LinkedUp: Linking Open Data for Education

Document status and date:
Published: 01/02/2014

Open Universiteit
www.ou.nl

LinkedUp: Linking Open Data for Education

4 February 2014 - 1:12pm

In the past, discussions around Open Education have tended to focus on content and primarily Open Educational Resources (OER), freely accessible, openly licensed resources that are used for teaching, learning, assessment and research purposes. However Open Education is a complex beast made up of many aspects, of which the opening up of data is one important element.

When one mentions open data in education a multitude of questions arise: from the technical (what is open data? What is linked data? How do I create open datasets?), the semantic (what is the difference between Open Education data and open data in education?) to the more philosophical (what exactly is Open Education anyway? How can we make sure 'open' means 'accessible to all'? How can opening up data be helpful?) All valid questions, yet not all with straight-forward answers; however exploration around what might purport to be answers to these questions is very much in scope for the LinkedUp Project.

The LinkedUp Project (Linking Web data for education) [1] is an EU FP7 Coordination and Support Action running from November 2012 to November 2014 which looks at issues around open data in education, with the aim of pushing forward the exploitation of the vast amounts of public, open data available on the Web. It aspires to do this by facilitating developer competitions and deploying an evaluation framework, which identifies innovative success stories of robust, Web-scale information management applications. The project comprises six pan-European consortium partners [2] led by the L3S Research Center of the Gottfried Wilhelm Leibniz Universität Hannover and consisting of the Open University UK, the Open Knowledge Foundation, Elsevier, the Open Universiteit Nederland and eXact learning LCMS. The project also has a number of associated partners [3] with an interest in the project

including the Commonwealth of Learning, Canada and the Department of Informatics, PUC-Rio, Brazil.
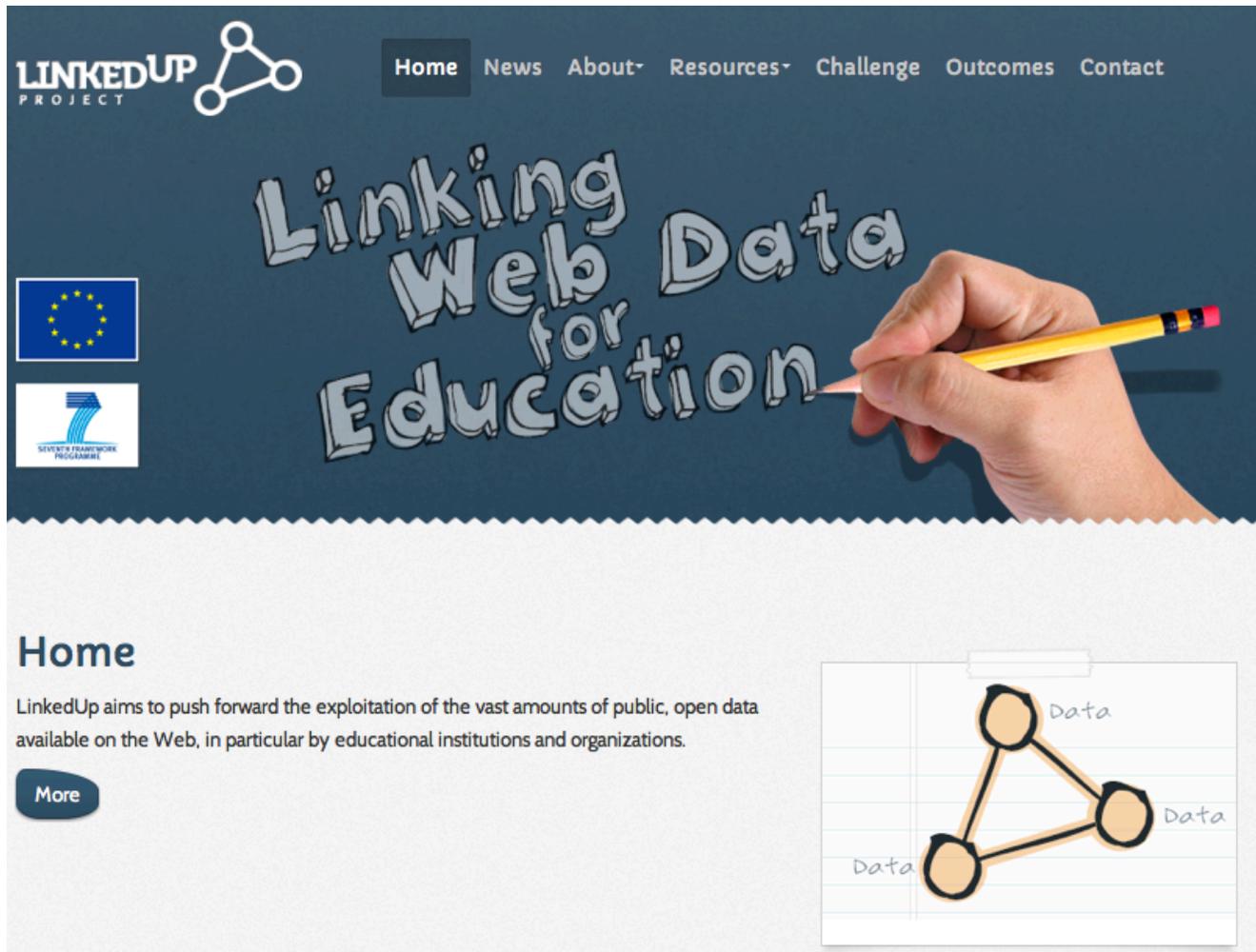

