

Nodalities

THE MAGAZINE OF THE SEMANTIC WEB



www.talis.com/nodalities

INSIDE

Issue 10

5 Data Mining with
Linked Data

10 Government and Data –
What's the link?

15 A Linked Data Platform
for Innovation

19 Growing Graphs:
Talis working toward
a Linked future

21 Linked Data APIs

22 Triplification Challenge

**Read Previous Issues of
Nodalities Magazine Online**

Want to contribute? Please visit www.talis.com/nodalities
or email nodalities-magazine@talismag.com.

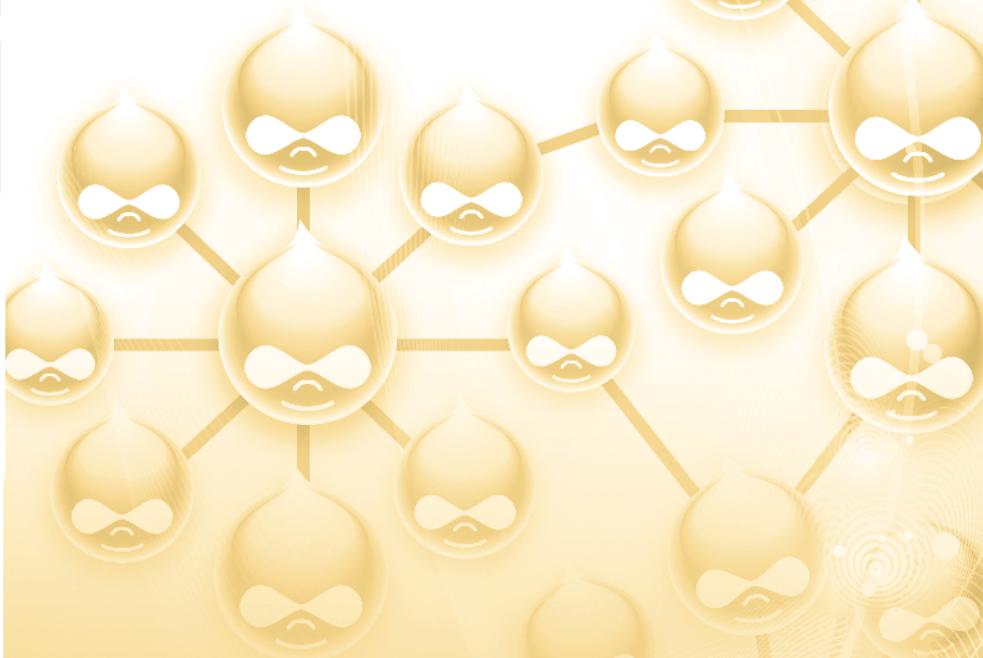


twitter

Follow Us On Twitter
[@nodalities](https://twitter.com/nodalities)

Drupal: Semantic Web for the Masses

By Lin Clark and Stéphane Corlosquet



In a 2008 keynote, the founder of Drupal, one of the most widely deployed content management systems on the Web, said that the Drupal developer community needed to start looking at the Semantic Web and Linked Data. In that talk, Dries Buytaert spoke about how the Web was at the inflection point where many of the pieces were coming together for widespread Semantic Web adoption and he encouraged the Drupal community to start bringing Semantic Web technologies, like RDF, to Drupal. He couldn't have known at that time that his eventual announcement of RDF functionality in Drupal's core would come in a week when that inflection point became clear to everyone else as well.

Between April 15 and April 21, 2010, three events in San Francisco redefined how many Web developers view the Semantic Web. To kick the week off, microblogging giant Twitter announced at its Chirp conference on Thursday April 15, that it would release a new feature called Annotations allowing users to add rich metadata to their tweets. The Twitter development team called it a first step in turning Twitter into a platform upon which developers could innovate¹.

continued on page 3

SUBSCRIBE NOW

Nodalities Magazine is made available, free of charge, in print and online. If you wish to subscribe, please visit www.talis.com/nodalities or email nodalities-magazine@talismag.com.

MEET US

For further information visit: www.talis.com/platform/events

**FOLLOW US**

Follow Nodalities Magazine on Twitter: @nodalities



Nodalities Magazine is a Talis Publication

ISSN 1757-2592

Talis Group Limited
Knights Court, Solihull Parkway
Birmingham Business Park B37 7YB
United Kingdom
Telephone: +44 (0)870 400 5000
www.talis.com/platform
nodalities-magazine@talismag.com



The views expressed in this magazine are those of the contributors for which Talis Group Limited accepts no responsibility. Readers should take appropriate professional advice before acting on any issue raised.

©2010 Talis Group Limited. Some rights reserved. Talis, the Talis logo, Nodalities and the Nodalities logo where shown are either registered trademarks or trademarks of Talis Group Limited or its licensors in the United Kingdom and/or other countries. Other companies and products mentioned may be the trademarks of their respective owners.

Editor's Notes

This is Nodalities' 10th edition, and it's been fascinating to watch the Semantic Web community grow and evolve. Looking back over previous issues, I'm struck by the shift from articles about possibilities—full of language like: "will enable" and "might bring about"—to strong examples and use-case articles introducing services, projects and mashups. It's been an interesting ride, and I'm very happy to be able to publish another edition full of examples and vision.

Massive projects like the launch of huge governmental datasets in the US, Britain and across the globe to the academic developments polishing and quietly innovating on open data have lead the way for bold moves. Drupal, one of the world's most successful publishing platforms has taken the lead among Content Management Systems by baking Linked Data publishing into the core of their soon-to-be-released Drupal 7. In this issue, Lin Clark and Stéphane Corlosquet introduce Drupal's linked future. The message is: just download the latest release of Drupal 7 and you're part of the Linked Data world.

Another potential game-changer covered in this issue is Facebook's toe in the Semantic Web water with their new RDFa-inspired Open Graph Protocol. There are many questions about what this might mean for the wider community, and what such a web giant like Facebook might do with Linked Data, but their 400million users will soon be "Liking" many thousands of linked items seeding potentially billions and billions of real-world triples. It will be an interesting story to unfold, and they've begun boldly with usable, linked data.

Moving from the world of social media to health, Mark Birbeck and Dr. Michael Wilkinson explore their innovative platform for linking medical data together and exposing them through widgets. Clinicians, patients and health professionals could all benefit through better joined-up thinking with the data behind the decisions in health. Elsewhere, Rob Lee from Rattle talks about an interesting data-mining application which uses DBpedia as a control in matching up notable names, facts and concepts—allowing for a huge number of potential matches through language parsing and concept matching.

As more of the world's data are published in reusable ways, it's becoming increasingly important for potential data publishers to understand at least the basics of licensing for reuse. I spoke with the Creative Commons' Prodromos Tsavos about licensing scientific data. I asked him: "What's Common about Science?" and his answer can be read inside.

Leigh Dodds and Richard Wallis also discuss some of the innovations Talis has been working on through projects ranging from government data through to BBC programme data—all growing the graphs of data.

As always, if you have a Linked Data story, I'd like to hear from you. You can contact me via twitter (@nodalities) or email (zb@talismag.com). Subscriptions are free for both print and online versions, and you can subscribe at talismag.com/nodalities.

Thank you for reading,
-Zach Beauvais

Continued from front page.



Just 6 days later, Facebook CEO Mark Zuckerberg announced at the f8 developer conference that the social networking megalith will be using the RDF-based Open Graph Protocol to semantically link sites to Facebook, turning the Web into one large graph for creating a more social, more personalised space. Zuckerberg said it would be "the most transformative thing we've ever done for the Web."²

Between these two surprise announcements, Buytaert gave his semiannual State of Drupal keynote. One of the primary topics was the new RDF functionality in Drupal and how it will change the Web; "I think we can help take the Semantic Web from sort of an academic environment and bring it to the masses. And I think if we do, it really changes the Web... I think it will even redefine companies like Google at some point because it's such a dramatic change," said Buytaert.³

In one week in April, with three events in one city, the Web took a huge leap towards fulfilling the Semantic Web vision. While other recent announcements have energised the Linked Data community, these three new developments enable and encourage the masses to engage with the Web of Data as both producers and consumers. And as these three thought leaders of the Web agree, this means big changes for the Web and how we use it.

This is where the Semantic Web and Drupal are a natural fit—delivering structured content, and enabling users to structure their own content, that's what Drupal does best.

How does RDF transform the Web?

"It will change the way the Web works, and it will mean better search, better search engine optimisation, more targeted advertising, deeper integration between different Web sites—things like mashups, data import, data export, and all of these things," said Buytaert in his speech. While these changes help large data publishers like Twitter to make their content consumable and large data consumers like Facebook to pull in data from other sites, these changes also help the individual data publishers and consumers that Drupal has been empowering for years.

When talking about RDF and what it means for the Web, people frequently say it moves us from a Web of Documents to a Web of Data. In the context of Web content management, this means that instead of linking to a person's Web page, site administrators can pull individual data elements out of that page and process and display those elements on other sites. This allows pieces of content to have a single

This means that a site administrator can create his own kind of page—for instance, a faculty profile page—and add different form fields to collect specific information, like phone number, department, links to courses that the faculty member teaches, or links to students the faculty member is supervising.

Because this information is structured—the individual data elements are kept separate in

Drupal already uses structured data and enables site administrators to create their own structures for data. It was the first of the big three open source content management systems to enable users to create their own content types and to do it without any programming necessary

home, where the content can be updated and maintained, while that same content is used in many places across the Web.

For instance, a person may want to include contact and biographical information in many places, such as on an employer's site, on a personal site, and on an alma mater's alumni page. If the sites use and understand RDF, on a personal site that person could say the equivalent of, "I am a person. I have this email address, this phone number, I am based in this location, and this is my bio." Then, simply by putting a link to this page, the employer and alma mater could pull out the information that they want to display (for instance, the bio and email address) and format it however they like. And the great thing is that if information changes on the personal profile page, subscribed sites can pull in the updated information.

RDF enables this by breaking down the information on the page into small, meaningful elements which have unique identifiers. Standardised yet flexible relationships between these elements can be set. HTML web pages can embed RDF in their markup, thereby exposing the structure of the page's information which is expressed in those standardised relationships.

So RDF makes the semantic structure of the information on a page explicit. This is where the Semantic Web and Drupal are a natural fit—delivering structured content, and enabling users to structure their own content, that's what Drupal does best.

