Putting Linked Data to Use in a Large Higher-Education Organisation

Mathieu d'Aquin

Knowledge Media Institute, The Open University, Milton Keynes, UK
{m.daquin}@open.ac.uk

Abstract. In this paper, we describe applications built on top of the Open University's linked data platform (data.open.ac.uk), from the point of view of the way they implement particular forms of interactions with linked data. We especially focus on the common advantages and pitfalls in interacting with linked data that these applications illustrate, from both the end-users' and the developer's perspectives. We conclude on suggested steps forwards regarding the ways to facilitate the realisation and adoption of applications interacting with linked data.

1 Introduction

Building and deploying linked data in a large organisation represents a challenge at many different levels. Many of the past research and development works have focused on the publication process for linked data: how to obtain data from legacy information systems; how to model these data according the linked data principles; how to link the organisation's data with external sources; how to expose the data online for wide accessibility. However, it is becoming more and more clear that there is a need to investigate another, possibly more important aspect – the other side of the coin: what are the issues related to interacting with linked data, from an end-user as well as a developer perspective.

In this paper, we rely on our experience in building, deploying and applying the Open University's linked data platform (http://data.open.ac.uk) to investigate this perspective. Data.open.ac.uk, developed through the LUCERO project (http://lucero-propject.info), was the first initiative to expose the public information of a university as linked open data, collecting and providing access to data from across the institution's departments[1]. One of the challenges

In: C. Unger, P. Cimiano, V. Lopez, E. Motta, P. Buitelaar, R. Cyganiak (eds.): Proceedings of Interacting with Linked Data (ILD 2012), Workshop co-located with the 9th Extended Semantic Web Conference, Heraklion, Greece, 28-05-2012, published at http://ceur-ws.org

[1] It has since then been followed by a number of other initiatives applying linked data in the higher-education sector, with high potential impact regarding the reuse

related to pioneering the use of linked data in a particular sector is the need to demonstrate the advantages that it brings to the users of the organisation. Several applications have been developed that are at different stages of their lifecycle, and have been deployed for different audiences (in size, technological awareness, etc.) Through these applications, common benefits as well as the challenges of providing linked data-based functionalities to 'real users' are emerging.

In this paper, we discuss some of these applications from the perspective of the way they provide means to interact with existing linked data sources. We discuss the lessons learnt from our experience regarding the issues, challenges and pitfalls of interacting with linked data. We not only consider here the perspective of the end-users in using the developed applications, but also the ones of the developers having a more direct interaction with linked data with the purpose of providing usable functionalities.

2 The Open University's Linked Data Platform

Data.open.ac.uk is a linked data endpoint that collects data from many different sources within the organisation, using a variety of different vocabularies and linking to external sources such as dbpedia.org or geonames.org (see [4]). Data collection is based on identifying streams of data inside the organisation and, in collaboration with the data owners, re-modelling the data to fit exposure as linked data. The architecture of the platform is based on a triple store providing a SPARQL endpoint, on ad-hoc mechanisms to extract and update data from the considered streams, and on a basic URI delivery mechanism. The process is continuous, with more data being exposed whenever new resources are made available. The current sets of data include:

Course information: This includes information about courses that are currently on offer at the Open University. The information includes a short description of the course, information about the levels and number of credits associated with it, the topics, and the conditions of enrolment (the countries in which it is available, the dates for registration and the student fees).
Example: `http://data.open.ac.uk/course/m366`

Research publications: This includes metadata for the research articles and other publications authored by Open University researchers, as available on the publication repository of the university (ORO, see `http://oro.open.ac.uk`). An article typically includes information about the authors, dates, abstract and venue of the publication.
Example: `http://data.open.ac.uk/oro/29916`

Podcasts: This includes the metadata for audio and video podcasts produced and made openly available by the Open University as open educational resources (see `http://podcast.open.ac.uk`). A typical podcast entity includes a short description, the topics, a link to a representative image and to

and interoperability of educational resources from across universities (see `http://linkeduniversities.org`) a transcript if available, as well as information about the course the podcast might relate to and license information regarding the content of the podcast. **Example:** `http://data.open.ac.uk/podcast/218dce44a4ed17b36ada50d18b866b03`

