Building the Open Elements of an Open Data Competition

Document Version:
Publisher's PDF, also known as Version of record

Document license:
CC BY-NC-SA

Open Universiteit
www.ou.nl

Building the Open Elements of an Open Data Competition

Mathieu d'Aquin
Knowledge Media Institute of The Open University, United Kingdom
m.daquin@open.ac.uk

Stefan Dietze
L3S Research Center, Germany
dietze@l3s.de

Hendrik Drachsler
ELSTEC institute of the Open University, The Netherlands
hendrik.drachsler@ou.nl

Marieke Guy
Open Knowledge, United Kingdom
marieke.guy@okfn.org

Eelco Herder
L3S Research Center, Germany
herder@L3S.de

Elisabetta Parodi
Lattanzio Learning, Italy
parodi@lattanziogroup.eu

Abstract

The European Union is increasingly committed to pushing forward open approaches as indicated by the G8 Open Data Charter, the Opening Up Education initiative, the launch of the Open Education Europa Portal for OER resources and other similar initiatives. The EU-funded LinkedUp Project (Linking Web data for education) aims to gather successful exemplars of the use of open web data in education, with the objective of pushing forward the exploitation of the increasing amounts of public, open data available online. It aspires to do this by facilitating developer competitions and deploying an evaluation framework, which identifies innovative uses of robust, web-scale information management applications. This article will look at how LinkedUp has moved beyond advocacy of linked open data and has begun encompassing open approaches in all areas of work. One key focus has been in bringing together the open elements of an open data competition and sharing them as widely and openly as possible. It is anticipated that these elements can then be progressed and built upon by others organising similar competitions in both academia and industry.

Introduction

In recent years, the European Union has made a substantial commitment to pushing forward open approaches. Events such as the agreement of the G8 Open Data Charter, the initiation of the Opening up education initiative and the launch of the Open education Europa Portal for OER resources, which all took place in 2013, clearly show that for the EU open is the way forward.

As a result of this move to openness, open data competitions and challenges are on the rise. They are an operational tool in demonstrating the power of open data and encouraging greater releases of data. Some offer insightful uses through combinations and visualisations of data, while others allow individuals to clean-up and upload data, possibly through crowd sourcing activities. These competitions are complemented by discrete events such as mashups, hack days and developer sessions. The last few years have also seen a rise of movement into business in events that offer seed funding for prototype development (see for example the Open Data Challenges series).