Drupal—Structured Data for Everyone

Drupal already uses structured data and enables site administrators to create their own structures for data. It was the first of the big three open source content management systems to enable users to create their own content types and to do it without any programming necessary (WordPress and Joomla! are just adding that functionality now).

different database columns—you can search, sort and filter on this information. For instance, you could create a list of faculty members that teach a specific course, or a list of students who are co-supervised by faculty in two particular departments. This kind of easy, sophisticated information handling, enabled in Drupal by the popular Views module and others, has been one of the features that has driven Drupal to

Drupal allows users to search, sort and filter their data and specify how and where it should be displayed, all without having to code anything.

success in the content management market.

So by enabling users to structure the information in their sites, Drupal allows users to search, sort and filter their data and specify how and where it should be displayed, all without having to code anything. This enables knowledge workers with a sophisticated understanding of their content model to dictate the data architecture, rather than relying on a programmer who has a limited understanding of the domain.

Drupal+RDF—Web of Data for Everyone

Drupal has opened up the Web to content creators across the world. It is used on hundreds of thousands of websites and powers 1% of the Web. Because Drupal enables these content creators to manage their own structure, the information that these sites contain is very well structured internally.

But so far this structured data has not been exposed to the Web and external applications in a meaningful way, and dynamically pulling content from other datasources into Drupal

is also still cumbersome for non-developers. While those with coding experience are able to connect to outside data sources using data source specific APIs, non-developers are limited to the data in their database.

Because RDF is standard, yet flexible, content administrators can focus on their content model, rather than dealing with site-specific APIs for content import and writing customized formats for content export.

With RDF in Drupal 7 core, this all changes. Because RDF is standard, yet flexible, content administrators can focus on their content model, rather than dealing with site-specific APIs for content import and writing customized formats for content export.

Just as creating dynamic sites with pages generated from a structured database backend used to be the province of Web developers alone but is now easy for anyone to do

Creating sites that add to and exploit the data in the Linked Data Cloud is coming within the reach of non-developers.

The easiest way to get started with Drupal and RDF is to simply download and install the latest Drupal 7 package

How to get started with Drupal and RDF

The easiest way to get started with Drupal and RDF is to simply download and install the latest Drupal 7 package⁴. By default, your posts' titles, dates and content will be marked up with RDF, as will the tags on those posts, the comments on those posts, and the users who posted the comments.

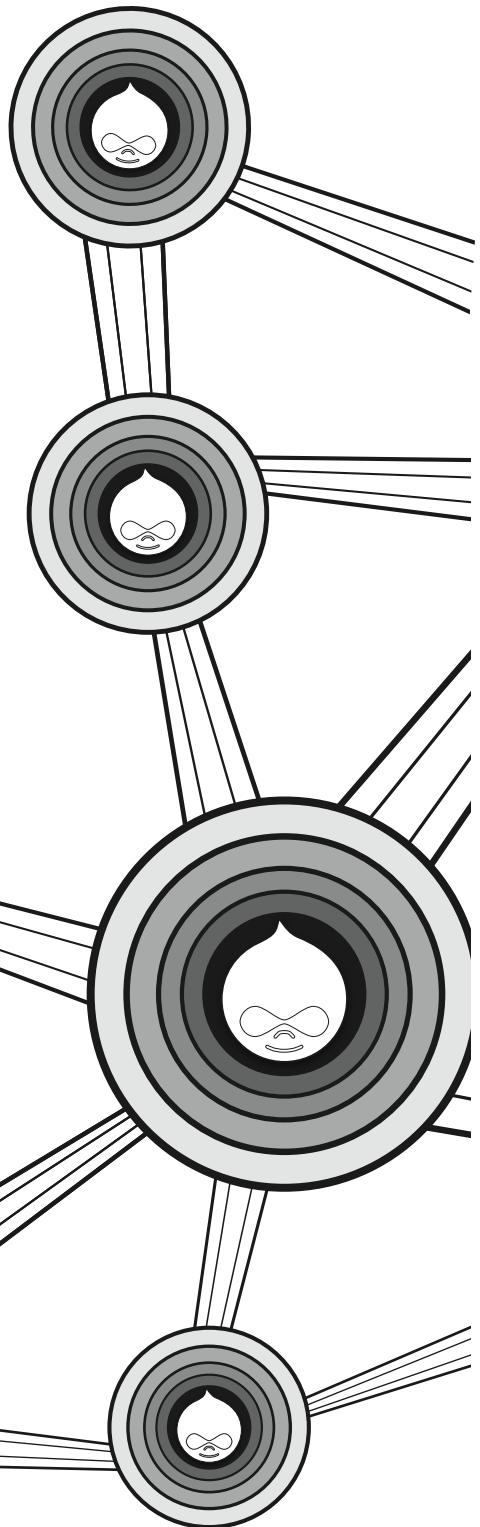
Drupal 7 includes generic RDF mappings for the most commonly used content types like articles, blog posts, and forums. It also allows Drupal based distributions to predefine their

data structure using RDF. These distributions enable one click creation of highly tailored sites for specific use cases like project management (Open Atrium) or online publishing (OpenPublish). As a result, sites powered by RDF enabled Drupal distributions will expose their content using the correct semantics "out of the box".

Semantic Web savvy site administrators will also be given the option to customize the RDF mappings their sites use to match the type of information that their sites contain. You can learn more at the newly launched Semantic Drupal website⁵, which features screencast tutorials and the latest happenings on the RDF in Drupal scene.

Lin Clark is a masters student at DERI, NUI Galway

Stéphane Corlosquet works at MIND Informatics, Massachusetts General Hospital



1 http://groups.google.com/group/twitter-api-announce/browse_thread/thread/fa5da2608865453

2 http://news.cnet.com/8301-13577_3-20003053-36.html

3 <http://sf2010.drupal.org/conference/sessions/state-drupal>

4 <http://drupal.org/project/drupal>

5 <http://semantic-drupal.com>

Data Mining with Linked Data

By Rob Lee, Rattle



Now that the Linked Data cloud has been firmly established, organisations can start to look at different ways to use and explore the data. At Rattle, one area we've been keen to examine is using Linked Data to extract meaning from existing unstructured content. This has involved looking at both Data Mining and Natural Language Processing (NLP) techniques. NLP enables machines to analyze, understand and generate natural (human) languages. By using Linked Data to form part of the NLP process, it's possible to start identifying key concepts in content and begin relating them to their Linked Data counterparts. These concepts can then be used as part of a larger data mining process, typically extracting, aggregating and linking to them to provide different views at the content being analysed.

Linking Concepts

The initial work has concentrated around DBpedia, a key part of the Linked Data cloud, derived from Wikipedia. DBpedia is used as a controlled vocabulary of notable 'things' or concepts. Anything in Wikipedia should by definition be notable, and by using DBpedia as our list of notable concepts, a comprehensive general purpose vocabulary that can be used to link to concepts in content.

This work has resulted in a web service that would accept text-based content and use DBpedia as the controlled vocabulary to link to the notable things in the content. In some respects it is similar to services such as Open Calais or Zemanta; however the service has attempted to blend traditional NLP techniques with the extra data available from DBpedia to improve the accuracy of the results. This service ('Muddy') has been released at <http://muddy.it>. Please try it and let us know what you think.

A common issue encountered when building NLP systems is that the English language is very ambiguous. This makes trying to find the correct sense of a term difficult.

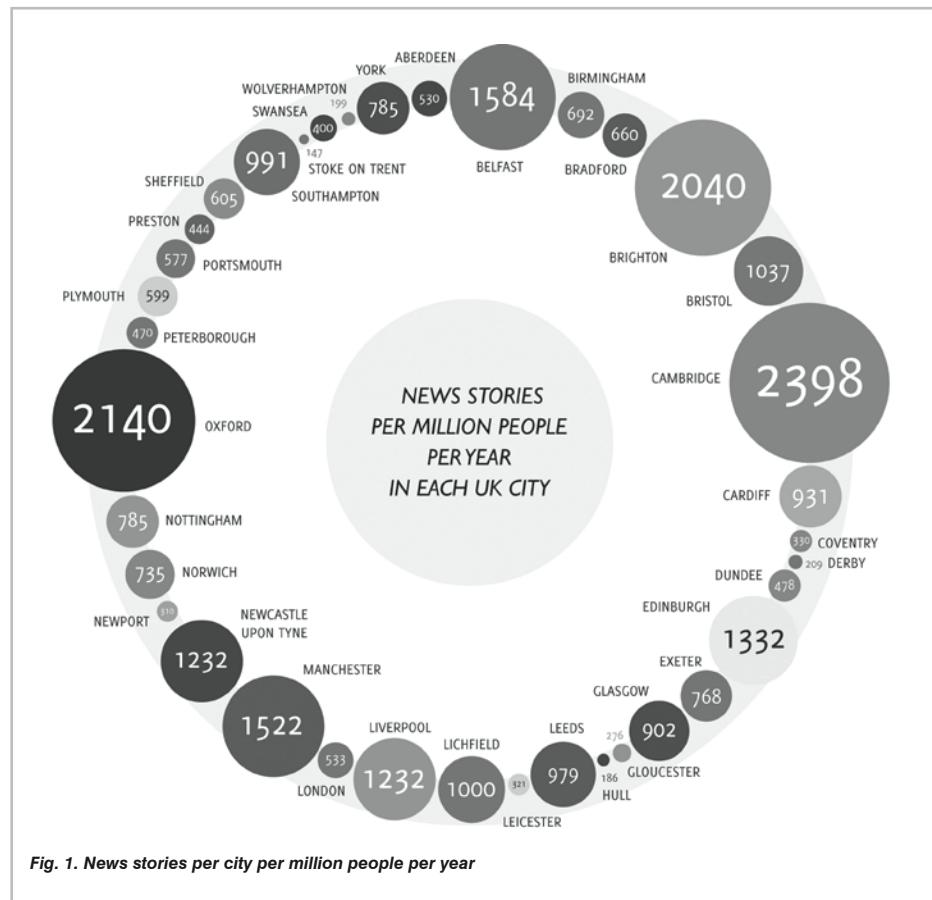
A common issue encountered when building NLP systems is that the English language is very ambiguous. This makes trying to find the correct

sense of a term difficult. Using Linked Data can help here, e.g. 'Apple Inc.' is described by http://dbpedia.org/resource/Apple_Inc, whereas 'Apple Records' is described by http://dbpedia.org/resource/Apple_Records. So if we are trying to describe the term 'Apple' in the statement "Apple produced the first iPhone.", as long as we disambiguate the term to the correct URI, we will have an unambiguous description.

DBpedia can also help with the process of disambiguation as it provides common disambiguations for a given term, via the associated Wikipedia disambiguation page. The challenge here is picking the correct disambiguation.