Open Educational Resources: This includes metadata about units of Open Educational Resources made available by the Open University through its OpenLearn system (see `http://openlearn.open.ac.uk`). A typical 'OpenLearnUnit' includes a short description of the units, the topics, tags used to annotate the resource, its language, as well as the course it might relate to, and the license that applies to the content.
Example: `http://data.open.ac.uk/openlearn/m366\_2`

Youtube videos: This includes metadata about videos published by the Open University on Youtube, as promotional videos or open educational resources. Such metadata include a short description of the video, the tags that were used to annotate the video, the collection it might be part of and a link to the related course if relevant.
Example: `http://data.open.ac.uk/page/youtube/A7BA7C1155BE887E/1E5D9A1BA21BDC51`

University buildings: This includes information about the building owned by the University. The Open University being a distance learning education, besides the main campus located in Milton Keynes, it also includes regional centres located in different locations across the UK territory. Building descriptions include their address (including links to the corresponding administrative areas in `http://data.ordnancesurvey.co.uk/`), a picture of the building and the sub-divisions of the building into floors and spaces.
Example: `http://data.open.ac.uk/location/building/rbedrb`

Library catalogue: This includes metadata about items available at the Open University's library that relate to Open University courses (textbooks and setbooks). The description of each item includes information about the topics, the authors, the publisher and ISBN, as well as the course it relates to.
Example: `http://data.open.ac.uk/library/406973`

Other specific data: Other datasets are also included that concern specific research projects (e.g., the Open Arts Archive – `http://openartsarchive.org/`) or specific departments of the Open University (e.g. the FOAF profiles of people from the Knowledge Media Institute – `http://people.kmi.open.ac.uk`).

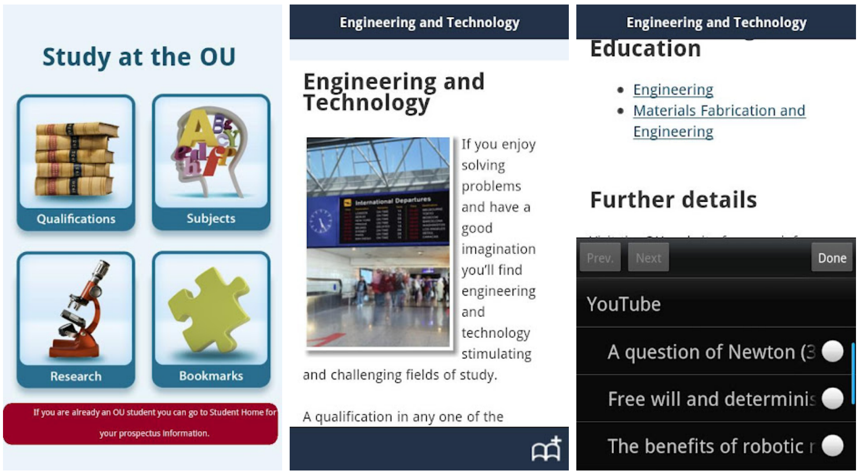
3 Applying Linked Data at the Open University

Amongst the many applications developed on top of the data.open.ac.uk platform (see e.g. [3]), we choose to describe here the ones that had a concrete deployment and impact on different categories of users of the Open University, focusing on our experience regarding benefits and issues at the level of interacting with the data, both from an end-user perspective, and from a developer's perspective.

3.1 The 'Study at the OU' Mobile Application

"Study at the OU" the website of the Open University that contains the description of the courses and qualifications that can be obtained from the University (see `http://www3.open.ac.uk/study/`). A mobile application was recently developed by the communication services of the University so that this course catalogue and additional information about the topics covered can be accessed from various types tablets and smartphones (see Figure 1). 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 `data.open.ac.uk`, simply querying resources that are directly related to the topic being considered, or for resources attached to courses that are related to this topic.

Since its launch in January 2012, this part of the application has been accessed more than 25,000 times.



If you are already an OU student you can go to Student Home for your prospectus information.

If you enjoy solving problems and have a good imagination you'll find engineering and technology stimulating and challenging fields of study.

A qualification in any one of the

Fig. 1. Screenshots of the 'Study at the OU' application on an Android phone.