The [LinkedUp Project](#) has brought together the open elements of an open data competition and shared them as widely and openly as possible. It is hoped that these elements have been shared in such a way that others can progress them by building upon them. This paper looks at the steps involved and lessons learnt in building a truly open data competition.
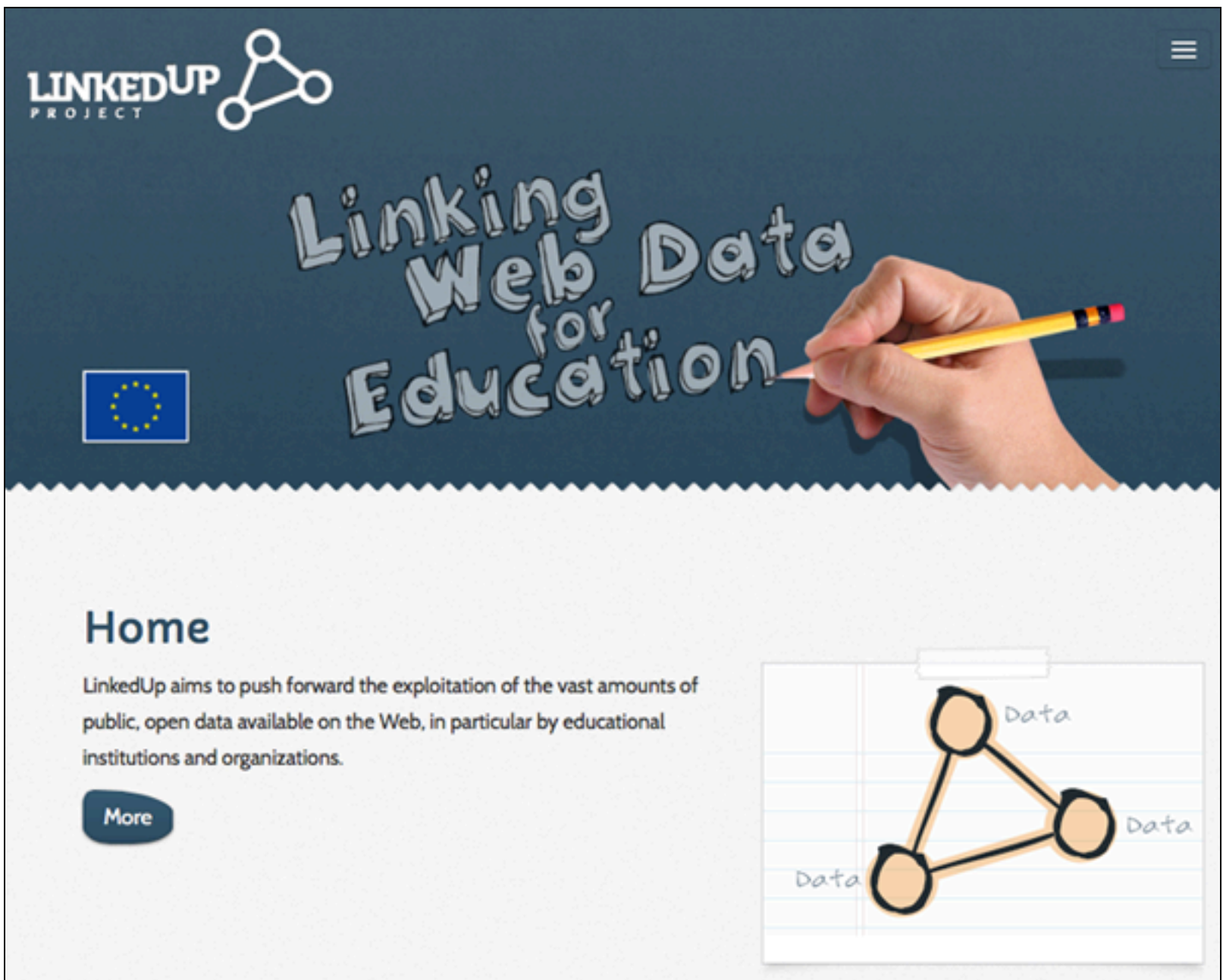
Figure 1: The LinkedUp Project Website

LinkedUp Project

The LinkedUp project (Linking Web data for education) is an EU FP7 Support Action running from November 2012 to November 2014. Its primary aim is to gather successful exemplars of the use of open web data in education, with the objective of pushing forward the exploitation of the increasing amounts of public, open data available online. It aspires to do this by facilitating developer competitions and deploying an evaluation framework, which identifies innovative uses of robust, web-scale information management applications. The project is comprised of six pan-European [consortium partners](#): the Open University (UK), Open Knowledge (UK), Elsevier (US), the Open Universiteit Nederland (Netherlands), Lattanzio Learning (Italy), and the project leader, the L3S Research Center of the Gottfried Wilhelm Leibniz Universität Hannover (Germany). The project also has a number of [associated partners](#) with an interest in the project including the Commonwealth of Learning (Canada) and the Data Archiving and Networked Services (DANS) based in the Netherlands.

The LinkedUp Challenge

The principal way in which the LinkedUp project intends to encourage engagement is through a series of open competitions aimed at eliciting web-data driven applications for personalised, open and online university-level studies. The [LinkedUp Challenge](#) is a series of three consecutive competitions looking for interesting and innovative tools and applications that analyse

and/or integrate open web data for educational purposes. The competitions are open to all: anyone from researchers and students, to developers and businesses. Each competition builds upon the previous, leading from innovative prototypes and tools 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 22nd May to 27th June 2013. By the closing date, 22 valid submissions had been received from 12 different countries (4 from the UK, 3 from France, 3 from Spain, 3 from the USA, 2 from the Netherlands and 1 from Greece, Bulgaria, Belgium, Italy, Argentina and Nepal). The abstracts are available from the LinkedUp Challenge website. 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.

During Veni, entrants to the competition had interpreted the specification "educational purposes" in a variety of innovative ways. A number of the entries had looked at MOOC 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 looked in particular at cultural heritage data and how museum data could be used in an educational context, the remaining submissions covered other educational related areas including use of conference publications, reading lists, mobile learning and annotation.