DBpedia can also help with the process of disambiguation by providing common disambiguations for a given term, via the associated Wikipedia disambiguation page. The challenge here is picking the correct disambiguation. The service uses a number of different metrics derived from DBpedia and from more traditional NLP methods in order to do this.

Disambiguation is all about context, so in an article about 'Apple Inc' you would expect to see mentions of things like 'Computer' or 'Steve Jobs' whereas in an article about 'Apple' the fruit, you expect to see mentions of terms like 'Braeburn' or 'Food'. A content-similarity metric is used to find the most relevant disambiguation, by comparing the DBpedia abstracts of the disambiguations to the source content. When combined with other metrics, this allows the most likely disambiguation to be chosen. The number of inbound pagelinks from DBpedia for a given disambiguation is another important statistic, the 'link popularity' of a concept being factored into how likely it is to be the correct disambiguation.



Hopefully, this illustrates how Linked Data can complement existing technologies, enhancing the methods and processes used in other disciplines.

Data Mining the BBC

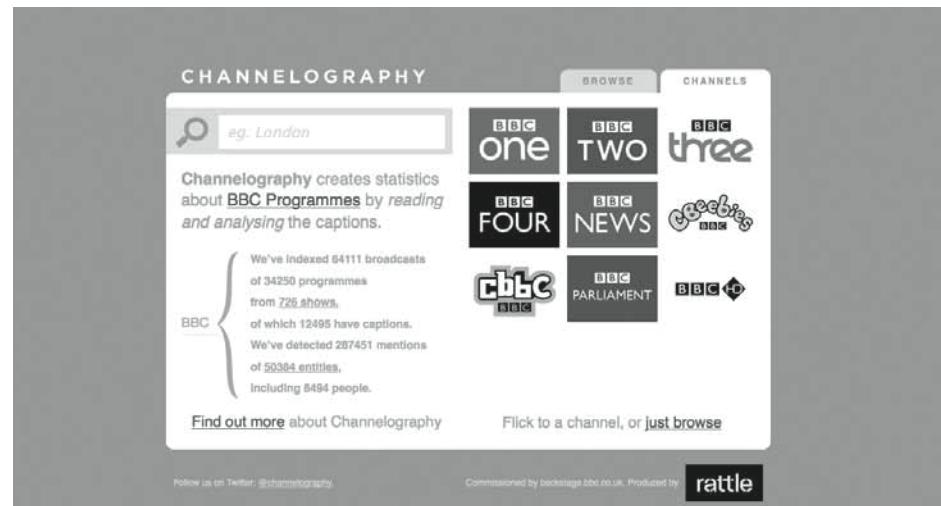
At Rattle, we've been interested in closed caption data for broadcast programmes and how they could be used beyond their original purpose. The BBC's iPlayer service provides closed captions for many of its programmes, and we have started to look at ways they could be used to say something about each of the channels that the programmes were broadcast on.

Channelography¹ was built to test some of these ideas. It combines the programme Linked Data published by the BBC² with the Muddy concept extraction service³ and the closed caption data from iPlayer to start describing and saying something about BBC broadcast television channels via the data mined and aggregated from these services.

Simple aggregation of the BBC's published

Simple aggregation of the BBC's published linked data alone provides interesting statistics, for example, looking at repeat broadcasts, aggregated at a channel level, approximately two-thirds of BBC Three's output is repeats.

Linked Data alone provides interesting statistics, for example, looking at repeat broadcasts, aggregated at a channel level, approximately two-thirds of BBC Three's output are repeats. When different Linked Data sources are combined, more revealing statistics can be produced. For example, by aggregating the people mentioned on a channel, we can determine how balanced gender representation is on each channel. As every person identified in the closed captions will be a DBpedia resource, linking to other data sources becomes possible. In this case Freebase is used to obtain the gender (as it has better coverage than DBpedia in this instance), the BBC's TV schedule data can be used to link a particular programme to a channel (or channels) and we then calculate the overall gender representation of people mentioned on that channel via the broadcasts. These figures can then feed into management information for a channel: is it meeting its public service requirements? Is it appealing to its target demographic?



Something we've discussed implementing for Channelography is examining the genres of the broadcast content, so what proportion of BBC One's output is 'soaps' and what proportion is 'factual' and how does this

As the figure shows, London if anything is under-represented, whilst we see some large figures for Oxford and Cambridge, presumably due to their famous universities.

Whilst this was a simple analysis, it starts to show how using Linked Data can be a useful tool in data mining, both from the perspective of having a curated set of data and the added value provided by the 'linked' nature of the data.

The next step for this project is to use geographic bounding data from other Linked Data sets in conjunction with increasing the size of the controlled vocabulary of places to any place in the UK, which would result in a more accurate representation of news coverage by encompassing towns and villages, rather than focusing on cities.

Where Next?

There is a range of data mining algorithms and techniques that could benefit from incorporating Linked Data in their approach, whether as controlled vocabularies, training sets or for algorithm development. Practical uses could include training classification algorithms, (e.g. Naive Bayes, Support Vector Machines) using existing taxonomies from Freebase or DBpedia. Conversely, there is a number of Linked Data sets ripe for further data mining and analysis, e.g. cluster analysis of data.gov.uk or visualisations of Pagerank within Freebase. Expect to see an increase in the fusion of these fields as Linked Data continues its rise in popularity.

Rob Lee is the CTO at Rattle (<http://rattlecentral.com>).

1 <http://channelography.rattlecentral.com>

2 <http://www.bbc.co.uk/programmes>

3 <http://muddy.it>

4 <http://www.rattlecentral.com/blog/2010/02/the-newsography-experiment>

What's Common about Science?

Creative Commons legal project lead, Pro Tsiavos talks with Talis.



In April, Talis published a podcast of Zach Beauvais speaking with Pro Tsiavos of the Creative Commons about data licensing and the Science Commons. The full podcast can be found on the Nodalities Blog (<http://bit.ly/protsiavos>). The following is an extract from that conversation, and begins with Pro's answer to the question: "What's common about Science?"

Pro Tsiavos: What's common about science is that there are many differences from scientific area to scientific area. It's hard precisely because there are such differences. It is important to actually be able to have, within the different sectors, the data to flow freely. Science is based on people being able to use other people's research and other people's data. To be able to use other people's research essentially means you have access to their published papers, and that's something which has been achieved a great extent through the Open Access movement, and through legal instruments like the classic Creative Commons licences.

So it is necessary for people to be able to share data in the most frictionless way, more frictionless than it would be for content.

However, increasingly, you need to be able to also share data, the data that are behind this research, and such data sharing poses an entirely different, or quite different if not entirely, challenge. So it is necessary for people to be able to share data in the most frictionless way, more frictionless than it would be for content.

Zach: OK. So in the same way that the Creative Commons licences have been applied to lots of different types of creative content—anything from millions and millions of Flickr photos, which anybody can make use of, to actual whole publications like our own Nodalities Magazine. So that works for content, but you're talking about data itself. So, publishing the stuff that you work with in order to get the publication, is that what you mean?

Pro: Yes, exactly. Also, other stuff as well,

experimental data, steps or procedures for conducting experiments. Also, increasingly you have a lot of what we call a backdoor of web pages that are being used in scientific projects, being fed by data which exists in raw form, like data sets. So you have increasingly this plethora of data that needs to be shared in the easiest possible way.

Zach: Right, so you're talking about it being frictionless, so that if I published a bit of information, somebody else who's doing a similar or even a dissimilar experiment can make use of the data that I'm publishing.

Pro: Precisely.

Zach: OK, so why can't they do that with Creative Commons licensing?

Pro: They can. But what the difference here is that when you have data there are several issues. One of the issues that you have has to do with the variations of the licences. So, for instance, some of the Creative Commons licences are for strict commercial use. We don't think from the experience we had with scientific projects, we don't think that's a good practice in the data context because it would cause all sorts of walled gardens and sort of data islands, and essentially would enclose the data in ways you wouldn't like them to be enclosed.

Similar considerations you have with the share-alike—at least to my opinion—provisions. So, it's not necessarily good to have copyright provisions in open data licenses.

Finally, you have the issue of attribution with the attribution stack problem. It's precisely because you may be getting data from multiple sources, you may end up having difficulties in how you attribute if you have to attribute all your sources.

So in other words, a Creative Commons licence could work, but the licence that would actually be the best for scientific data would be licences that do not impose any restrictions. So public domain indication type of licence sources are zero type of licences. That's consistent with the other principles for open data and data in science.

Zach: So that allows you to essentially publish information, the raw information, behind the project that you're working with in a way which you can guarantee that you're not going to sort of hold anybody into account, whether or not they attribute it, how they use it or reuse it?

Pro: Exactly.

Zach: OK. So are there any implications in that, from your perspective, for provenance, things like not knowing where data is coming from?

Pro: That's a really, really important issue. I think right now with the CC licences, you don't have any provenance problems because you have to attribute. So you always have to say where your data came from.

Creative Commons license could work, but the license that would actually be the best for scientific data would be licences that do not impose any restrictions.

Now, let's say that you have a data set that has been altered substantially after a series of rounds or fluctuations, you can always know which was the original data set by going to the data.gov web page. So if someone wants to assert whether this data set is the original or not, it's quite easy to do. You can always go to the source. There's a single place where you can find the data and they're indexed.

If we go back to your question about what constitutes open, and I mentioned one of the key issues is that data are indexed, that's why it's important to have a point where the data depart from and a point where the data are actually stored and preserved.

Of course, that's the situation right now. It could be true in the future by providing services that actually can trace the data even through voluntary schemes or services where people can deposit original data sets and somehow stamp them. We have such services in the area of non-data content, and they're quite successful.

I haven't seen anything similar in the case of data as yet, but I'm quite confident that we're going to see those services in the future. I think that's something to be watched. The people that are most probably going to do that, I think, are going to be libraries or archives in relation to their metadata.

What we've seen recently is a great discussion about this issue of provenance and how you trace provenance in the context of Europeana, the European digital library where

there's a discussion about how the metadata about all of the European culture resources are to be shared and how you're going to achieve provenance. But we don't have any conclusive results as yet.

Zach: OK. It sounds to me like if I have an interesting data set that I've either come across or created and I want everybody to know: "This is what I've done; this is the original, this is mine." If I don't actually publish that myself online, then it's going to be difficult to check whether or not the data set is derived from it.

Pro: If you don't have it at the fixed point where it can be indexed, clearly not.

By having them in a place where it's kind of permanent, you facilitate other users in their search of the source and whether they want to verify some qualitative aspect of data, how they're being collected, or who has collected them, and what they involve.