Figure 1: The LinkedUp Web site

Open and Linked Web Data

The LinkedUp Project focuses on open Web data and has its roots in the linked data movement. The project recognises that while World Wide Web began as a global space in which to link primarily documents, through the development of Web standards and the inclusion of semantic content in Web pages there is now an increasing need for access to raw data that sit separately from documents. Publishing these data in a structured way as linked data, through the use of URIs and RDF, provides an opportunity for these data to become much more useful.

As Tim Berners-Lee explains:

> 'The Semantic Web isn't just about putting data on the web. It is about making links, so that a person or machine can explore the web of data. With linked data, when you have some of it, you can find other, related, data.' [4]

The Semantic Web is an overarching concept, a common framework that allows data to be shared and reused across applications, and Linked data is one part of this framework. Many use the terms semantic web synonymously with linked data. For a lay-person the easiest way to understand both the Semantic Web and linked data is through the idea of a 'web of data'. So for example if you searched for lecturers at an institution you could find their name, but you could also find all the papers that they have written, where those papers were published, definitions of all the topics they cover, details of all the other universities involved, and so on. Searching is considerably enhanced by semantics: you would know that the author with the very commonplace name John Smith you searched for was the lecturer at 'University A' rather than the one at 'University B' because all key elements have unique identifiers. The result is a

much more intelligent system than the current Web. There are many better concrete examples out there, the Ordnance Survey linked data video [5] offers a good introduction.

So are people already creating linked data? The book: *Linked Data: Evolving the Web into a Global Data Space* [6] reports that there are 50 billion facts published as linked data on the Web today, while the W3C notes that in September 2011 there were a billion RDF triples, interlinked by around 504 million RDF links [7]. These data cover most academic domains, cross-domain datasets (such as DBpedia, a crowd-sourced community initiative to extract structured information from Wikipedia [8]) and governmental data [9]. In addition to the more technology-centric publication of data, many initiatives have emerged recently that follow the more general principle of open data, from governments and public institutions to local and private institutions (eg the *Guardian* data blog [10]).

Linked data must adhere to the four principles of linked data outlined by Tim Berners-Lee in his Design Issues Linked Data note [11]. They could be summarised as:

- use URIs to denote things;
- use HTTP URIs so that these things can be referred to and looked up;
- provide useful information about the thing when its URI is dereferenced;
- leverage standards such as RDF, SPARQL and;
- include links to other related things (using their URIs).

But when they are published on the Web, open data do not necessarily have to be structured, but they have to be open. The Open Definition [12], developed by the Open Knowledge Foundation [13], sets out principles that define 'openness' in relation to data and content. It can be summed up in the statement that: 'A piece of data or content is open if anyone is free to use, reuse, and redistribute it — subject only, at most, to the requirement to attribute and/or share-alike.' The LinkedUp Project advocates use and creation of linked data and recognises the value of open data.

Open and Linked Data in UK Higher Education

In the education sector, the benefits of using open and linked Web data are starting to show with several universities engaged in the deployment of linked data approaches. In the UK this has been driven by a requirement for transparency and accountability in public institutions, directed by government. However there is also a relatively recent acknowledgement that sharing data not only allows comparison between individual institutions and cluster groups but can also inform decision-making. The creation of innovative tools, as supported through LinkedUp activities, can bring together different datasets and offer new perspectives. These new perspectives are nicely illustrated in a UK project on sharing equipment data. The UK University Facilities and Equipment Open Data project [14] was funded by a UK research council in response to a need to facilitate discussions around equipment sharing among UK universities. It delivers a national 'shop window' for research equipment and supports searching across UK HE equipment datasets. The service forms part of the data.ac.uk initiative formed by the community of UK university open data projects. One institution that has been doing some really interesting work in this area is the University of Southampton [15]. Staff there have created an open data service and brought together datasets on many aspects of university life such as locations, course details, people and travel information. The result is many interesting views on the data that allows you, for example, to find out what coffee shops are open on campus, what they sell and how to locate them. Or what courses are on at what time and whether students rate them highly. Another interesting project in this area is mEducator [16] which provides linked data on open educational resources and has led to community platforms such as LinkedEducation.org [17], an open platform which promotes

the use of linked data for educational purposes. One of the project's aims is to identify best practice and potential links between individual resources in order to contribute to a well-connected educational Web of Data.