From the users' perspective, while connecting to relevant resources could be a very valuable feature, especially for prospective students, there is no indication that this has been realised with linked data. In other terms, without knowing the underlying information architecture, there is no reason not to believe that this functionality was realised using more common technologies. The added value here is however that such a simple and straightforward feature requires the combination of information coming from different, and mostly isolated systems (the course database, the podcast system, etc.), which are seamlessly integrated through linked data.

It is from **the developers' perspective** however, that the benefits of interacting with linked data to build such an application is appearing obvious. Indeed, providing the same feature using the usual information infrastructure of the Open University would have required accessing and connecting to many different systems that use different platforms, technologies, formats, conventions, etc. It would also have required an ad-hoc integration of the data, with an additional level of complexity.

It is still a problem however for developers used to more common technologies to make use of linked data technologies such as SPARQL (`http://www.w3.org/TR/rdf-sparql-query/`), and to integrate them with their usual development environment. The solution adopted here was to create an intermediary view generating an ad-hoc XML descriptions of the relevant information from the results of pre-established SPARQL queries. This also allowed the inclusion of caching mechanisms, to avoid adding unnecessary overhead to the SPARQL query engine to process identical queries from potentially thousands of users.

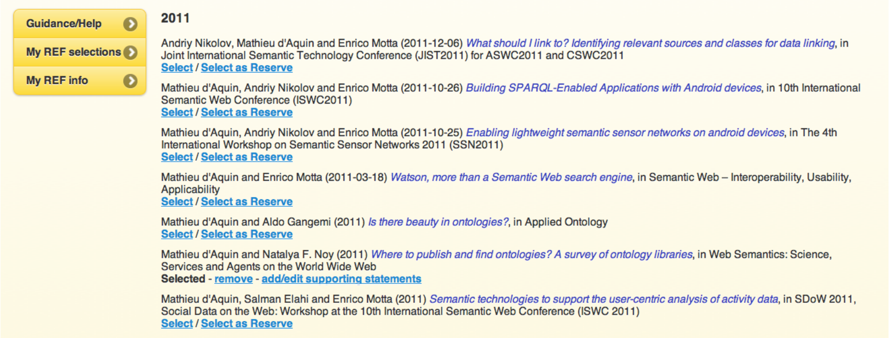
3.2 Supporting the 'Research Excellence Framework' Activities

The 'Research Excellence Framework' (REF) is the process applied to evaluate and assess the quality of research in UK universities (see `http://www.hefce.ac.uk/research/ref/`). As part of this process, each university is required to submit to their corresponding funding body a report summarising the research carried out at the university in various disciplines. In order to achieve this, 18 different panels have been formed at the Open University (to cover the disciplines in which the Open University is carrying-out research), in charge of identifying individual researchers with a selection of their publication to be part of the submission.

To support this work, an application was developed (see Figure 2) to be used by individual researchers to select what they considered to be their 'best' publications (since 2008), to indicate to which discipline they are associated, and to annotate their selected publications to include supporting statements for their selection (describing their significance, originality, etc.) This application has already been accessed by about 600 researchers concerned with the REF at the Open University, and a similar application is being developed at the moment to support the work of the 18 panels.

The application uses linked data to obtain for each individual researcher, the list and description of their publications in the recent years, as well as to connect them with information regarding their role in the organisation and the faculty/department they relate to. The captured information (selection of publications and their annotations) is also processed according to linked data principles and technologies, creating another (private) triple store

From the users' perspective, the main advantage of using linked data in this application is that information about their publications is directly obtained and integrated with other information, without them having to provide any additional input. Their linked data identifier (i.e., their URI) is directly derived



2011

Andriy Nikolov, Mathieu d'Aquin and Enrico Motta (2011-12-06) *What should I link to? Identifying relevant sources and classes for data linking*, in Joint International Semantic Technology Conference (JIST2011) for ASWC2011 and CSWC2011
Select / Select as Reserve

Mathieu d'Aquin, Andriy Nikolov and Enrico Motta (2011-10-26) *Building SPARQL-Enabled Applications with Android devices*, in 10th International Semantic Web Conference (ISWC2011)
Select / Select as Reserve

Mathieu d'Aquin, Andriy Nikolov and Enrico Motta (2011-10-25) *Enabling lightweight semantic sensor networks on android devices*, in The 4th International Workshop on Semantic Sensor Networks 2011 (SSN2011)
Select / Select as Reserve