Zach: OK. So do you think it's important to not only have these licenses around your data, but also to publish the sort of raw data yourself, to take responsibility, to have the data out there in a form that you want people to know?

It's good practice, both for the subsequent users and the data producer and originator to have the data at the point where they can be indexed and they can be found, and the original quality may be asserted.

Pro: It depends on the case, but my personal feeling is that that's what you should be doing. You should take responsibility for all the data you're publishing. By having them in a place where it's kind of permanent, you facilitate other users in their search of the source and whether they want to verify some qualitative aspect of data, how they're being collected, or who has collected them, and what they involve. I think it's good, also, for the data producer to do that because that's the way my data are always going to be... Other people are going to find better ways to use my data than I have, even photographs. So it's good to put my data out, but these people that are reusing my data and they may do very novel things with my data, they may have an interest in getting attached with me.

They may have an interest to know where the data came from. So it's good practice, both for the subsequent users and the data producer and originator to have the data at the point where they can be indexed and they can be found, and the original quality may be asserted.

Zach: It also sounds like you get a lot more sort of rich use out of it because, as you said, if I'm publishing data but I'm also publishing my methods and my own context and reasoning, then whatever you're going to do with my data is going to be that much more rich. Because you can say: "This is a data set; it was collected in these ways, from this person, during this time period," rather than just: "There's some data

that we happen to have."

Pro: Yes, exactly. I think that's why services like myExperiment are, I think hugely successful in the set up or at least they have huge potential.

Zach: Tell us about that.

Pro: So myExperiment is a project which is run by Southampton University, among others. The idea is that you conduct an experiment and there are certain stages that you have to follow in order to conduct such an experiment. Then, once you do, you actually share the workflow, you share the data, you share the results, you share the papers, and you choose which licence you want to share things under. It could be that you've decided to use Creative Commons Attribution for the content, or CC 0 for the data, et cetera. But you essentially put them out there as a package or components that would be compiled into packets and other users can take this packet and disassemble it into components and reuse it again. That's very useful for the scientific community.

Zach: I was going to say it sounds very scientific. From my understanding of science, the whole point of an experiment is that you can replicate it.

Pro: Precisely. That's why I think it's such a good project.

Zach: OK. So this licensing is quite important to get right then because if I want to replicate a project and I find differences in it, I need to be

able to compare my data to the original and not worry that what I'm doing infringes on any sort of attribution.

You need to have a technology platform that allows people to do things in a very easy way. You also need to have access to the relevant scientific communities that are going to use the platform, so you have to engage them in order to use your platform.

Pro: Yes, the way the project runs right now is that it gives you a set of licences and pretty good guidance as to how you're going to license things. So there is a platform. It gives you all kinds of tools that you want to have in order to actually both produce and share your scientific results. I also have to mention that Southampton and Manchester Institutes funded this project.

Zach: OK.

Pro: So this project is a social networking—essentially—project for science. I think it shows that in order to get it right, you need to have a lot of components besides the licences. So licences are important, but you should have guidance as to how they're going to be used. You need to have a technology platform that allows people to do things in a very easy way. You also need to have access to the relevant scientific communities that are going to use the platform, so you have to engage them in order to use your platform. If you get all these things right, then you can extract value from data. I think just pulling data out of the air is useful, but it's just very first step. It's important to create the structures for allowing the people to share data.

Also, it's important to create the processes. I think that's another thing that we have to work on, particularly in the public sector, on the process of how you release data openly and how then you allow people to reuse them.

Zach: So how do you think this is going to change how we do science?

Pro: I think, first of all, you have to be much more vigorous in what you do. Precisely because of increased transparency, you're constantly being checked. So, secondly, I think

what it should change is the way you're being evaluated. So right now, you're evaluated on the basis of your publications. But we need to incorporate evaluation schemes for people who are using raw data and making them available. So it's not just how many times you're quoted, but also how many times your data are being used. I think that will be a major push for people to share more data.

If you see most of the studies about innovation in the country, they count the patent applications per university, and that has been fed into the assessment of research in the various national assistance projects in the UK.

Zach: So that's actually adding a sort of quality of the information itself then. Because you're not just measuring how well-put your arguments are, or how you formulated your hypothesis; you're actually putting a metric on how often your data is used, which means your data has to be good, doesn't it?

Pro: Yes, exactly. I think if we really want to promote science, we need to promote such practices. I think that's a task for our research consults and how they decide who to fund. So we could leave funding options of publications to release of data and evidence of reuse of data. I think that would be a really important step to encourage organizations to open up data. The second thing we have to do internally within the universities is called the "exploitation unit." So the IP officers in the universities normally, and research assistants, tend to think in terms of patents, how they're going to patent stuff. So there is a culture of enclosure right now because that's how innovation is measured traditionally. If you see most of the studies about innovation in the country, they count the patent applications per university, and that has been fed into the assessment of research in the various national assistance projects in the UK.

Now if you have a different approach, which is to assess research not just on the basis of patents, but also on the basis of data being used, produced, and being reused, and is quoted, then you also have people as IP officers that have a different mindset, which is how you're going to facilitate your researchers so that they open data and make their work known, rather than closing data.

At the same time, I think on a national basis, you can have a more incremental way of developing science, which is how it should be in the first place.

Zach: It sounds like it would change the whole way that one goes about getting grants, as well. So if I'm applying for a grant to do something, I don't have to just show my track record of how I've been quoted on things. I can also say, "Well I've produced these data sets, and what we're going to produce at the end of this is data that can be used by all these projects."

Pro: Precisely. I think this is why we tend to think I'm going back to your first question of what's common in science or in general, what are the Commons or how the Science Commons differ from Creative Commons or cultural commons, or governments, or data. I think we should think of the whole thing more realistically and see what the role of the administration or the granting agency is. This is one thing that research institutions do and how the whole life cycle works. So the government has the responsibility to facilitate the opening of data, and in order to do that, it has to link grant applications with opening up of data and the reuse of data sets, and also provide the infrastructure to do that.

We should think of the whole thing more realistically and see what the role of the administration or the granting agency is.

Zach: So does this change the nature of science itself in relation to other things? I'm thinking journalism, for example, which makes use of scientific publications all the time in reporting facts and figures.

Pro: I think it does precisely because we see people that are not experts in the sense that they haven't received any formal education, in the specific fields. Because of their ability to have access to data and content, and also to be able to evaluate it, and practise, they may provide feedback in different ways to scientists. Also, in terms of how journalists do work, there are some interesting projects right now about the way of assessing the quality of journalistic work, so assessing the quality of an article. Does this person adhere to certain principles? These principles normally refer to unbiased coverage of an event, or where you obtain your data from. In that respect to the question, where

did you obtain your data from, then it becomes relevant how the data have been produced. So what we see is you have this kind of constant evaluation of the work, of people by other people who are not necessarily within their same working area, working fields. This means that you have cross-pollination, which I think is very, very valuable.

An area in which I think we're going to see and we're already seeing things happen and we're going to see even more in the future is medical research, and medicine in general, precisely because you have patients that discuss their diseases with each other or their health problems with each other, and they form groups.

We're already seeing things happen and what we're going to see even more in the future is medical research, and medicine in general, precisely because you have patients that discuss their diseases with each other or their health problems with each other, and they form groups.

They frequently know more about side effects of drugs than their physicians. They know more about potential treatments of a particular ailment than their GPs. So when they go to their doctor, they already have a knowledge that they have obtained through their discussions and interactions on the net. Where the doctor is to provide them with treatment, it is no longer the expert that talks to someone who knows nothing. It's an expert that talks to someone who's very informed, which means the doctor's practice is going to be very different.

So this is an example where we see clear interaction between the experts and the non-experts in a more equal fashion. We may see similar things in science, as well. I don't want to be very prosaic, it could be this is not necessarily every stream of interaction. It may mean it's more difficult to convince the patient to receive a treatment which may be good for her.

But at the same time, this kind of dialogue between experts and non-experts, I think, reintroduces controls that were not there before. In a sense, it makes the work of the experts more rigorous and better. At the end of the day, it's good for science.

Government and Data – What's the link?

By Emer Coleman, Greater London Authority



The past six months have been a good time for the Open Data movement. It's seen the launch of the London Datastore and Data.gov.uk and when you throw in the Ordnance Survey consultation and at least the partial resolution of the issue of derived geographic data, it seems like things are really changing.

But before we step too far ahead, it's probably worth taking a few steps back to really question the context in which this is happening. In the Greater London Authority (GLA), we have maintained for some time that the release of data, like the introduction of new technologies and social media into the public sector, is not about bolting 21st century technology onto 19th century governance. Its about nothing less than a fundamental change in the paradigm of government, and attempting one without the other represents both a flawed approach and the great possibility of things remaining the same.

For the past twenty years in the public sector across the globe, government has practised a system of management known as New Public Management. You will recognise it: a target-driven environment with a mixed market economy and a rational approach to policy making known as Evidence Based Policy. The latter is based on the assumption that policy makers can amass statistical evidence to assist them in naming the problem. It also assumes this problem can be analysed and a rational course of action taken to solve it.

But it might as well be called Policy Based Evidence, because the most difficult social problems that the public sector face cannot be addressed with this approach. This is because it is based on the assumption of rational man and we know that while policy makers think of us as Dr Spock, who will do what is advised and in our best interests, the reality is that we are all more like Homer Simpson. How often do we promise that we are going to go to the gym and take the exercise that we know we should and then simply go home, sit on the couch eating pizza with beer aloft?

Obesity, for example, is not a problem that can be tackled through rational action because what drives people to overeat and under exercise are complex desires and values that are not often amenable to the instrumental action of the state. What is needed is a much greater understanding of the human condition and the acceptance that for policy makers, the greatest learning opportunity exists in

failure. But, the one thing one must never do as a public servant in the world of New Public Management is fail. The public and the media will simply not tolerate governments experimenting through trial and error to see what course of action might have the best chance of success.

The public and the media will simply not tolerate governments experimenting through trial and error to see what course of action might have the best chance of success.

Think how ludicrous this approach would be in the scientific, technological or business world? How many experiments, moments of failure, trial and error approaches did it take to make something as complex and wonderful as an iPhone?

In the GLA we are arguing that the state needs to move away from a New Public Management approach to one of Communicative Governance. One where there is an acceptance that the state does not know best and that the state and citizen need to engage in much more meaningful dialogue if we are to solve together the increasingly complex public policy issues we face.