For the second competition (Vidi), which ran from 4th November 2013 till 14th February 2014, developers were once again asked to develop apps and prototypes for educational purposes but it was noted that while their tool may contain some bugs it needed to have a stable set of features and some proof that it can be deployed on a realistic scale. For Vidi there were also two focused tracks running alongside the open track. In these tracks developers were asked to design a solution to one of the following problems:

- **Simplificator** calls for applications easing access to complex information by summarizing them in a simpler form.
- **Pathfinder** requires applications easing access to recommendation and guidance when choosing appropriate curriculum of courses and related resources.

There were 14 submissions to the Vidi competition with authors from 12 countries. A shortlist of 9 submissions will be showcased at the European Semantic Web Conference (ESWC) in Crete, Greece in late May 2014 and awards allocated. The final competition, Vici, will officially launch at ESWC and will run from May to September 2014.

Open Data

The European Union has for years been stressing the goal of opening up data as a resource for innovative products and services and as a means of addressing societal challenges and fostering government transparency. In June 2013, the EU endorsed the G8 Open Data Charter and, with other G8 members, committed to implementing a number of open data activities in the G8 members' Collective Action Plan. These activities include making data available in an open format; enabling semantic interoperability; ensuring quality, documentation and where appropriate reconciliation across different data sources; implementing software solutions allowing easy management, publication or visualisation of datasets; and simplifying clearance of intellectual property rights.

The Open Definition states that: "Open data is data that can be freely used, reused and redistributed by anyone — subject only, at most, to the requirement to attribute and sharealike." The key features of open data are:

- **Availability and Access:** The data must be available as a whole and at no more than a reasonable

  reproduction cost, preferably by downloading over the internet. The data must also be available in a convenient and modifiable form.
- **Reuse and Redistribution:** The data must be provided under terms that permit reuse and redistribution including the intermixing with other datasets. The data must be machine-readable.
- **Universal Participation:** Everyone must be able to use, reuse and redistribute — there should be no

discrimination against fields of endeavour or against persons or groups. For example, 'non-commercial' restrictions that would prevent 'commercial' use, or restrictions of use for certain purposes (e.g. only in education), are not allowed.

The LinkedUp project focuses on open web data and has its roots in the linked data movement. In the education sector the benefits of using open and linked web data are now really starting to show with several universities engaged in the deployment of linked data approaches (see Linked Universities). In the UK this has been driven by a requirement for transparency and accountability by 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 data sets and offer new perspectives.

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 which has the potential to introduce a paradigm shift in the way educational services are provided, and substantially improve educational processes and lower the costs of offering higher education. However, 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, or do not provide users with an appropriate level of context and filtering for the vast amounts of heterogeneous information and content retrieved to be adequately exploitable. 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 organizations, and to facilitate the development of innovative applications produced by the LinkedUp community and challenge participants and their deployment in real-world use case scenarios.

To support the LinkedUp Challenge the LinkedUp team is curating a dataset catalogue of educationally relevant datasets from the Linked Open Data cloud. The LinkedUp Dataset Catalog can be used in many ways. It is first and foremost a registry of datasets. Datahub.io, a data management platform from Open Knowledge based on the CKAN data management system, is probably the most popular global catalog of datasets, and forms the heart of the Linked Open Data cloud. In the interest of integrating with other ongoing open data efforts rather than developing in isolation, the LinkedUp Dataset Catalog utilises dataset information from Datahub.io. Any dataset in Datahub.io can be included in the Linked Education Cloud group (providing it is relevant), and the datasets in this group are also globally visible on the Datahub.io portal. Every dataset is described with a set of basic metadata and assigned resources. This makes it possible to search for datasets; for example, one could search for the word "university" in the Linked Education Cloud, and obtain datasets that explicitly mention "university" in their metadata. These results can be further reduced with filters, for example to include only the ones that provide an example resource in the RDF/XML format.

Complementing the contribution to the global Datahub.io registry of data sources, the LinkedUp Dataset Catalogue utilises the information from the Linked Education Cloud group to provide richer descriptions and search facilities. Indeed, the LinkedUp team has developed a set of techniques which automatically extract information about the content of the data, their structure and topics. This enables application developers to find and discover data of use for their application based, for example, on the type of resources they are interested in (documents, people, organisations, etc.), or on the discipline they wish to cover (biology, computer science, etc.) In other words, through this growing and evolving catalogue, the LinkedUp team's goal is to provide an easily accessible and usable resource to reduce the barrier of entry for application developers to find and actually exploit the wealth of open educational data available on the web.