Mathieu d'Aquin and Enrico Motta (2011-03-18) *Watson, more than a Semantic Web search engine*, in Semantic Web – Interoperability, Usability, Applicability
Select / Select as Reserve

Mathieu d'Aquin and Aldo Gangemi (2011) *Is there beauty in ontologies?*, in Applied Ontology
Select / Select as Reserve

Mathieu d'Aquin and Natalya F. Noy (2011) *Where to publish and find ontologies? A survey of ontology libraries*, in Web Semantics: Science, Services and Agents on the World Wide Web
Selected - remove - add/edit supporting statements

Mathieu d'Aquin, Salman Elahi and Enrico Motta (2011) *Semantic technologies to support the user-centric analysis of activity data*, in SDoW 2011, Social Data on the Web: Workshop at the 10th International Semantic Web Conference (ISWC 2011)
Select / Select as Reserve

Fig. 2. Partial screenshot of the interface for researchers to annotate their publications for the REF.

from the login name they use to access the application (as well as all other systems on the Open Universities intranet), meaning that the relevant publications are displayed directly as they access the application. The use of linked data is in principle, as in the previous section, hidden from the user. However, some elements can sometimes create confusion due to the different modelling of the data from the original sources. In the original source for example, while each author is associated with a unique ID, their name is recorded separately for each publication. In the linked data version, the names are all aggregated under the person's ID, meaning that each publication might appear slightly differently than in the original source (ORO). This also means that errors that can be very localised in the original source (a typo in the name of an author in one of their publications), might have a larger impact in the linked data version. Similarly, the application relies on the fact that each person and each publication is associated with a stable and unique identifier. This is however a strong assumption which, even considering well curated sources, is often hard to achieve (e.g., people changing user IDs, publications entered multiple times, etc.)

Here again, it is **from the developers' perspective** that the advantages of relying on linked data technologies and principles are the most obvious. Indeed, this allows to build on top of existing data and process various sources in an homogeneous way. Also, in producing new data (selections and annotations of publications), linked data technologies allow more flexibility and agility than with traditional, relational database systems: 'adding a field' in the data is trivial and does not require any database administration task, just adding data according to the newly considered ontological property. This is even more facilitated with the appearance of robust implementations of the SPARQL Update (`http://www.w3.org/TR/sparql11-update/`) language (on which this application relies), allowing the homogeneous use of the HTTP protocol both for querying and updating a triple store. Adding and integrating new sources of data is also made easier, as long as this information is provided using URIs consistent with the ones already in use, and the produced data is naturally reusable to build further applications.

Many issues appear here however that are not usually present when using more traditional technologies. One of them concerns the way to deal with incomplete data. Indeed, in applications like the one considered here that rely on a set of established queries, assumptions are made regarding certain properties of the data, which are never made explicit. For example, it is expected that every publication is associated with the list of authors. In case this assumption is not valid, and some publications do not have a list of authors, these publications would simply not appear in any result, making it difficult to recognise that a potentially problematic issue have emerged. This adds to the difficulty of knowing whether a problem that is identified at the level of the application originates from the application, the linked data representation or the original sources. Dealing with such issues generates more complexity in the querying process, as well as in the development/maintenance of the application.

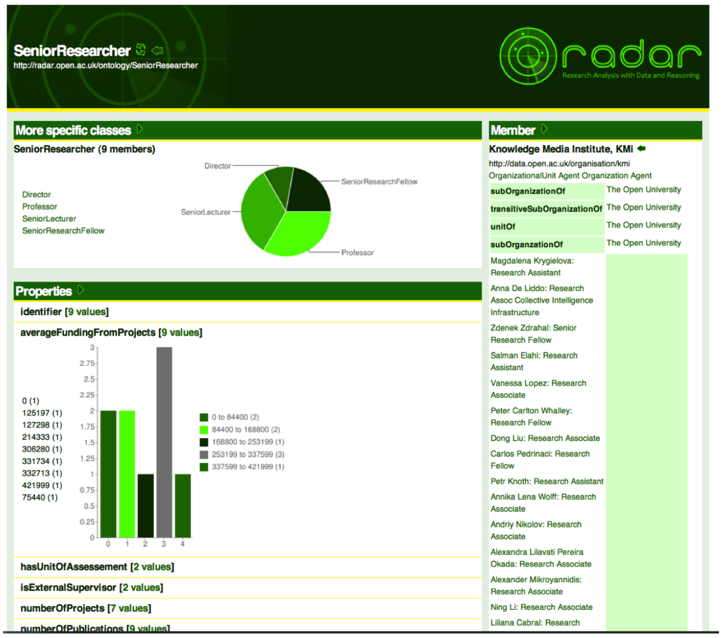
3.3 Understanding Research Communities at the Open University