So the release of data and the use of new technologies in the public sector needs to be understood in the context of a fundamental shift in the paradigm of governance. We must move away from the broadcast mode of telling

We must move away from the broadcast mode of telling people what to do, toward the engagement model of understanding why they do the things they do, and work consensually to reach mutually agreed goals.

people what to do, toward the engagement model of understanding why they do the things they do and work consensually to reach mutually agreed goals. At the core of this is the acceptance that while the state may have the monopoly on the collection of certain data, it no longer has the monopoly on the interpretation of that data nor the right to dictate how the

data should be used for fear that others might see things differently to what we do as public servants.

And it may mean—and most certainly will—that concerned and committed citizens might point out mistakes, might unmask failure. Surely that should be welcomed as a good thing? Might it be possible to envisage a grown-up citizenry who can engage in mature debate and who could help us to do things differently and better?

The evolution of the London Datastore is an attempt not just to release London's public sector data but to put into concrete practice the theoretical model of communicative governance. We started from the assumption that releasing our data would be a good thing. We set a simple rule: if the information was available under a Freedom of Information request, it should be available without asking for it.

From the outset, we wanted to work in collaboration with the community most likely to use this information. In the first instance, this is the software development community in London who have been calling for this data release for many years. We sent out a call on Twitter and on the web to "come help us free London's data." And, guess what? They did. In the spectacular setting of London's Living Room in City Hall sixty of them gave up their Saturday morning to tell us what they wanted, advise us on what to do and what not to do and exhorted us to simply get on with it.

Seven months on, I have learned some things about software developers in London.

Firstly, never schedule a workshop early on a Saturday morning (hackers don't do early mornings). But much more importantly, I learned about the amazing innovation and creativity they bring to the table and the

incredible generosity of spirit that exists to do good things and make people's lives better. Their advice, suggestions and most importantly support has been eye opening.

In our model we have our colleagues in the GLA, but I see no difference between them and the developers who work outside the organisation. On this project, we hold each

other equally to account and I am clear in my understanding that my role in pushing agencies to deliver their data into the Datastore is to deliver for both the Mayor of London and for the people who have been part of the project from the beginning.

In the past seven months, while meeting with Transport for London, the Metropolitan Police, the London Development agencies and others, I have brought the developers with me to participate in the discussions. This in itself has been an interesting example of communicative governance in action. On one side of the table sit the developers as zealots

there. When we get a good dataset, I can call them up and they will remind me how far we have come. And when it goes cold and nothing seems to be happening, I can Twitter out and let them know that just because there are no new datasets coming in right now, it doesn't mean nothing is happening.

For me this is about making the process of governance visible by showing the workings of policy development. Otherwise, how will anyone be able to stare through the glass dome of City Hall and call me to account? How will they know that though the wheels of government move slowly (and boy do they move slowly

The coming years for the public sector will see huge changes—data release will be a big part of this change but only if the necessary cultural changes happen in tandem. Its time for the public sector to understand that the future is life, Jim – but not as we know it.

and evangelists for open governance, while on other sit public servants constrained by both culture and history. Bringing both to a middle point has been an interesting journey. When I get immensely frustrated at the slow speed of progress, I can call the developers and rant and rave, and they will remind me that we will get

sometimes), that they are nonetheless moving? Who would have thought that there is such good will to help and support a changing approach to governance and that there are citizens—because developers are also citizens of London—who are willing to give of their time and expertise for free to enable progress?

Think for a moment what the public sector might look like were central and local government to begin to open up. Yes, their data in the first instance—in a spirit of openness that encourages use, debate, dialogue, challenge and criticism. How quickly might we in government be able to adapt and change to improve things? We might, through the technological innovation we can leverage, stop doing things we have always done. We might cede power and let the citizen do things for themselves without our help and we might therefore be able to manage scarce public resources to focus on working with the most vulnerable. We could provide the services they determine they most need themselves.

The coming years for the public sector will see huge changes—data release will be a big part of this change but only if the necessary cultural changes happen in tandem. Its time for the public sector to understand that the future is life, Jim – but not as we know it.

Special thanks to: @jaggaree, @countculture, @chychau, @dominiccampbell, @paul_clarke, @hubmum, @tobybarnes, @MadProf, @osbor nec, @tomskitomski, @londondastore

Emer Coleman is London Alliances Project Director at the Greater London Authority.

Read Previous Issues of Nodalities Magazine Online

Want to contribute? Please visit www.talis.com/nodalities or email nodalities-magazine@talismagazine.com.

The screenshot shows the homepage of Nodalities Magazine. At the top, there's a navigation bar with the Talis logo, links for 'About Nodalities', 'Previous issues', and 'Nodalities Blog'. Below the header, the title 'Nodalities' is prominently displayed with the subtitle 'THE MAGAZINE OF THE SEMANTIC WEB'. To the right, a thumbnail of an issue of the magazine is shown, featuring the title 'Open Government Data' and a photo of the US Capitol building. The main content area includes sections for 'Contact us' (with phone number +44 (0) 870 400 5000 and 'Email editorial'), 'Previous issues' (with a thumbnail of another magazine cover), 'Nodalities Magazine' (describing semantic web technologies moving beyond the laboratory), 'Want to subscribe?' (with a 'Subscribe for free' button), 'Want to contribute?' (with a 'Share your ideas for future articles' button), and 'Updates' (with a 'Follow us on twitter' button).

`@prefix`
`datatype#`
`?RDE.`
`sparql#`
`:name @dc#`
`api.talis.com`
`foaf:mbox#`
`?RDfa`
`?vcard`

We're always on the lookout for talented and pioneering people
to join with us and help us push our boundaries even further.

If you're interested now, or could be in the future, please visit our
website where our vacancies are advertised or to send a speculative
application, please use the email address below.

talis.com/careers
catherine.mayo@talis.com

talis[®]Platform[™]
data, connected

Facebook and the Open Graph

By Zach Beauvais



In April, I was watching the twitterverse explode during the Facebook's f8 conference, watching a steady stream of links and gasps and applause, as well as intentions to delete profiles. My initial reaction to quickly-scanned third-hand reports was essentially: "Oh no." The message I was getting was that public profiles would be made more public, and that more places would start sporting the little fb: "like" box you see on sites using Facebook Connect. I was concerned because there have been many conflicting messages around facebook and privacy¹, and hearing about this movement to include a wider presence online would essentially pull more people into a huge walled garden.

Facebook Links Data?

Watch the f8 video sessions², though, and some interesting things begin to emerge. The main announcement at f8 is an update to Facebook's API. Indeed, it wasn't so much an update as a rewrite, moving from an older and complicated SOAP architecture with Facebook Connect to a more RESTful approach, giving services a simpler and more straightforward way to interact with content within Facebook using http. OK, so this isn't particularly ground-breaking, nor is it very exciting in itself. What is far more interesting, is that Facebook's engineers are using this word graph³ to describe their ecosystem with the launch of the Open Graph API.

Their new Open Graph API is built to enable social plugins which let users on other sites pull content into Facebook.

Firstly, their new Open Graph API is built to enable social plugins which let users on other sites pull content into facebook. So, little "like" boxes will let someone authenticated on

Facebook but viewing, for example, a movie site click to identify a film they, well... like. This information is recorded on their Facebook profile. But the interesting thing here is that the social plugin is identifying items and objects within pages, and the engineers who introduced the plugins are talking about linking to these individual things. They identify the fact that when someone indicates that they like a movie, that's exactly what they're doing. They're not "liking" the page which contains the review, but the film itself. These social plugins tell Facebook that a person is expressing some kind of relationship with objects from the wider web. They talk about these people (Facebook users), and the things they're interacting with (content, objects... things), as existing like points in a graph. There are people, and there are objects, and there are relationships between these nodes. In essence, they're talking about linking data.

What happens with thousands of developers are taught something that's by definition boundless?

To make this possible, they've written the Open Graph Protocol⁴, based on RDFa. Site owners can begin marking-up their content, flagging their pages as Open Graph objects, that Facebookers can start to like. The protocol contains a vocabulary of object types, and addresses physical location and contact information. So, I can now type in a few lines of metadata in my header, and start declaring the objects in my content. It's all starting to sound very Linked Data, isn't it? The Tetherless World blog⁵ even has a post showing a mapping between the Open Graph Protocol and RDF, exposing metadata to the wider Linked Open Data cloud. The long and short of this is that anyone who wants people to be able to join their content with the ecosystem of Facebook users, can do so using a very simple semantic markup process.

So, Facebook's nearly half-a-billion users will soon start to make use of semantic links, and millions of sites will begin to mark-up their content using Linked Data. Indeed, they had a reported 50,000 sites implement social plugins within the first week!⁶ This is properly exciting, because it will dump billions of triples out there on the web, and give more developers a boost in dealing with machine-readable information. It hasn't, however, completely negated my initial feeling about Facebook and the sprinkling of the web with thousands and thousands of likes.

But...

Many people have renewed their concerns of Facebook's stance on privacy. Some of the Open Graph API-accessible fields are now defaulted to be public until a user opts-out. Marshall Kirkpatrick talks about the vulnerability of this centralisation⁷. Despite the more "open" direction in which the Open Graph points, it's very clear that all this data of users' graphs of likes and relationships will be a valuable asset, and Facebook holds the keys to this personal data. They've already begun partnership deals with Microsoft, Yelp and Pandora, so users' data will start to flow more freely between Facebook-selected organisations.

Liz Gannes over on gigaom points out that Facebook is making itself a single point of failure for the web⁸, and illustrates in another post that a Facebook outage on 23rd April also took down partner-site plugins⁹. Robert Scoble, while admiring their ambition¹⁰ also points out that the move requires a lot of trust in Facebook. We need to trust Facebook with our own personal information, and trust it to look after the information we're feeding it about our interests, relationships and tastes. It also raises questions about security: the stakes are higher if an account is hacked, or (as happened to the Scobleizer himself) disabled.

So the main impetus for using the Open Graph Protocol is to tie in with the Facebook ecosystem.

1. <http://bit.ly/c0XuGP>
2. <http://apps.facebook.com/feightlive/>
3. http://en.wikipedia.org/wiki/Graph_%28mathematics
4. <http://opengraphprotocol.org/>
5. <http://tw.rpi.edu/weblog/2010/04/28/putting-open-facebook-data-into-linked-data-cloud/>
6. <http://www.wallblog.co.uk/2010/04/29/50000-websites-adopt-facebook-like-button-in-one-week/>

Thoughts:

So the main impetus for using the Open Graph Protocol is to tie in with the Facebook ecosystem. This is not a Linked Data evangelism project, or the combined efforts of thousands of Semantic Web developers, but the logical move of a huge company to better manage its data. They'll be creating billions of nodes on a huge, social graph; and for developers the initial purpose will be to join a group, and cash in on a quick win (if you happen to have use for social networking in your app/service/site.) We'll see that little "like" button appear all over the place, even on sites which seem a bit odd "Share your next auto check-up on Facebook!" as the bandwagon sets off.