LinkedUp Data Repository   Home   Browse Data   On Datahub   SPARQL Endpoint   LinkedUp DevTalk

Endpoints
Endpoints in bold are the ones currently responding. Endpoints in italic are down.

data.southampton.ac.uk with SPARQL   *Ege University Linked Open Data*
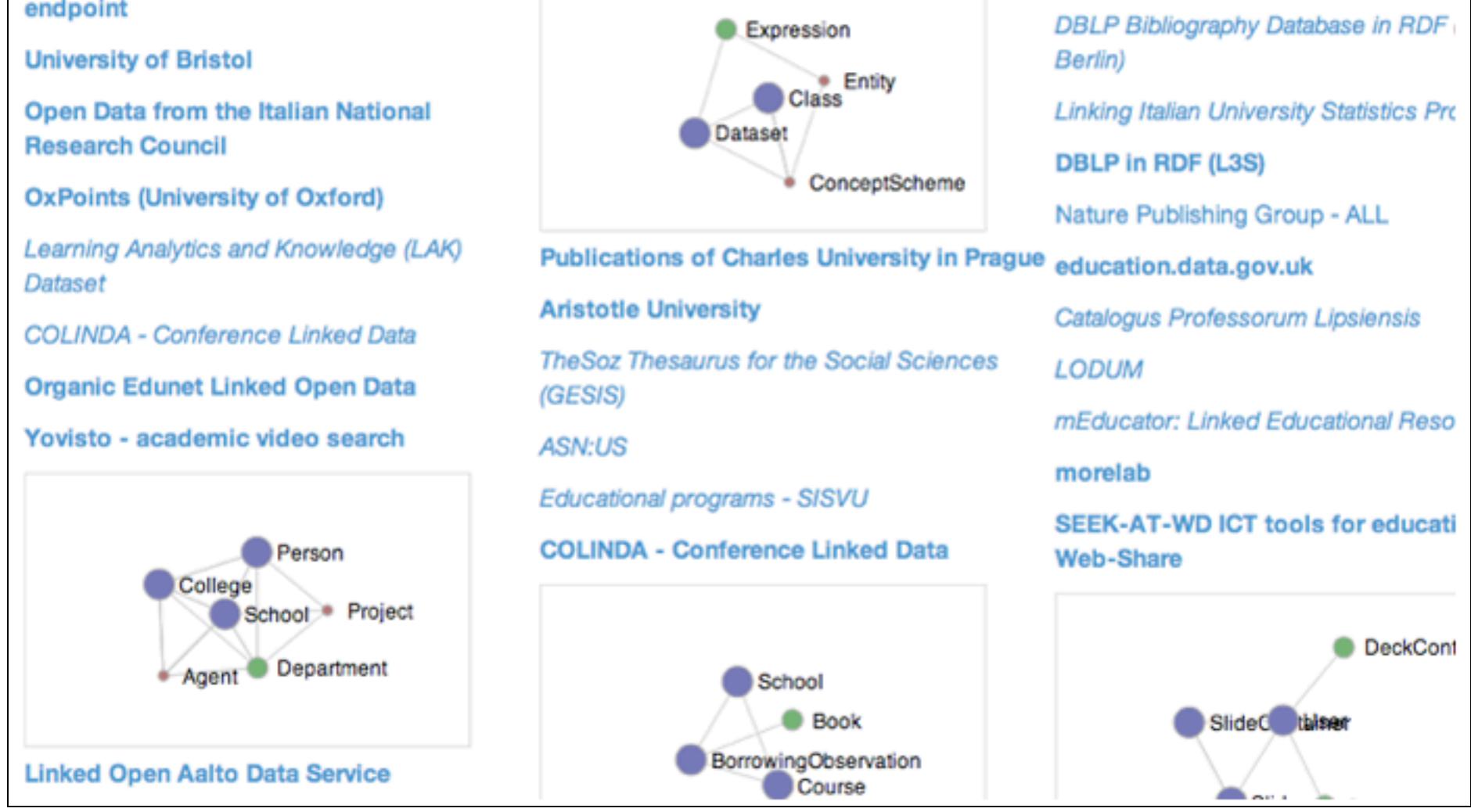
Figure 2: LinkedUp Catalogue