In the continuity of the application described in the previous section, another application was developed to help research managers within the university in understanding and monitoring their research communities. Called RADAR (Research Analysis with DAta and Reasoning), this application makes use of information about research publications from ORO (as above), as well as other sources regarding the positions of researchers, their projects, funding, supervision history, etc. to visualise different indicators of research activities for individuals and groups.

Two parts of the application were developed. The first one relies on a generic framework for the visualisation and exploration of linked data sets, which is parametrised by an ontology of the particular domain of the application. It uses basic ontological reasoning to classify individuals into different classes, and uses automatically generated charts and tag clouds to visualise the distribution of values of the properties in individual classes (see Figure 3). In our case, the classes correspond for example to different categories of academic staff (senior/junior researchers, lecturers, professors, etc.) or different types of projects (internal, national, european, etc.) The indicators being visualised here correspond for example to the amount of funding received, the number of publications, of projects or the number of supervised students.

The second part of the application uses the same data and indicators, but displays graphs specifically conceived for the application in research community analysis, considering for example the distribution of the number of publications per year, or the overlap between the publications of a group of co-authors (see Figure 4).

From the users' perspective, the advantages and issues related to the generic part of the RADAR applications appear very clearly. On the one hand, the application is driven by an ontology, which means that only integrating more



SeniorResearcher
http://radar.open.ac.uk/ontology/SeniorResearcher

radar
Research Analysis with Data and Reasoning

More specific classes

SeniorResearcher (9 members)

Director
Professor
SeniorLecturer
SeniorResearchFellow

Properties

identifier [9 values]

averageFundingFromProjects [9 values]

0 (1)
125197 (1)
127298 (1)
214333 (1)
306280 (1)
331734 (1)
332713 (1)
421999 (1)
75440 (1)

hasUnitOfAssessement [2 values]

isExternalSupervisor [2 values]

numberOfProjects [7 values]

numberOfPublications [9 values]

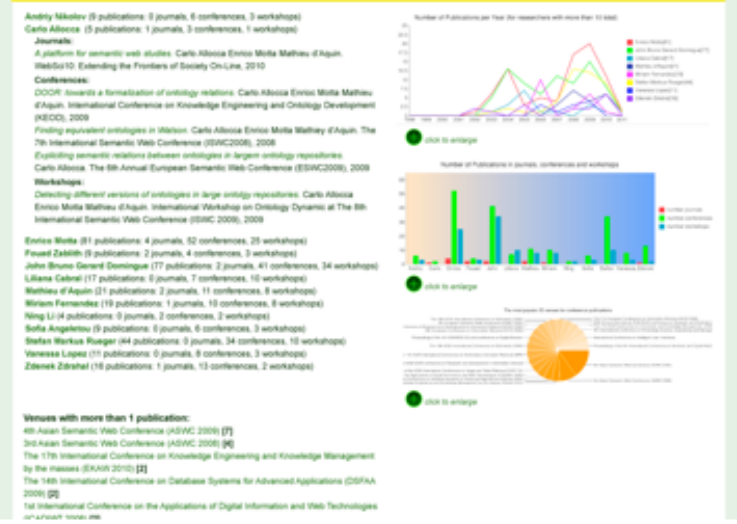
Member

Knowledge Media Institute, KMi
http://data.open.ac.uk/organisation/kmi
OrganizationalUnit Agent Organization Agent

subOrganizationOf	The Open University
transitiveSubOrganizationOf	The Open University
unitOf	The Open University
subOrganzationOf	The Open University

Magdalena Krygielova: Research Assistant
Anna De Liddo: Research Assoc Collective Intelligence Infrastructure
Zdenek Zdrahal: Senior Research Fellow
Salman Elahi: Research Assistant
Vanessa Lopez: Research Associate
Peter Carlton Whalley: Research Fellow
Dong Liu: Research Associate
Carlos Pedrinaci: Research Fellow
Petr Knoth: Research Assistant
Annika Lena Wolff: Research Associate
Andriy Nikolov: Research Associate
Alexandra Lilavati Pereira Okada: Research Associate
Alexander Mikroyannidis: Research Associate
Ning Li: Research Associate
Liliana Cabral: Research

Fig. 3. Screenshot of the 'generic' part of the RADAR application, showing distributions of indicators for sub-classes of members of KMi.