This means we'll start seeing a lot of RDFa-like semantic metadata on pages all over the web, and this increase will be almost exclusively using JUST the Open Graph. But it will mean more folks will be asking for RDFa, and more developers will begin learning it as a new string

to their bow. I wonder how long it'll take before they start asking what else they can do with all this graph data? Teaching people the value of machine-readable data (through a popular, specific application like Facebook) has the benefit of increasing developer knowledge and inquisitiveness.

Teaching people the value of machine-readable data (through a popular, specific application like Facebook) has the benefit of increasing developer knowledge and inquisitiveness.

think a lot of this RDFa will be "wasted" as it's implemented only for the purpose of joining in with the Facebook sphere, and is under-utilised. But, I think the interesting stuff will emerge as more innovators quickly find the limitations of Facebook's controlled vocabulary and data-hoarding ambition and begin to see the potential the bigger Graph brings to their repertoires. What happens when thousands of developers are taught something that is by definition boundless?

So, we're left with a question of what we'll build, and what the Linked Data community does in reaction. For my part, I think the most important message to raise is to mix your data freely. When people begin to see the existing ecosystem of Linked Data, and that it's not just Facebook's own-branded metadata, we'll start to see innovative mashing, and thousands of new services. What will you build?

This could be catalytic; allowing a rapid change in the direction of the Semantic Web. From a Linked Data perspective, though, I

- 7. http://www.readwriteweb.com/archives/facebook_centralization.php
- 8. <http://gigaom.com/2010/04/21/facebook-makes-itself-a-central-point-of-failure-for-the-web/>
- 9. <http://gigaom.com/2010/04/23/facebook-sees-major-outage-takes-out-partner-site-plugins-too/>
- 10. <http://scobleizer.com/2010/04/22/facebook-ambition/>



Liberate your data and let a thousand ideas bloom.

Talis provides a Software as a Service (SaaS) platform for creating applications that use Semantic Web technologies. By reducing the complexity of storing, indexing, searching and augmenting linked data, our platform lets your developers spend less time on infrastructure and more time building extraordinary applications. Find out how at talism.com/platform.

shared innovation™

A Linked Data Platform for Innovation

By Mark Birbeck and Dr. Michael Wilkinson



Like many organisations, the UK's National Health Service (NHS) is data rich and information

poor. However, there is a pressing need to improve access to high-quality, focused information in order to improve efficiency, effectiveness and quality in the NHS.

UK Government data is increasingly being made available via data.gov.uk and the Department of Health (DH) has contributed over 360 datasets, more than 10% of the total on the site.

However, whilst it's important to publish information, it's also crucial to be able to make use of it, and the DH's National Innovation Centre (NIC)¹ is developing a range of widgets and applications for use in the field of healthcare technology, that allow data to be viewed and manipulated in a variety of ways.

However, whilst it's important to publish information, it's also crucial to be able to make use of it

In order to put these tools on a solid foundation, the NIC has developed a modular, extensible OWL ontology that describes the process of innovation and those involved in it. This ontology has been designed in such a way that it will also provide the basis for describing innovations more broadly, not just those involving healthcare technologies.

The NIC is committed to using Semantic Web technologies as a way to significantly improve the speed and quality of decision-making in the area of health technology innovations. Its tools will be used to assist:

- decision-making clinicians, managers and commissioners in the NHS;
- healthcare technology industries;
- other government departments

At the same time the NIC is committed to contributing back to the community from which

these technologies have come, and is therefore making its platform available in the form of the Linked Data Platform².

What are we building?

In this article we look at some of the key components of the NIC's platform.

A widget platform

To help the NIC achieve its key goals of flexibility in use and agility in build, it has eschewed the traditional notion of an application. Instead it is building a platform that makes it possible to create different types of visualisation against different sets of data, and then to combine those visualisations onto a page.

The platform is essentially a widget platform, allowing users to browse, categorise, rate, and discuss widgets and datasets, and then to deploy visualisations.

By providing social networking facilities, it is easier for people to discover and share techniques and interesting insights on the data. It's also possible for the more technically-minded to create different widget types and add them to the catalogue.

Once a data/widget combination has been chosen and configured (i.e. that is, once a visualisation has been created), the platform also makes it possible to deploy the live visualisation to the web, perhaps in a blog or journal article, or to embed a snapshot of the visualisation in a business case or report.

Applications as visualisation collections

Since the data in the platform can be presented in many different ways and used by many types of user, it's almost impossible to create a single application that will appeal to everyone. So instead the focus whilst building the platform is on creating as many visualisations as possible, and then allowing those visualisations to be combined in a variety of ways.

To make it easy to combine visualisations that are based on a common piece of data, the platform makes use of the W3C's XForms standard. XForms is more commonly used as a means of marking up data entry forms, but it also provides a number of key features for use in the platform. The first is a UI framework into which it's possible to drop custom visualisations; in essence each visualisation becomes an XForms control, and the XForms

engine then becomes responsible for the control's lifecycle.

The second benefit that XForms provides is the dependency engine, which automatically keeps the state of all related controls up-to-date, based on changes in other controls. This makes it easy to link a number of visualisations together based on a common value, such as a year, a type of treatment, or a particular country.

The platform is essentially a widget platform, allowing users to browse, categorise, rate, and discuss widgets and datasets, and then to deploy visualisations.

Architecture

Triple store

The foundation of the entire architecture is a triple store. The core data about innovations, people and organisations is obviously held in this store, but so too is other information that has been imported from other sources, such as those on data.gov.uk. The store also holds categorisation information about widgets and data sources, so as to be able to help users create visualisations by matching data with appropriate widgets.

The triple store used could be one of a number of available Open Source stores, but for our work we've focused on using 4Store. There are a number of reasons for this, but the key benefits were that it's fast, it scales, and its feature set is limited to importing triples and querying them with SPARQL.

Natural Language Processing (NLP)

Sitting alongside the triple store is the NLP module, whose job it is to import information from other sources, such as medical journals, patient forums, news articles, and so on. Other organisations' websites are a key source of information for use in the platform. For example, there are over 50,000 scientists working in the NHS, and the only realistic way to keep the system up-to-date on the projects and research those people are undertaking, is to import data

1 <http://www.nic.nhs.uk/About>

2 <http://linked-data-platform.googlecode.com/>

from their individual websites and those of their departments. The NIC is involved in a number of projects aimed to adding RDFa to key sites that can then provide useful information to the platform.

SPARQL end-point

The key interface to the triple store is the SPARQL end-point. This allows queries to be run over the web, against the data in the store. The SPARQL end-point will shortly be made available to other developers, for use in their applications, meaning that other departments and organisations will have access to the same data that is driving the NIC's own applications.

Visualizations

There is now an enormous number of high quality graphing and charting libraries available, each with their different strengths. Since this landscape is constantly changing and improving, it was important for the platform to be able to leverage the best of these components, as well as to be able to take advantage of new components as they become available. For this reason effort was put into ensuring that data from the SPARQL end-point could be consumed by components from libraries such as Google's Visualization API³ and MIT's Exhibit⁴.

However, even this is not enough, since the aim of the platform is to make it as easy as possible for authors to consume and display the data, and to build applications. The next step was therefore to wrap the widget and data combination into an XForms custom control. These controls have been built on top of backplanejs⁵, an Open Source JavaScript library which provides support for XForms and RDFa. To embed a Google Visualization map now requires nothing more than authoring a simple control:

```
<xf:output bind="mapdata"
mediatype="visualization/GeoMap" />
```

The XForms foundation is important for another reason which is that its spreadsheet-like dependency engine allows a number of visualisations to be connected to the same data. For example, three visualisations might be showing different views on data for the year 2021, such as a map, a bar chart and a

pie chart. The XForms dependency engine ensures that if the data is changed to 2021, all three visualisations will be notified and updated without the programmer having to do anything.

By combining visualisations on the same page, it is possible to use these components to create applications.

UI management

The UI management layer of the platform is provided by Drupal 7. Using a Content Management System (CMS) seemed obvious, since there is no point in reinventing many of the components that are needed, such as forums, comments, ratings, user registration, and so on. However, Drupal was chosen specifically because of the well-known commitment to the Semantic Web of leading members of its development community. Drupal's own data is stored in a traditional SQL database, but there are a growing number of modules that allow RDF and SPARQL to become part of the mix.

In terms of the functionality needed for the platform, each dataset and widget has its

Drupal's own data is stored in a traditional SQL database, but there are a growing number of modules that allow RDF and SPARQL to become part of the mix.

own page, which allows it to be a focus for comments, ratings and categorisation. Further pages are then available when users create visualisations — i.e. a combination of data and widget.

The need for a linked data platform project

The architecture described here has applicability beyond Health, and even beyond innovation, and many organisations would benefit from having a ready-made platform such as this one. For this reason the NIC has initiated an Open Source project to create a set of guidelines, best practice and tools that it hopes will help anyone who wants to make their data part of the Linked Data cloud.

Everyone should benefit as more people join the project and contribute to the code and guidelines, and more organisations put their data into the Linked Data cloud.

As the work progresses we hope to see a community grow around the platform, sharing tools and visualisations that make it easy to build applications. And we hope to see an increasing number of user groups benefit from this work.

The future

The NIC Linked Data project provides a genuine opportunity to provide tools that will assist decision-makers and healthcare professionals, and to enhance the process of developing and capitalising on innovations. But building these tools on Open Source platforms and in a collaborative manner means that others can benefit from the work, as well as contribute to it.

As the work progresses we hope to see a community grow around the platform, sharing tools and visualisations that make it easy to build applications. And we hope to see an increasing number of user groups benefit from this work.

Mark Birbeck is a Semantic Web evangelist.

Dr Michael Wilkinson is the Business Development Manager for the NHS National Innovation Centre (NIC).

They will be presenting on Monitoring Healthcare Innovation: A Case Study in Using OWL, Linked Data and RDFa⁶ at SemTech 2010.

Upcoming Events:

Talis Open Day:
Linked Data and Libraries
Wednesday 21st July 2010
British Library Conference Centre,
London

Talis Open Day:
Linked Data and Health
Thursday 19th August 2010
No 76 Portland Place, London

3 <http://code.google.com/apis/charttools/index.html>

4 <http://www.simile-widgets.org/exhibit/>

5 <http://backplanejs.googlecode.com>

6 <http://semtech2010.semanticuniverse.com/sessionPop.cfm?confid=42&proposalid=2956>

Excellence at the click of a button

By Richard Wallis



A simple statement in a Prime Ministerial speech often belies the efforts that enabled it to be uttered. At the London Global Investment Conference in February 2010, Prime Minister Gordon Brown, whilst describing the way the Department for Business Innovation and Skills (BIS)¹ promotes innovation, said: "From today you will be able to identify centres of excellence at the click of a button." The button in question was to be found on a—just launched prototype—web application: the BIS Research Funding Explorer².