Open Evaluation

The [Evaluation Framework for Open Web Data Applications](#) (a deliverable in month 6) is a complete framework for assessing large-scale open Web data applications, taking into account educational aspects. It consists of predefined evaluation procedures and benchmarking criteria for the ranking of the participating projects during the LinkedUp competitions. The requirements include interdisciplinary coverage, integration of high-quality web data, integration with local data, context and filtering, scalability and performance and multilingualism. The framework consists of a transparent and distinct list of evaluation criteria that enable the review panel to measure, based on quantifiable criteria and qualitative assessment, the impact and appropriateness large-scale web information and data applications. It also helps identify strengths and weaknesses of particular projects and submissions. These ratings are being used by LinkedUp to analyse 'gaps in knowledge' that will be covered as part of the open training approach. The evaluation framework is being reviewed after each stage of the challenge on its validity, and possible improvements needed, to reach the most sustainable and practical evaluation instrument by the end of the project.

As part of the sustainability of the LinkedUp Project a 'LinkedUp Toolbox' will also be released. The toolbox is likely to contain various 'tools' appropriate for two audiences: open data competition organisers and open data competition entrants. The tools will include the evaluation framework in various formats and versions, for example there will be lists of questions, evaluation processes and evaluation categories, best practices, case studies and technology transfer modules.

The LinkedUp team will also be working with the [Apps4Europe Project](#), a support network that provides tools to transform ideas for data based apps into viable businesses. The Apps4Europe Project have created supporting materials to help people get started with releasing data and organising competitions. Documents on the following areas are likely to form part of the toolbox:

- Guidelines for reusers of data
- Guidelines for data owners

- Guidelines for organisers of app challenges
- Guidelines for data owners and data publishers

It is anticipated that the toolbox will also cover legal and privacy aspects which are of importance when exposing and using publicly available Web data.

Open Voting

In order to involve the public in deciding who will win each of the LinkedUp Competitions, an open vote runs in parallel to assessment using the evaluation framework. The open voting system has been named the 'People's Choice'. Early on a specification was written outlining requirements, which included the need for voters to register and be restricted to one vote each, and for each submission to have its own People's Choice URL. The URL was required because one of the main objectives for the People's Choice idea was to create a 'buzz' around the competition. The developers would be able to encourage friends and colleagues to vote for their submission, which in turn would be good publicity and create interest in the project. A number of open voting systems were analysed and Ideascale, a cloud-based crowdsourcing service, was chosen for the Veni voting. Further details of the open voting approach are given on the remote worker blog (Online Voting: the highs and lows). The scores from this system were combined with poster voting at the Open Knowledge (OKCon) festival in Geneva to decide a People's Choice winner. For the second competition, Vici, the GNOSS community platform, which allows votes and comments, is being investigated as a possible option.

Open Ideas

To support the LinkedUp Challenge, suitable use cases are being collected by the LinkedUp consortium and associated organisations, including representatives of renowned industrial, academic and higher education institutions such as Elsevier, the BBC, and the Commonwealth of Learning. The use cases present actual, real-life challenges that the related institutions are facing and addressing. Their aim is to provide challenge participants with inspiration and examples of the kinds of problems that submissions can address.

One example is the Educationalizer use-case supplied by Elsevier. The use case notes that most content and data is not created, designed or formatted specifically as learning objects for educational purposes. But much of it has value for educational purposes if given the appropriate context, threaded together with a larger curriculum, and described in a manner that is meaningful to the user. Achieving this can be particularly valuable to an educator who is approaching a new field for the first time and needs to find material for a curriculum. This can also be valuable for an educator who is approaching a field that is inherently interdisciplinary and candidate data and content can come from disparate resources.

Open Training

One of the important goals of the project is to ensure 'technology transfer in the education sector', i.e., demonstrating and promoting the benefit of open Web data technologies in education, and providing a reusable testbed in this domain. Developing applications based on open web data requires specific skills. The project, therefore, employs an open training approach to support developers in acquiring these skills, including code clinics,

offering one-to-one support, and sharing code recipes and 'how to' guides.