For those involved in the LinkedUp Project it is clear that the availability of open teaching and education-related data represents an unprecedented resource for students and teachers. It has the potential to introduce a paradigm shift in the way educational services are provided, substantially improve educational processes and lower the costs of providing Higher Education. Nonetheless, so far, the potential of using educational Web data has been vastly underexploited by the educational sector. Applications and services often only make use of very limited amounts of data and distributed datasets; nor do they provide users with an appropriate level of context and filtering for the vast amounts of heterogeneous content retrieved to make it possible for such information to be adequately exploited. The LinkedUp Project hopes to engage with communities working in this area, and also with others who have yet to see the potential of open and linked data for educational purposes. Its aim is to encourage more activity in the open and linked data arena, in particular by educational institutions and organisations.

The LinkedUp Challenge

One of the principal ways it intends to encourage engagement is through a series of open competitions designed to elicit Web data-driven applications for personalised, open and online university-level studies. The LinkedUp Challenge [18] is a series of three consecutive competitions which seek interesting and innovative tools and applications that analyse and/or integrate open Web data for educational purposes. The competitions are open to anyone, from researchers and students, to developers and businesses. The second and third competitions will build upon their predecessor; leading from innovative prototypes and tools through to large-scale deployable systems. Participants are required to solve critical issues with respect to Web-scale data and information discovery and retrieval, interoperability and matchmaking, data quality assurance and performance. The challenge builds on a strong alliance of institutions with expertise in areas such as open Web data management, data integration and Web-based education.

The first competition (Veni) ran from to 22 May to 27 June 2013. Extensive promotion was carried out using Twitter, blog posts and pan-European mailing lists. By the closing date 22 valid submissions had been received from 12 different countries (four from the UK, three from France, three from Spain, three from the USA, 2 from the Netherlands and 1 each from Argentina, Belgium, Bulgaria, Greece, Italy, and Nepal). The abstracts are available from the LinkedUp Challenge Web site [19]. The majority of entries were from teams based at universities or from start-up companies, but there were also a few from independent consultants. Some entries were developed by large teams, for example one had 9 people listed as authors and others had authors spread across different countries and organisations, while other entries had sole authors.

Entries and Judging

The entrants to the competition had interpreted the specification 'educational purposes' in a variety of innovative ways. A number of the entries had looked at Massive Open Online Courses (MOOCs) and course data and offered cross-searching mechanisms while others had concentrated on discipline-specific data and offered new pedagogical approaches for learners to explore and understand subjects. Two of the submissions focused on cultural heritage data and how museum data could be used in an educational context. The remaining submissions covered other educationally related areas including use of conference

publications, reading lists, mobile learning and annotation.

The submissions were judged using two different approaches. An evaluation framework was used to assess entries and the wider public were also given the opportunity to vote on entries in the People's Choice. The People's Choice was operated using Ideascale [20], a cloud-based crowdsourcing service combined with poster voting at the Open Knowledge (OKCon) festival in Geneva [21]. The online voting approaches adopted in the People's Choice are explained in a blog post on the Remote Worker blog [22]. After entries were reviewed by the evaluation committee, led by the LinkedUp advisory board, a shortlist of eight was agreed on 16 August 2013. The shortlist consisted of:

- **Knownodes** is a collaborative Web site that supports relating, defining and exploring connections between Web resources and ideas, making use of graph visualisations. Knownodes scored high on educational innovation.
- **Mismuseos** connects museum data with sources including Europeana, Dbpedia and Geonames. With Mismuseos, learners can browse and explore the backgrounds and relations between objects from multiple Spanish museums.
- **ReCredible** is a browsable topic map with Wikipedia-like content alongside. The topic library showcases interesting topics varying from dog breeds and alternative medicine to nanotechnology and information systems.
- **DataConf** is a mobile mashup that enriches conference publications. The reviewers applauded its attractive and effective design. DataConf is especially useful at the graduate education level.
- **We-Share** is a social annotation application for educational ICT tools. We-Share can help educators to find tools to support teaching at all educational levels, and received high scores on educational innovation.
- **YourHistory** is a Facebook app that makes history tangible by showing historic and global events that are related to your own life events and your interests.
- **Globe-Town** is a 'fun-to-use' tool that enables users to find out the most important trade partners, migrant populations and airline routes in their own countries. It also provides infographics on issues regarding society, environment and economy.
- **Polimedia** connects transcripts of the Dutch parliament with media coverage in newspapers and radio bulletins. Polimedia employs innovative information techniques and provides an attractive front-end that invites exploration and browsing.