Roll the calendar back a month to a conversation with BIS about the benefits of using Linked Data principles to bring together data from different sources. The then recently launched, openly accessible data.gov.uk³ Linked Data stores (hosted on the Talis Platform)⁴ were already demonstrating their usefulness in creating data mashups, and the team at BIS were interested in showcasing the UK's leading role in this field. As the UK is also at the leading edge of many other technology research fields, the idea of showcasing one technology by identifying centres of excellence in others soon emerged. At the time, none of the source data for such an application had been published in any form, let alone RDF, and the expected relationships between these data was only assumed.

As the UK is also at the leading edge of many other technology research fields, the idea of showcasing one technology by identifying centres of excellence in others soon emerged.

Within a few days, a recipe for a prototype application project was formed: take research funding data from two research councils⁵ and the Technology Strategy Board⁶; combine with information about patents from the Intellectual Property Office⁷; mix using RDF and Linked Data principles in a Talis Platform store; and serve up using an animated visualisation on a web application that shows the connections between funders, projects, patents, institutions and regions. A plan hatched, and a few

commercial details settled brought us to a project kick off with only three weeks to go before a certain Prime Minister was due to take to the podium.

Being a short timescale prototype project, it was not feasible to deal with all aspects of UK research. So a representative group of four subjects was chosen: RFID, Plastic Electronics, Regenerative Medicine, and Advanced Composite Materials. As is often the case, the source data was mostly to be found in spreadsheets with differing interpretations of research topic categorisation, and the naming of organisations, locations and regions. The team at Talis launched into a period of intense activity creating a data model to contain these data, whilst recognising the wider need to become part of the growing set of data available in RDF from data.gov.uk.

Modelling the data was only the start of the exercise how it would be visualised and navigated was equally important. The project was joined by Amsterdam-based visualisation company Iconomical⁸. With their background in producing visualisations for the London Gazette

and Where Does My Money Go, they were able to hit the ground running. After analysing the data, it became apparent that in addition to the relationships between research projects and organisations, and organisations and patents,

Modelling the data was only the start of the exercise how it would be visualised and navigated was equally important.

funding details, dates and geographic locations would be a rich source to drive a visualisation.

It was concluded that the visualisation should be constructed in Flash from three main components: a timeline control to identify when between 2000 and 2018 research funding was allocated; a cumulative bar chart showing how the research investment and patents build up over time and a map showing where in the

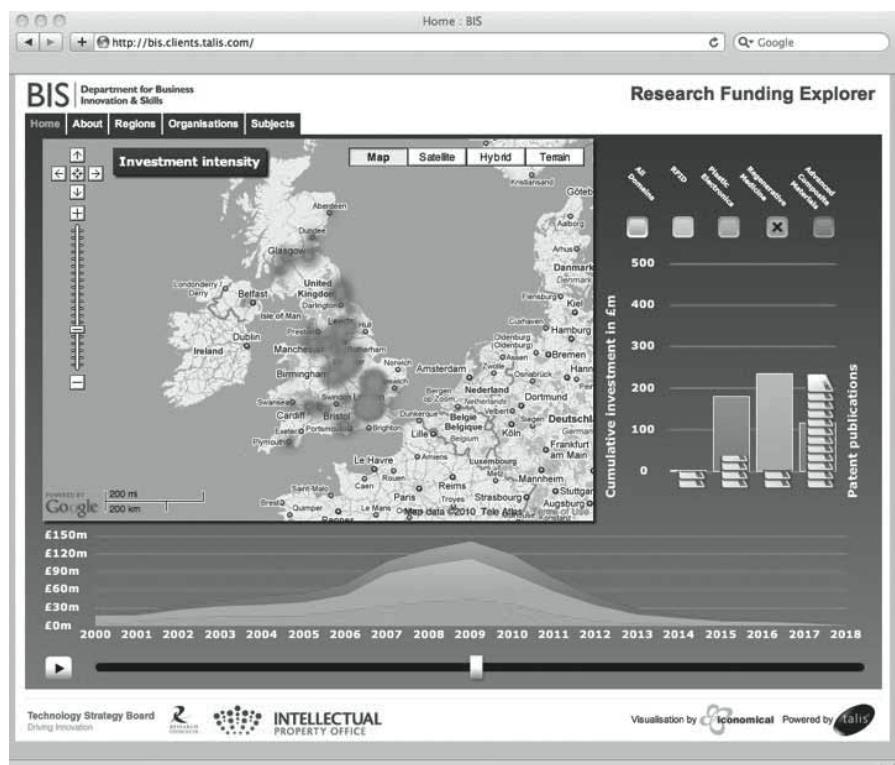


Fig. 1. The BIS Research Funding Explorer

UK research investment funding has been allocated.

Whilst Iconomical started to build the visualisation components based upon assumptions about the data, the Talis team worked on scripts and regular expression algorithms to extract the data from the source spreadsheets. This was an iterative process involving the owners of the data to tease out the categorisation of projects. What looks like a well structured informative spreadsheet for the internal use of a single research council to track research projects and funding, is not necessarily the best source for populating a shared resource. Much script wrangling, regular

into a region or smaller area and seeing representations of individual icons displayed for each organisation's location. Then, of course, the design of map pins to identifiably represent commercial and academic research organisations became an issue in its own right.

Although the animated visualisation is a core part of the BIS Explorer, it is not the only way to explore the linked data that drives it. It was soon apparent from the data modelling that there were some obvious user navigation start points. Starting with organisations—both companies and academic institutions—or individual subjects or, geographic regions, you can easily view the projects and patents in your

commented upon by the team at Iconomical:

"We joined the BIS project with two weeks to go. At the outset, it appeared an impossible task—uncertain objectives, unevaluated data from multiple sources, and a deadline so tight it didn't seem realistic. Over the next fortnight, everyone worked together to make an elegant and convincing application, and we even had time to develop some new features for our visualisation library. All in all, it was a stimulating and enjoyable collaboration with Talis, who never once lost their sense of humour."

The project was completed and launched at <http://bis.clients.talis.com>, in just over two weeks, allowing the speech writers time to work it in, a few post development tweaks to be made and an accompanying screencast¹¹ to be recorded. Since its launch, and addition to the data.gov.uk apps listings, it has been referenced many times as one of the few demonstrable Linked Data applications.

As a prototype it forms a basis for possible further development of the application itself to cover more topics, utilise OS data for better location information etc., and as an exemplar as to the way data can be successfully brought together from separate organisations to deliver a beneficial outcome.

There were a few lessons learnt from this project that are worth sharing:

- Never underestimate the differences between organisational approaches to the same problem (and tracking it in a spreadsheet).
- The links between concepts go a long way in identifying user navigation paths.
- You do not have a monopoly on the ideas for using your data.
- A Prime Ministerial event has a great influence on meeting delivery dates.
- A short, sharp successful project with a group of motivated bright people, is stimulating, satisfying and fun.

Richard Wallis is a Technology Evangelist at Talis

expression wizardry, and data model tweaking from the Talis team took place over the next couple of weeks. Often late into the evening you could drop in to the project's IRC channel to find conversations about the nuances of data extraction, interpretation, and the most suitable ontology to use—as well as a lot of humorous banter!

Representing research activity (as measured by funding input over time) on a map has its own challenges. One option was to locate a point on the map for each running project for the point in time selected. This worked well when zoomed in on a region of the UK, but at the national level, you ended up with an unhelpful forest of overlapping map pins. The alternative was to produce a heat map with coloured areas of varying density, looking a bit like clouds, overlaying the map and give a general impression of activity in a region. But this only really worked well at national level. The pragmatic solution was to do both: starting off at national level with the heat map, but zooming

area of interest. Selecting a project displays description, dates, grant values, participant organisations, and region(s). Following-your-nose style navigation enables you to explore other linkages within the data, such as projects associated with an organisation, and patents associated with a subject within a region. Although a prototype, with a few rough edges, it clearly demonstrates how natural this form of navigation through linked data feels to a user. As the underlying data is also accessible for search and SPARQL queries using the Talis Platform API⁹, it also can, and has been¹⁰, used to for mashups with other data.

Nevertheless, things were not always as easy as hoped. For instance, lack of availability of other public datasets, such as the Ordnance Survey geographic data at the time, did hold back some development. The detailed result of this being that government region is not recorded for patents—the OS data was needed to map from postcode to region.

This was a enjoyably challenging project, as

1. <http://www.bis.gov.uk>

2. <http://bis.clients.talis.com>

3. <http://data.gov.uk>

4. <http://www.talis.com/platform>

5. EPSRC: <http://www.epsrc.ac.uk> & MRC: <http://www.mrc.ac.uk>

6. <http://www.innovateuk.org>

7. <http://www.ipo.gov.uk>

8. <http://www.iconomical.com>

9. http://n2.talis.com/wiki/Platform_API

10. <http://blogs.cetis.ac.uk/wilbert/2010/02/25/linked-data-mashup-on-a-string/>

11. <http://tinyurl.com/bisdemo>

Growing Graphs: Talis working toward a Linked future

By Leigh Dodds, Programme Manager of the Talis Platform



It's been an interesting year for Linked Data, and there's every sign that momentum is continuing to build. You only have to look at Facebook's adoption of RDFa to see Semantic Web technologies beginning to enter the mainstream.

At Talis we've spent a great deal of time over the last twelve months working with a wide range of different organisations, helping them explore Linked Data and Semantic Web technologies. It doesn't take long to go beyond initial evangelism and introductory presentations before organisations begin asking more specific questions:

- How can we acquire the necessary skills to explore further?
- We're wrestling with this specific problem. Can a Linked Data approach help?
- Can a Linked Data approach let us integrate our internal datasets, and what tools or services could we build on them if we did begin exposing data as RDF?
- Can consuming Linked Data let us innovate around a specific product?
- How can we maximise usefulness of our data so that it meshes well with other sources?

The answers to these questions are different for every organisation. But gaining those answers, and a deeper appreciation for the technology, invariably involves some form of focused exploration. There's a common journey that we've seen several organisations, (and individual teams within specific organisations) follow, and which we're now actively supporting through our newly founded Semantic Web consultancy team.