Andriy Nikolov (9 publications: 0 journals, 6 conferences, 3 workshops)
Carlo Allocca (5 publications: 1 journals, 3 conferences, 1 workshops)

Journals:
A platform for semantic web studies. Carlo Allocca Enrico Motta Mathieu d'Aquin. WebSci10: Extending the Frontiers of Society On-Line, 2010

Conferences:
DOOR: towards a formalization of ontology relations. Carlo Allocca Enrico Motta Mathieu d'Aquin. International Conference on Knowledge Engineering and Ontology Development (KEOD), 2009
Finding equivalent ontologies in Watson. Carlo Allocca Enrico Motta Mathieu d'Aquin. The 7th International Semantic Web Conference (ISWC2008), 2008
Exploiting semantic relations between ontologies in largem ontology repositories. Carlo Allocca. The 6th Annual European Semantic Web Conference (ESWC2009), 2009

Workshops:
Detecting different versions of ontologies in large ontolgy repositories. Carlo Allocca Enrico Motta Mathieu d'Aquin. International Workshop on Ontology Dynamic at The 8th International Semantic Web Conference (ISWC 2009), 2009

Enrico Motta (81 publications: 4 journals, 52 conferences, 25 workshops)
Fouad Zablith (9 publications: 2 journals, 4 conferences, 3 workshops)
John Bruno Gerard Domingue (77 publications: 2 journals, 41 conferences, 34 workshops)
Liliana Cabral (17 publications: 0 journals, 7 conferences, 10 workshops)
Mathieu d'Aquin (21 publications: 2 journals, 11 conferences, 8 workshops)
Miriam Fernandez (19 publications: 1 journals, 10 conferences, 8 workshops)
Ning Li (4 publications: 0 journals, 2 conferences, 2 workshops)
Sofia Angeletou (9 publications: 0 journals, 6 conferences, 3 workshops)
Stefan Markus Rueger (44 publications: 0 journals, 34 conferences, 10 workshops)
Vanessa Lopez (11 publications: 0 journals, 8 conferences, 3 workshops)
Zdenek Zdrahal (16 publications: 1 journals, 13 conferences, 2 workshops)

Venues with more than 1 publication:
4th Asian Semantic Web Conference (ASWC 2009) [7]
3rd Asian Semantic Web Conference (ASWC 2008) [4]
The 17th International Conference on Knowledge Engineering and Knowledge Management by the masses (EKAW 2010) [2]
The 14th International Conference on Database Systems for Advanced Applications (DSFAA 2009) [2]
1st International Conference on the Applications of Digital Information and Web Technologies (ICADIWT 2008) [2]

Fig. 4. Screenshot of the 'specific' part of the RADAR application, showing distributions of indicators for a particular group of researchers.

knowledge into it is sufficient to make it better structured and more comprehensive. There is an advantage also in having an homogeneous representation for different types of objects, and in having a comprehensive view of all the different indicators and types of data available. It is however a lot more demanding for end-users, as the interface would often include irrelevant elements. Its generic aspect (meaning that it can be applied similarly to other datasets, in other domains) also means that he organisation and navigational structure of the application is guided by the modelling in the data, which is often not natural to end-users. Finally, it makes the visualisations presented harder to interpret, as they do not necessarily refer to the notions considered in the domain.

The specific part of the application was developed to counter the deficiencies of the generic view, by showing visualisations more directly relevant, more understandable and often more complex than what the generic part could do. While this has clearly been welcomed by users, the main issue with this part is that it is necessarily limited not only to the specific data and domain being considered, but also to the views that the developers of the application implemented.

