can be exploited that spread across several organisations and systems. In this spirit, the Solvonauts\textsuperscript{16} search engine is an example of the way connecting metadata about resources available in several online systems can help resource discovery, integrating in an homogeneous way results from different origins, and in different formats.

However, linked data can also make possible much more sophisticated scenarios. One of them, not unrelated to the applications described above, is resource recommendation. Typically, resource recommendation is based either on the user profile in the system holding the resources (and the profile) or on some form of similarity (of content or usage) between resources of interest to the user. However, this is not an easy task to achieve when trying to recommend resources of a certain type, held in one particular system, from resources of another type, held in another system. Through linked data however, it is possible to process information about these different resources homogeneously, independently of the systems, and to relate them conceptually using other, reference data.

\footnotesize\textsuperscript{16} http://solvonauts.org/
Figure 5: Screenshot of DiscOU (overlay window) used on iPlayer, showing 10 recommendations of open educational resources relevant to the currently shown programme.

This is the scenario addressed by the DiscOU application (see d’Aquin et al., 2012; and Figure 5). DiscOU is designed to provide recommendation of open educational resources from the Open University, based on a (potential) learner’s interest in a TV or radio programme on the BBC. Open educational resources at the Open University include pieces of course material and small articles from the OpenLearn system, as well as multimedia material in the form of audio and video podcasts. These, resources are described through linked data on the Open University platform (data.open.ac.uk), including links to the full content (and transcripts in the case of podcasts). Programmes are also described as linked data on the BBC website, which are accessible from the programme page, or the iPlayer web system (the video player of the BBC). DiscOU takes the form of a bookmarklet, a link in the bookmark bar of the browser that can overlay functions and interfaces over the currently displayed website. When activated while displaying a BBC programme or iPlayer page, the tool extract from the corresponding linked data information about the programme, connect this information to concept in DBpedia, and find in the open educational resources of the Open University resources that also relate to these concepts, through pre-established connections between DBpedia and their linked data-based descriptions.

17 http://discou.info
18 http://www.open.edu/openlearn/
19 http://podcast.open.ac.uk/
20 http://www.bbc.co.uk/iplayer
What is interesting here is not only that linked data can help establishing a connection between two completely different information systems, held by two different organisations and with completely different purpose, and use these connections to facilitate access to (open, free) educational resources. It is also that the whole process of recommending resources, is generic and based on the semantic relationships that exist between the resources. It is therefore possible to make explicit these relationships to the user - in DiscOU, in the form of the explanation under each recommendation: “Also about X”. What’s even more interesting is that the user (learner) can even interact with these relationships, tuning the weight of the concepts resources are connected to based on their interest, and obtaining personalised recommendations (see Figure 6).

![Figure 6: Customising the search in DiscOU by choosing the importance of the concepts of relevance to the displayed programme.](image)

Another common scenario where the web has introduced massive changes, and where linked data has a strong potential to help for open and distance learning, is collaboration. Indeed, linked data represents a natural model through which different people and organisations can contribute information that connect with each other, therefore creating a global base of information out of these smaller contributions. This is
the scenario addressed by the *We-Share* application\(^\text{21}\) (Ruiz-Calleja et al., 2013), for ICT tools in education (see Figure 7). Indeed, the aim of this application is basically to build a tool repository to help educators select the right tool in the right situation. This might include any type of ICT tool, and the description of each tool includes information not only about the tool and its features, but also about the contributor and the educational scenario in which it has been used. The information is then available for others to reuse, through a web interface and direct access to the created linked data.

![Screenshot of the We-Share application to explore ICT tools and their use in educational scenarios.](image)

**Figure 7:** Screenshot of the We-Share application to explore ICT tools and their use in educational scenarios.

The benefit of using linked data in data collection here is not only that the pieces of information that are being produced can be easily put together and re-distributed to others (or linked). It is also that it makes it easier to reuse information that already exist. Indeed, each tool can be described once, and whenever a new educational scenario applies, be simply pointed to (and possibly the description extended), contributing only the part of the information that is specific to the considered used. This is valid internally as well as externally: basic information about the tools are available in other datasets such as DBpedia. Each tool description in the system can not only connect to to existing, internal tool descriptions, but also to these external tool descriptions, enriching the information provided by contributors to the system with general information built for other purposes.