The LinkedUp DevTalk blog regularly shares insights into how to build upon open data and linked data included in the catalogue. The resources page points to a set of other useful resources, such as tutorials, that can help developers enter the competitions. A dedicated support contact form makes it possible to interact with the technical expert team of the project with specific questions regarding the implementation of open data-based applications.

The project team have also worked with EUCLID, an EU Project which is building a Linked Data curriculum. The curriculum will be realized as living learning materials on a community website, and will be evaluated, refined, and extended in a webinar series, face-to-face training, and through continuous community feedback and contributions coordinated by a designated community manager. A significant share of the materials will consist of examples

referring to real-world data sets and application scenarios, code snippets and demos that developers can run on their machines, as well as best practices and how-tos.

In order to support the Vidi Competition the LinkedUp team ran a webinar on Adobe Connect which introduced the competition and gave developers the opportunity to ask technical and administrative questions. After the initial presentation the LinkedUp technical team answered questions from the attendees and discussed some possible applications. There were 22 webinar attendees and the recorded webinar is available online along with the transcript of the discussion and slides.

It is worth noting that the support provided is not just technical but also more general regarding legal, exploitation and dissemination issues related to participants applications. These processes will ensure that, by the end of the project, the results of the challenge can be taken up by educational organisations, commercial organisations or the development community with a clear legal and exploitation framework for each application.

Open Education Handbook

Open data in education, or open education data is still a relatively new area of interest. It is slowly starting to feature in discussions around open education which have tended to previously focus on content and primarily Open Educational Resources (OER), however there is much more dialogue to take place. One important deliverable of the LinkedUp Project is the LinkedUp Handbook on Open Data in Education, a collaboratively written living web document targeting educational practitioners and the education community at large.

The original intention was only to cover open data use in education, but it was felt that a broader scope would be a more useful output for the primary audience. It would enable readers to have a better understanding of how different aspects or facets of open education, such as resources, data and culture, fit together. It would also allow exploration of how open education can benefit from open and linked data approaches. The handbook is now referred to as the Open Education Handbook.

The Open Education Handbook has so far been written through a series of mini-booksprints. The initial booksprint was held in London on Tuesday 3rd September 2013 and was attended by over 20 open education experts from many different sectors (commercial, academic, government, not-for profit). The initial draft of the handbook was made and good headway was made in a number of sections including data, learning and teaching practice, and OERs. In October 2013 the handbook was switched from Google docs to Booktype, an open source platform for writing and publishing print and digital books. A wide call was made for contributions to the handbook. A second Open Education Handbook booksprint was held in Berlin in late November 2013. Much of the discussion at this event revolved round the structure of the handbook and it was 'chunked up' into question areas. In early 2014, as part of Education Freedom Day, the Open Education Handbook was translated and adapted to Portuguese and released on Booktype and in EPUB format. Work has also begun to create a set of openly licensed slides based on the handbook in Slidewiki. Plans are being made for more focused handbook events, similar to the Open Education Timeline event, where a physical, and later digital, timeline of open education was created. An open education mapping activity to fit in the handbook will take place at the Helsinki Learning Festival.
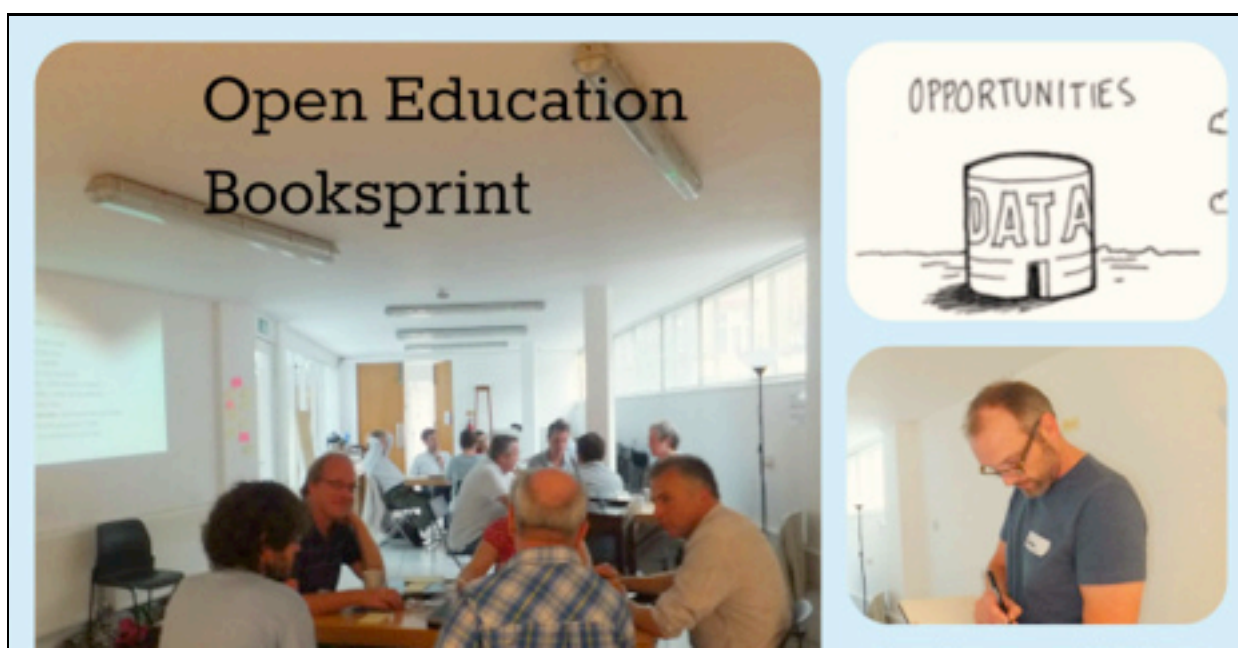
Figure 3: Open Education Handbook Booksprint