From the developers' perspective, it is natural to think that the generic version of the application is more difficult to build, as it requires to abstract from the domain and data-specific assumptions that can be made with the specific version. It is however also more customisable, as many changes can be brought into it by modifying the data or the ontology on which it relies. This introduces a number of issues however, at the level of the usability of the application as mentioned above, as, without such assumptions being explicitly made, many of the results being shown to the user cannot be properly interpreted.

The specific part of the application naturally suffers from the inverse issues: while its interface design is guided by the requirements of the specific task, it requires significant efforts to be extended to support other visualisations or tasks. It is in this sense closer to the application presented in Section 3.2.

3.4 Investigating the Presence of the Open University in the Media

This last application is based on data collected in addition to what is available from `data.open.ac.uk`. It relies on systems used by the Media Relation services to collect clippings from news items mentioning the Open University and its members. The data collected concern the publication/channel where a news item has been issued, general metadata about the news item and possible additional information regarding researchers, lecturers or other members of staff of the Open University cited or who contributed to the news item. Links are also created to `data.open.ac.uk` (regarding people) and to `dbpedia.org` (regarding the publications and channels that provided the news items).

An application has been created that allows members of the Open University to create charts and reports on top of this data. A 'chart generation' interface is provided that allows to create filters and identify categories to be visualised based on properties and values in the data. Once configured, the interface creates a linkable and embedable chart that can be customised, and is dynamically updated based on changes in the data (see the example Figure 5).

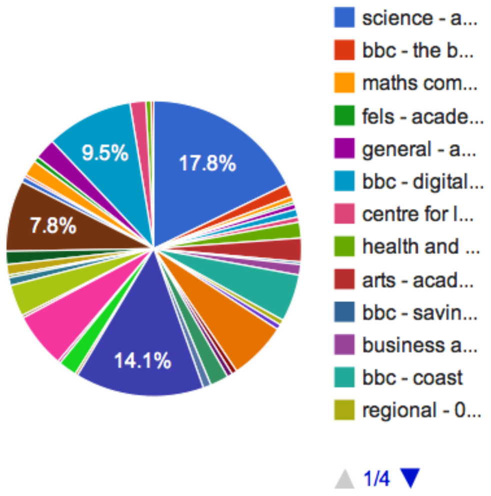


Fig. 5. Charts showing the distribution of topics of news items mentioning the Open University and published through channels owned by the BBC.

Here, two different types of users need to be considered. For the **users of the chart**, the results are reasonably straightforward, as the charts can be embedded into dedicated interfaces that implement dashboards and navigation mechanisms that are understandable in the partiular context. This application also clearly demonstrates the benefits of linked data, especially through exploiting the links with external sources of data (in the example Figure 5, the information about topics and number of corresponding news items is provided by the local data, while the information regarding the channels owned by the BBC is provided by `dbpedia.org`).

The other type of users are the ones **creating the charts** using the configuration interface. This interface requires to be able to understand and follow the properties used in the data for different types of objects. In other terms, while it is generic and can be used to generate charts from any linked data endpoint, it needs for the user to be familiar with the schema used in the data, the particular properties of the data, and with the specificities of their linked data-based modelling.

From the developers' perspective, this application represent an example of a generic application, but which is used to generate specific, customisable visualisations. In this sense, it completely abstract from any specificities of the data being considered, but on the other hand, cannot provide much guidance to the users with respect to the use of the data, and to the interpretation of the results.

4 Conclusions: Challenges and Pitfalls of Interacting with Linked Data, and Steps Fowards

The four applications presented above represent concrete experiences of developments relying on linked data that have been actually deployed and used in an environment more accustomed to proprietary, corporate software relying on more common information systems and data management solutions. As such, they illustrate common challenges and pitfalls that interacting with linked data can generate, from both the end-users' perspective and the developers' perspectives. In this section, we summarise the general notions that appear from such experiences to require attention, which we believe help identifying important directions for research into interacting with linked data.

First, it appears clearly that linked data should be hidden from the end-users. While this might appear trivial, this is not an easy goal to achieve: from our experience, we can see that most of the advantages of linked data should appear obvious to the application developers, but should as much as possible not need to be understood by the end users. This of course concerns purely technical elements such as URIs, RDF and SPARQL, but also more conceptual considerations, such as the integration of multiple sources of data or the use of reasoning. This is the case of our "Study at the OU" app and of the REF support application, where little issues appear on the end-user side. In RADAR and the media relation applications on the contrary, whenever the technology is too present, and even if it is to provide advanced features, it introduces confusion for the users. In other terms, while it is often still needed to convince stakeholders of the value of linked data, applications that are technological demonstrators tend to have little value to the users, and the technology should be essentially be considered from the point of view of the developers.