Other systems have used a similar approach to collaboration and reuse on the basis of linked data at a much larger scale. These include the GNOSS platform in Spain\(^\text{22}\), that acts as a social network of students and teachers, to exchange, share and describe (educational) resources; or the VIVO platform\(^\text{23}\) (Borner et

\(^{22}\) [http://www.gnoss.com/home](http://www.gnoss.com/home)
al., 2012) which exploits linked data technologies to collect information within universities, that can then be opened through linked data to others, forming a network of resources from many different universities.

Of course, a way in which linked data can help is also in providing easy access to information for learners about the subject matter of their study. While it is hard to find a generic application of this sort, they are many examples of visualisations and information exploration tools that use web, open data to support learners in a particular domain. For example, Globe Town24 (Townsend et al., 2013) focuses on sustainable development and uses open data to “explore the intersections, tensions and trade-offs between the ‘three pillars’ of sustainable development: the environment, the economy and society”. PoliMedia25 (Kleppe et al., 2013) is another example where linked data is used not necessarily in a way that is explicitly related to education, but that presents information about a certain specific topic online, through aggregating and combining information available as linked data, and therefore helps learners in their study of this particular topic. It is designed to facilitate large-scale, cross-media analysis of the coverage of political events, focusing on the meetings of the Dutch parliament. It achieves this by automatically generating links between the transcripts of those meetings, newspaper articles and radio bulletins.

Finally, another area where linked data is having a growing influence is in the emerging domain of learning analytics (d’Aquin et al., 2014a). Learning Analytics is about the processing of data about learners and their environments for the purpose of understanding and optimising learning (see for example, Ferguson, 2012). A lot of both the research-oriented and the practical work in this area is dedicated to the methods employed for collecting, analysing, mining or visualising such data in relation to various levels of models of learning, from the basic information models used to structure the data, to the cognitive models that are expected to be reflected in the learners activity patterns found in the data. As such, learning analytics introduces specific challenges with respect to the management of the data used. In particular, Learning Analytics might require to combine and connect various data sources from different origins, through trying to gain insights on the connections between different aspects of the learning experience and the learner’s context.

However, the benefits of using linked data in learning analytics are just starting to be considered, mostly within the research community dedicated to learning analytics (see d’Aquin et al, 2014a; d’Aquin and Jay, 2013). Indeed, while linked data simplifies access to information, integration with the tools required for analytics is still preliminary. Interesting examples can be demonstrated, including for example using R26 to do statistical analysis on linked data sources (see Figure 8a for an example) or, with ad-hoc development of visualisations (see Figure 8b for another example). In many cases however, linked data can be used to enrich the data to analyse or to provide background information for its interpretation (as for example in d’Aquin and Jay, 2013).

24 http://globe-town.org/
25 http://www.polimedia.nl/
26 http://www.r-project.org/
Figure 8: Distribution of courses in different topics, using R and the data.open.ac.uk endpoint (see d’Aquin et al., 2014a) (a); Gender inequality related to education: yellow countries indicate countries where men stay more in education, blue where women stay more in education and green where they are approximately equal. Built using UNESCO data on the uis.270a.info endpoint and D3.js (see Tiddi et al., 2014) (b).
The State of Linked Data in (Open) Education

There is a growing base of open educational content being made available online, both in education-specific resource repositories and in general information sources of relevance to learning and teaching. Having such content accessible and discoverable on the web has a significant potential impact on the way education is delivered and received. Information consumed by educational environments can take many different forms, beyond the base material for teaching and learning (including for example repositories of research articles and descriptions of educational facilities), and the open exposure of such data following the principles of open, linked data is expected to make educational information easier to address, aggregate and reuse for various purposes.

Accordingly, linked data principles are increasingly being adopted by educational institutions in different contexts. Indeed, institutions and initiatives have started using these approaches to expose open data of educational relevance. These include universities, schools and research centres (see LinkedUniversities.org), government agencies (see for example education.data.gov.uk), and projects around specific domains (see for example mEducator27).