Collaboration with other Open Competitions

The LinkedUp Project is aware that it is not working alone in the open data challenge space and has been collaborating with others setting up similar or complementary challenges.

The EU-funded Open Education Challenge is asking for innovators to submit project ideas in the area of open education. A shortlist will then receive mentoring and seed funding through the European Incubator for Innovation in Education, and get direct access to investors. There is now a formal collaboration between the Open Education Challenge and LinkedUp, and the OEC has been listed as an associate partner. Members of the LinkedUp team will sit on the OEC board and may offer support in mentoring approaches. It is also hope that entries from the Vidi competition will follow on to compete in the OEC.

The UK Open Data Institute Education Open Data Challenge is one in a series of seven challenges. The aim of the challenges is to generate innovative and sustainable solutions to social challenges using open data. In the education challenge they are looking for teams to create products and solutions using open data to help parents make informed choices about their children's education. LinkedUp have been talking to the Open Data Challenge team about key datasets for the UK that are currently available in the LinkedUp catalog. The two challenges hope to work together on dataset preparation and identification, which will reduce the amount of work needed to be carried out independently. Another area for collaboration is marketing; cross advertising ensures participant and community growth and will potentially lead to an increase in the pool of competitors for competitions.

Open Community

One core activity of the LinkedUp project is to establish a network of open Web data and resource evangelists (in particular in the area of education) who will raise awareness of legal and technical best practises in a variety of different domains, facilitate conversation and collaboration between technologists in the Open Educational resource community and engage end users in teaching and learning. This is happening through establishment of the Open Education Working Group. Open Knowledge defines the working groups it hosts as collaborations of individuals, who meet virtually and in-person, to focus on a particular area of open knowledge and its effect on society. Another important aspect of working groups is the opportunity for cross organisation collaboration by engagement with pre-existing groups. The Open Education Working Group has a broad remit but is working closely with other groups already active in this area, such as the OER community.

The group itself has been built around open and transparent working practices. The structure follows that described in the membership charter, the wording of which is still being discussed by the group:

Broadly speaking, the Open Education Working Group is made up of three groups:

- A members group that collectively decides on the direction of the Working Group and its priorities. Working Group members are the key evangelists of the initiative and they focus on concrete actions that help to

generate more openness within the domain.

- A larger group that gathers around the public discussion list to share information on events and news from and around the field. Open Knowledge working group mailing lists are open for anyone to join.

- An Advisory Board which contains high-profile Open Education advocates who are experts in the field. The Advisory Board provides thought leadership about the direction of the working group and helps to raise the profile of the working group by talking about the group and their work at conferences and events. Members of the working group have been asked to nominate people to sit on this board. The group is keen to have a diverse Advisory Board and so will be considering factors such as location and sector when deciding on participation.