Second, there is an elusive trade-off to be found between developing generic, reusable frameworks that can be applied on a large variety of datasets and domains, and specific applications that are meant to work only with certain datasets. Indeed, while the value of reusable components is quite obvious, and making this possible is one of the strong benefits of linked data, the RADAR application clearly shows that achieving an intuitively useable application that relies on a generic template for navigation and presentation of the data is close to impossible. Most applications of linked data nowadays are closer to the "Study at the OU" app or to the REF application: applications working in a close environment with a clearly defined and understood set of datasets and queries. These applications tend to be disappointing as they cannot benefit from the openness of linked data and the possibility to integrate data which might originate from other organisations, possibly at run-time. As a conclusion, rather than a complete, generic application framework like what was attempted with the generic part of RADAR, we believe that what is needed are libraries of reusable and highly customisable interface components that rely on generic linked data resources, but can be flexibly integrated to create specific application interfaces in specific scenarios. The chart creation feature represented by the media relation application can be seen as an example of such a library. Other initiatives exist that provide initial building blocks, such as the SPARK (`http://km.aifb.kit.edu/sites/spark/`) javascript library, the SIMILE timeline and map widgets (`http://simile-widgets.org/`), as well as more general visualisation components, such as Fusion [2] or the Linked Data API[2].

Third, it appears clearly that one of the obstacles to building reusable interface components based on linked data is the openness and flexibility of the data model on which linked data relies. At a higher level, integrating data from external sources represent a major challenge, as interaction needs to implement a trade-off between control and the potentially infinite possibilities that opening the interface to unknown data can bring. More concretely, many tasks require assumptions related to the data, which are rarely made explicit and formalised. For example, most of the applications we have considered make the assumption that each entity appearing in the interface are associated with a human readable label. It is often the case that, if not requiring that there is only one human readable label for each entity, the application will make a choice between the ones available, either randomly or using domain- and data-specific criteria. Similarly, while operational and functioning in most situations, it is clear that the charts generated by the media relation application can only make sense under certain conditions. For example, if entities can be associated with more than one identifier, this would generally lead to these entities being counted as if they were multiple entities. Also, charts showing distributions of data would be misleading if not all the entities considered have the same number of values for the visualised property or if some of them do not provide values. In other terms, while the fact that the formalisms underlying linked data do not make the closed-world assumption or the unique-name assumptions is at the basis of their openness and flexibility, it makes them less exploitable as part of generic data processes. Our suggestion related to this process would be to create a way to annotate datasets (including such annotations for example in the Void [1] descriptions of the datasets) related to the particular 'characteristics' of the datasets that can be exploited by data processing/visualisation mechanisms, including for example "local unique-name assumptions" indicating that the instances of a certain class are non-redundant, as well as expressions similar to integrity constraints (e.g., that there is necessarily a value of a given property for the instances of a given class). Being able to rely on such characteristics would make it more feasible for generic interface components, as previously suggested, to provide some levels of guaranties regarding their interpretability and usability when applied on particular data.

References

1. Alexander, K., Cyganiak, R., Hausenblas, M., Zhao, J.: Describing Linked Datasets – On the Design and Usage of voiD, the 'Vocabulary of Interlinked Datasets'. In: Proc. of Linked Data on the Web, WWW 2009

[2] `http://code.google.com/p/linked-data-api/`

2. Araújo, S., Houben, G.-J., Schwabe, D., Hidders, J.: Building Linked Data Applications with Fusion: A Visual Interface for Exploration and Mapping. In: ISWC Posters and Demos (2010)
3. Zablith, F., d'Aquin, M., Brown, S., Green-Hughes, L.: Consuming Linked Data Within a Large Educational Organization. In: Proc. of the Second International Workshop on Consuming Linked Data (COLD), ISWC 2011
4. Zablith, F., Fernández, M., Rowe, M.: The OU Linked Open Data: Production and Consumption. In: ESWC Workshops Proceedings, Linked Learning 2011