In (d’Aquin et al., 2013), we described how, as part of the LinkedUp Catalogue of Datasets for Education28, we collected and analysed datasets that explicitly reference education and are exposed as linked data. At that time, the catalogue contained 146 datasets from 22 different sources. Some datasets originate from universities (e.g. data.open.ac.uk, data.southampton.ac.uk, data.aalto.fi); others from publication repositories (e.g. data.nature.com, dblp.l3s.de), government agencies and standardisation bodies (e.g. education.data.gov.uk), or as the output of specific projects (e.g. meducator.open.ac.uk, data.organic-edunet.eu). All together, the datasets use 588 different types of entities and 719 relations.

While the numbers above do not, by themselves, give us much insight into the current state and growth of linked data in education, they show that, even at this early stage, initiatives are being formed that try to individually exploit the benefits of linked data to expose information of educational relevance. More importantly, as discussed in details in (d’Aquin et al., 2013) and shown through the graphs of Figure 9, each of these initiatives constitutes a contribution to a growing network of data on the Web, which can be used for educational applications, and making open and distance learning easier to implement. Interestingly as well, the LinkedUp catalogue is also growing, moving from 22 sources in the initial study to 42 now, reflecting the increasing adoption of linked data in the broad context context of education, with a specific emphasis on open data, and open learning.

27 http://www.meducator.net/
28 http://data.linkededucation.org/linkedup/catalog/
Figure 9: Connection between datasets in the LinkedUp catalogue that share types of objects (a); Most common types of objects, connected based on their co-occurrence in datasets of the LinkedUp catalogue (b).
Adopting Linked Data: Technologies and Tools

While the “what is Linked Data” section above presents the basic principles of linked data, several applications have been described and the current state of the use of linked data in open and distance learning has been discussed, it is worth spending a bit of time before concluding this report to describe more precisely the technologies that are associated with the implementation of linked data. The goal here is not to give a complete guide to these technologies (books such as Heath and Bizer, 2011 can be used for this), but to give an overview of the types of tools and formats that one might have to encounter when adopting linked data, as an application developer or as a publisher of open data.

RDF29 (Resource Description Framework) is the base format for the representation of linked data on the web. Since it is meant for the representation of data that can connect across different sources on the web, RDF naturally follows a graph model, where data is represented as nodes, connected through edges. Nodes can be either resources, or literals (i.e. values such as a string or a number) and edges connect these resources or literals. An important aspect here is that each resource and each relationship is identified by a URI, i.e. a web address that point to the information about the corresponding entity. Taking the example of a unit of open educational content from the OpenLearn repository, we can represent the fact that it is a document, has a title (“Machines, minds and computers”), is published by the Open University and relates to a course called “M366: Natural and Artificial Intelligence”, through the set of triples:

<http://data.open.ac.uk/openlearn/m366_1>
  <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://data.open.ac.uk/openlearn/m366_1>
  <http://purl.org/dc/terms/title> "Machines, minds and computers". 
<http://data.open.ac.uk/openlearn/m366_1> <http://purl.org/dc/terms/publisher> 
  <http://education.data.gov.uk/id/school/133849>.  
<http://data.open.ac.uk/openlearn/m366_1>
  <http://data.open.ac.uk/openlearn/ontology/relatesToCourse> 
    <http://data.open.ac.uk/course/m366>. 
<http://data.open.ac.uk/course/m366> <http://purl.org/dc/terms/title>
  "Natural and artificial intelligence"

While these triples might relate information from various sources across the web, they need to be stored and managed within dedicated systems. Triple stores are the equivalent to database management systems in relational databases. They are software systems that provide functionalities to load, store, update and query data in RDF. They are called triple stores as their underlying data model is a graph made of RDF triples. Contrary to usual database systems, they interact with external applications through linked data standards (RDF and SPARQL, described next), therefore hiding the details of their specific implementation and making it possible to use different triple stores in a homogenous way. Existing systems differ mostly with respect to their hardware requirements, their scalability and performance, as well as whether they provide additional features beyond what is strictly required to realise the linked data standards.