While the working group is co-ordinated by LinkedUp staff there are many opportunities for members to get involved. In January 2014 the first working group call took place using Google hangout and was attended by over 15 people, with many others watching the video and reading the minutes afterwards. It is anticipated that calls will take place once a month. The purpose of these calls will be two-fold. Firstly, they will be an opportunity for members of the open education community to get to know each other better. During the first half of the call people will be able to introduce themselves and give a brief overview of the open education related project and activities they are involved with. Secondly, they will be a chance to agree on the structure and focus of the group, and on activities for the group to work on.

Those interested can also participate in the group by sharing activities, participating in discussions on the mailing list and writing blog posts; there is a series of posts on open education around the world being published. Areas of interest for the Open Education Working Group include moving forward the debate on the opening up of MOOC data, building up evidence and case studies around open data use in education, looking at multilingual issues, and building lists of resources and best practice.

Conclusions

One of the core elements of the LinkedUp Project is openness. The LinkedUp Challenge not only builds on open and linked data, but all the elements of the challenge are also being openly released and shared. Apart from the competition framework, which can be used as a starting point, there is also the legacy of promotional material (website, social media, posters, flyers) and competition outcomes (presentations, demos, publications) that can be used as input and inspiration.

It is anticipated that the experience and lessons learnt during the running of the LinkedUp Challenge can offer real benefit to others running open data competitions. For example, preliminary project research noted the importance of having a balance between academic and industry-oriented approaches. Academic competitions are typically organised at conferences and workshops, and involve the submission of papers, while more industry-oriented competitions usually focus on the competitive aspect, promoting prototypes aiming to win prizes, and paper

submission is typically not necessary. During the LinkedUp Challenge there have been efforts to keep this balance in the design and wording of the promotion material, as well as in the requests to participants and the selection of venues.

Another lesson learnt relates to the importance and value of planning and continuous monitoring and update, for example timelines need to be arranged in advance. Thanks to the feedback and monitoring of participants it was possible to "push" promotion and marketing during the competitions. The judges' comments also allowed updating of the Evaluation Framework across competitions.

Through the open approaches taken by the LinkedUp Project a 'template' is being created for others to reuse and build upon when creating their own Open Data Competition.The main LinkedUp "legacy" will be an open data competition toolbox available from the main project website by the end of 2014.

About the Authors

Mathieu d'Aquin (Data and Support Coordinator) is a research fellow at the Knowledge Media

Institute of [The Open University](#), and his research activities focus on the Semantic Web, and especially on methods and tools to build intelligent applications exploiting online knowledge. He has been involved in the organisation of events such as the IWOD series of workshops and the SSSW summer school.

Stefan Dietze (Project Coordinator) is a Research Group Leader at the [L3S Research Center](#) (Germany). His research interests are in Semantic Web and Linked Data technologies and their application to Web data integration problems. Stefan currently is coordinator of two European R&D projects (LinkedUp, DURAARK) and he has been involved in the organisation of numerous events, such as ACM Web Science 2012 or the Linked Learning workshop series.

Hendrik Drachsler (Evaluation Coordinator) is Assistant Professor at the CELSTEC institute of the [Open University](#) in the Netherlands. He research area is the personalisation of learning with information retrieval technologies, especially recommender systems. He is interested in research on educational datasets, linked data, data mashups, data visualisations and learning analytics. Learn more [here](#).

Marieke Guy (Community Coordinator) works for [Open Knowledge](#), a non-profit organisation dedicated to promoting open data and open content. Marieke has an MSc in information management and prior to her current employment spent 13 years as a research officer at the University of Bath. Her main areas of interest include research data management, digitisation of cultural heritage works, web archiving and digital preservation.

Eelco Herder (Challenge Coordinator) is a Senior Researcher at the [L3S Research Center](#). His research areas include Web personalization, user modeling, usability and HCI in general. He organized several workshops at UMAP, ESWC and IUI. He is program chair for Hypertext 2014 and was a member of the organization committees for UMAP 2013, 2012 and 2011, CHI 2012 and Adaptive Hypermedia 2008. Learn more [here](#).

Elisabetta Parodi works for [Lattanzio Learning Spa](#) which offers consulting for management, training, operational and managerial outsourcing, assistance with internationalization and communication in Italy and abroad. She supports several projects: LinkedUp, weSPOT and INTUITEL.