---

29 [http://www.w3.org/RDF/]
**SPARQL link** (Simple Protocol and RDF Query Language) is the query language for RDF and generally Linked Data. It has a similar role in triple stores as SQL in relational database management systems. Major differences with SQL however include that SPARQL is designed to query graph-based data representations. It is therefore essentially based on the definitions of “triple patterns” corresponding to particular conditions in the data graph. For example, the following query (adapted from the examples at [http://data.nature.com/query](http://data.nature.com/query)) gives the title and first author of up to 25 articles that contain the word “an” in their title:

```sparql
select ?title ?name where {
    filter (regex(?title,"an","i")) .
    ?article <http://purl.org/ontology/bibo/contributorList>
    [<http://www.w3.org/1999/02/22-rdf-syntax-ns#first>
     [<http://xmlns.com/foaf/0.1/name> ?name ]].
} limit 25
```

Another important aspect of SPARQL is that it does not only include a query language, but also a protocol to create data endpoints on the Web. The idea is that a SPARQL endpoint (such as the one from nature.com mentioned above) should only require standard web mechanisms in order to be accessed and used. In other terms, one only needs a web connection to query a SPARQL endpoint. Also, results are given in standard web formats such as XML, or even RDF, making them Web accessible and linkable. Finally, the SPARQL language includes several different types of queries that can be used to interrogate the data in a triple store (“Select” queries as above, or “Ask” queries) or to extract a sub-graph from the data (“Construct” or “Describe” queries).

These technologies represent the basic, lower level layer of linked data - dedicated to data management, and that is necessary to understand and possibly adopt linked data in an open and distance learning context, or any other context where it might be relevant. Going beyond this however, a layer above these basic data management mechanisms include the one of vocabularies, that relates to conventions and schemas for modelling linked data in a way that makes them easily interoperable. Indeed, while in principle data on the web and the RDF format, do not require a schema in the typical sense of a relational database system, vocabularies and ontologies in linked data play a similar role, with the aim to define shared ways to structure and meaningfully organise data in RDF so that these data can be reused. The base format for representing such vocabularies is **RDF Schema**\(^{30}\), which makes it possible to declare the types of objects (the classes) and the relationships (properties) that might apply between objects of different types. OWL (the Web Ontology Language\(^{31}\)) goes a step further, making it possible to define these classes and properties using logical statements. Such logical statement should represent the shared meaning, the semantics, of these classes and properties, therefore providing a way to attach this meaning to the data for others to reuse, and possibly to reason upon.

Several of such web vocabularies and ontologies exist that are specifically dedicated to the representation of data about education, learning and teaching (see the vocabularies page of [linkeduniversities.org](http://linkeduniversities.org) for details). For example AIISO (the Academic Institution Internal Structure Ontology\(^{32}\)), is dedicated to the representation of information about departments, faculties and other divisions of an educational institution, as well as their relationships with courses and qualifications. LRMI (Learning Resource Metadata Initiative\(^{33}\)) is another of such examples. It has been designed as an extension of Schema.org\(^{34}\),

---

\(^{30}\) [http://www.w3.org/TR/rdf-schema/](http://www.w3.org/TR/rdf-schema/)

\(^{31}\) [http://www.w3.org/2001/sw/wiki/OWL](http://www.w3.org/2001/sw/wiki/OWL)

\(^{32}\) [http://vocab.org/aiiso/schema](http://vocab.org/aiiso/schema)

\(^{33}\) [http://www.lrmi.net/](http://www.lrmi.net/)
and as such can be used as a linked data vocabulary. Its aim is to provide a common representation schema for metadata related to educational resources.

One important set of tools are the ones that can help transforming existing, legacy data into linked data. Some platforms specifically dedicated to education and eLearning already provide features to export metadata about their content in RDF. For example, in Moodle, the RDF description of the content of a page can be obtained by adding ".rdf" at the end of its address. The Fedora Content Management System, commonly used for multimedia repositories, includes a similar feature, and is actually based on a triple store providing SPARQL-based querying mechanisms. The ePrints open source digital repository also includes features to export data into a variety of formats, including RDF.

When not provided directly by the platform that originally holds the information to be extracted and transformed, tools to realise these tasks might be available separately. For example, Triplify and D2RQ are two of the most popular tools used to create mappings between relational databases and specific RDF representations. Other tools such as Open Refine (a tabular data cleaning and manipulation tool) also include extensions to export data that can originate from a large variety of tabular formats (Excel, Google Spreadsheet, Comma Separated Value files) into RDF.

Last, but most definitely not least, an important aspect of adopting linked data is, of course, linking. Links between different datasets can take two forms: the reuse of external URIs, or declaring that some resources in the considered data are the same as resources in others. The first approach is somehow the most natural, creating links by simply relating resources in one source to another somewhere else on the web using a dedicated property. For example, the description of courses at the Open University include information about the countries in which they are available. These countries are referenced by using URIs from the Geonames dataset. When this might not be suitable, one can use the second approach, relying on the ‘owl:sameAs’ property of the OWL language to declare that two different linked data entities (having two different URIs) represent the same real-world object. For example, in the Linked Data platform of the University of Southampton, entities representing bus-stops (for example http://id.southampton.ac.uk/bus-stop/SN120684) are linked through ‘owl:sameAs’ to corresponding entities in the UK government's transport linked dataset (for example http://transport.data.gov.uk/id/stop-point/1980SN120684).

In both cases, if not created manually, the links are often created through ad-hoc scripts and tools. However generic tools such as SILK and LIME exist that can be set up to automatically discover such links between datasets.

34 http://schema.org/
35 https://moodle.org/
36 http://www.fedora-commons.org/
37 http://www.eprints.org/
38 http://triplify.org/Overview
39 http://d2rq.org/
40 http://openrefine.org/
41 http://refine.deri.ie/
42 http://wifo5-03.informatik.uni-mannheim.de/bizer/silk/
43 http://aksw.org/Projects/LIMES.html
Future: Open and Distance Learning
Smarter and More Connected

As shown above, while behind when compared to other areas such as eGovernment, open and distance learning is showing an increasing adoption of linked data, as a base technology and principle to publish, share and reuse data across the web in an open way. The number of initiatives creating and publishing linked data of relevance to education is increasing rapidly, as demonstrated in community efforts such as the LinkedUp catalogue of educational linked data. Also, more and more examples of innovative use of the available linked data are appearing, showing how relevant tasks and services are becoming easier and less costly to deploy, and how new applications that can only be developed with the availability of web scale, connected information are now a reality.

This is an exciting time as, as impressive as these developments might be, we are still only at a very preliminary phase. Indeed, open and distance learning has a lot to benefit from linked data. After all, both have for goal to enable the sharing and distribution of information and knowledge as widely as possible, with the least possible obstacle so that all can benefit and contribute (d’Aquin and Dietze, 2014). However, we are still a long way away from having every provider of open and distance learning, every MOOC, or every OER repository publishing linked data about their courses and contents, and every student having access to the range of innovative applications that this would enable. This is a slow process, but one which is sure to happen, as surely as the web has slowly transformed the area in such a way that, nowadays, it is almost impossible to distinguish “open and distance learning” from “online learning” (although considering them equivalent would still be significantly inaccurate).

While the future cannot be predicted, it appears that, for what concerns data management and information sharing, linked data is generally the way to go for open and distance learning. One aspect not much covered in this report relates to how such an approach can also bring more intelligence to the area, enabling smart applications to support both learners and teachers. Indeed, linked data as a technology-related area is strongly connected to the semantic web - the vision of a web where information and knowledge are available with clearly defined meaning, so that they can be manipulated, processes and reasoned upon by software agents. Both linked data and the semantic web have made a lot of progress as academic disciplines recently. Thanks to this, applications that can intelligently support learners in organising their studies, finding the right courses, connecting these courses to relevant cultural resources, creating study pathways and relating them, for example, to an employment objective on the basis of vast amount of knowledge directly accessible online can be foreseen as becoming reality, and even commonplace, in a not too distant future.
References


d'Aquin, M. (2012b) Putting Linked Data to Use in a Large Higher-Education Organisation, In Interacting with Linked Data workshop at Extended Semantic Web Conference, ESWC