Often the first stage in an organisation's journey towards adopting Linked Data is a prototype or pilot project. These typically follow the form of taking a candidate dataset or collection of datasets and exploring how to model, convert, and publish them as Linked Data. The end results are typically showcased

through a visualisation or web application that shows the data in actual use. This lets decision makers quickly see the potential of the technology to mesh together a range of different data sources. Not everyone can see the potential in the data alone; a visualisation or compelling UI helps to make that data come alive.

Often the first stage in an organisation's journey towards adopting Linked Data is a prototype or pilot project.

We've put together a number of these prototype applications over the last few months. The majority of these have involved merging in-house data assets to create a new unified view of information being collected across the organisation. The BIS Research Explorer¹, discussed elsewhere in this issue of Nodalities, is a good example of this kind of experiment. In two weeks we were able to bring together data from across the Department of Business, Innovation and Skills to create not just open

there, and value to be gained from the existing sources.

These kinds of experiments can quickly showcase the potential for a Linked Data approach, in a way that can help gain buy-in across an organisation. While Linked Data is ultimately about the power of publishing data to the web, in the short term there are a lot of internal factors to overcome. Showing the viability of the technology is one of these. But there are many other, non-technical, issues to be faced too.

Moving beyond the pilot project phase, the issue for most organisations is how to build internal capability around Semantic Web technologies. There's a lot of great material available that provides some good quality primers on Linked Data, RDF, SPARQL, etc. Jeni Tennison's series of blog posts on publishing Linked Data is a particularly good resource², and Semantic Web for the Working Ontologist is a great general reference and primer on modelling with RDF Schema and OWL.

But there's also a lot of cruft. Some of the published material is now dated; the technologies have moved on, and best practices while appearing are not yet always clearly articulated. In response to requests from a number of organisations, Talis now offers a "deep dive" training course that over the course

While we're all concerned about growing the size of the Semantic Web, it's important to realise that there's already some good quality data out there, and value to be gained from the existing sources.

data, but also an interesting visualisation. Some of these prototypes have also been about trying to show the usefulness of existing Linked Data sources in creating innovative user experiences that can enrich an existing product. While we're all concerned about growing the size of the Semantic Web, it's important to realise that there's already some good quality data out

of two days provides an introduction to all of the key technologies. These courses have been successfully delivered a number of times, and we've found that developers can very quickly get to grips with the basics. SPARQL, for example, has had some bad press recently as being overly complicated, but we've found that an example-led exploration works very well for

1. <http://bis.clients.talis.com>

2. <http://www.jenitennison.com/blog/node/135>

3. <http://workingontologist.org/>

demystifying its capabilities.

Moving beyond the foundations, we've also discovered a need for a mentoring phase to help support a project or product team in developing their skills. We've found that a good strategy is to follow up an initial training course with a separate workshop session that involves working with a team to provide useful, constructive feedback on their initial work. These kinds of open, joint workshops, involving domain experts from within an organisation are a great way to check basic assumptions and understanding, as well as a forum for exploring specific modelling issues, prototyping data conversions, and exploring ways to query and explore a populated data model.

to this here's a brief run-down of some recent
engagements.

BBC: Music data in the Talis Platform

Last year we worked with the BBC to host their Programmes and Music data in the Talis Platform⁴. This was a experiment to explore whether a SPARQL endpoint would be a useful addition to their existing APIs and methods of publishing open data. This public facing project resulted in our working with a number of BBC product teams, mainly running training sessions and workshops across a number of different areas of the organisation. We've been able to provide guidance on data modelling, share our experiences with building Linked Data

Our involvement with the data.gov.uk project has taken a similar trajectory. While we were initially providing technology⁷, we've also been offering advice and support on a range of different issues including helping to identify and document best practices around publishing Linked Data for Public Sector information. The data.gov.uk team are taking a principled and responsible approach to publishing open data, wanting to ensure that provenance and versioning issues are addressed from the start. This has presented some interesting challenges which aren't addressed in many recipe-lead approaches to publishing Linked Data.

The work we've been doing, along with Jeni Tennison (TSO) and Dave Reynolds (Epimorphics), on the Linked Data API⁸ also stems from the data.gov.uk requirements. Specifically the need to deliver not just Linked Data, or SPARQL endpoints, but a range of data access and consumption options that can fit a wide variety of different skillsets, tool chains, and integration options.

The BBC and UK Government are both early adopters of Linked Data, our work with them and others has one goal: to grow the web of data as widely and as quickly as possible. The bigger the Semantic Web grows, the more everyone benefits.

Moving beyond the foundations, we've also discovered a need for a mentoring phase to help support a project or product team to develop their skills.

Furnished with in-house skills, and provided with support in climbing the learning curves, we've seen several teams begin to successfully apply Semantic Web technologies to solve realworld problems and increase the speed of their development cycles. To add some colour

applications (some, like FanHu.bz⁵, against the BBC's own data), as well as expertise in specific areas. For example we were lucky enough to be able to help define the vocabulary for publishing the rich set of RDF data exposed by the wonderful BBC Wildlife Finder application⁶.

4. <http://blogs.talis.com/nodalities/2009/06/the-bbc-the-graph-and-linked-data-stores.php>
5. <http://fanhu.bz>
6. <http://www.bbc.co.uk/wildlifefinder/>
7. <http://www.talis.com/platform/news/>
8. <http://purl.org/linked-data/api/spec>

Talis Linked Data events

Check out the events page for upcoming opportunities to learn about Linked Data, apply Semantic Web technologies to your project and join a range of discussions. talis.com/platform/events

Upcoming Events:

Talis Open Day: Linked Data and Libraries
Wednesday 21st July 2010
British Library Conference Centre, London

Talis Open Day: Linked Data and Health
Thursday 19th August 2010
No 76 Portland Place, London



Linked Data APIs

By Keith Alexander



Converting a dataset to RDF, publishing it as Linked Data, and providing a public SPARQL endpoint, will make many things possible. But sadly, as those who have followed the UK Government Data

Developers mailing list will know, it won't make everyone happy.

For experienced Semantic Web developers, RDF is great for joining together data from different places and making it relatively easy to use new datasets. However, for many developers, RDF is a barrier to entry. They want to perform uncomplicated tasks, like doing a search and pulling a subset of values from the results. They resent being forced to learn a data model, query language, and possibly multiple serialisation formats, for the sort of thing that, with traditional RESTful APIs, can be quickly figured out from a source code example or a few lines of documentation. If you have to learn them from scratch, RDF and SPARQL are an abstraction too far for these purposes.

Even for the experienced and the willing to learn, SPARQL endpoints are often something of a black box, and discovering what kind of queries are relevant to run on a given endpoint often requires some exploration, experimentation and documentation.

So a number of dataset publishers are concluding that the best way to lower these barriers that might be stopping 3rd party developers from using their data (while still preserving all the benefits of RDF technologies) is to layer a simple RESTful API on top of their SPARQL endpoints, providing developers with shortcuts to the most useful sorts of queries, and presenting the results in a simpler format.

Realising that this might be a general concern for Linked Data sets with a user base of developers unfamiliar with Linked Data, the team behind the UK Government's Linked Data deployment (including Talis' Leigh Dodds) came up with the Linked Data API Specification (<http://code.google.com/p/linked-data-api/wiki/Specification>).

The specification defines the behaviour of a web application that uses a Linked Data API definition to create a RESTful API to the data behind a SPARQL endpoint, mapping request URIs to SPARQL queries and serving the results in a choice of formats including simplified JSON and XML serialisations, as well as turtle and RDF/XML.

The Linked Data API definition is an RDF graph (probably created as a turtle/n3 document) defining the available APIs, the

endpoints associated with them, the queries that retrieve the results for those endpoints, and what properties of the items matched by the query should be included in the results document.

For example, the API definition might define:

```
/schools/secondary.json?_page=52&_sort=size
```

to translate to a SPARQL query searching for items with an rdf:type of sch:SecondarySchool, sorted by the sch:size property. The results will be formatted as simple JSON objects like:

```
{
  "_about" : "http://example.org/schools/secondary/StCustards" ,
  "type" : "http://example.org/def/SecondarySchool" ,
  "name" : "St. Custards",
  "size": 223,
  "motto": "Tandem novum quidem et inauditum consilium capit",
  "description" : "ane skool is a bit of a shambles ..."
}
```

There are currently two implementations of the Linked Data API specification:

- a Java application available from the Google Code project (<http://code.google.com/p/linked-data-api/source/checkout>)
- a PHP application, Puelia-PHP (<http://code.google.com/p/puelia-php/>)

Those who wish to deploy a Linked Data API should configure one of the above applications by creating an API definition, and installing the application on their web server.

For dataset publishers, Linked Data API is an easy way to add a RESTful API to an RDF dataset

For dataset publishers, Linked Data API is an easy way to add a RESTful API to an RDF dataset; in fact it allows them to complement the already-RESTful item-level view they already provide when Linked Data resources are dereferenced, with RESTful collections

of resources. If <http://example.org/schools/secondary/StCustards> returns a view of a single school, it is in keeping with REST convention that <http://example.org/schools/secondary> returns a collection of schools.

For many developers, Linked Data API will be a way to query Linked Data without learning SPARQL, but even for seasoned RDFicians, the API may be a useful overview of the focus and contents of the dataset, providing a shortcut to SPARQL for the most common tasks.

From the point of view of those of us already convinced of the utility of Linked Data, it is to be hoped that ultimately more and more developers will learn RDF and SPARQL and exploit the potential of Linked Data for building applications that integrate lots of different datasets across the web. In the medium term however, Linked Data APIs can lower the barrier to entry into the web of Linked Data for third party developers, but also for dataset publishers. If Linked Data APIs prove satisfactory, more data owners may be encouraged that they can deploy Linked Data and enjoy the benefits of RDF without alienating 'regular' developers who want to consume their data.

Triplification Challenge

By Juan Sequeda



In the past several years, Linked Data has been gaining traction from the academic and media communities and now the open government community. Linked Data is offering the possibility of easy and natural data integration on the web. Organisations expose data as Linked Data on the web so others can make use of it by combining it with other data. Also, from the perspective of the enterprise, Linked Data is another solution of data integration.

The rise of Linked Data can be attributed to several aspects. Initially the Linked Open Data Project exposed important datasets on the web as Linked Data such as DBpedia. Further along, media companies such as the BBC, New York Times, and Thompson Reuters got involved. Currently, we are seeing the effort of several open government initiatives that are publishing their data as Linked Data. However, I also believe that the Triplification Challenge deserves some attention.

The main objective of the Triplification challenge is to overcome the chicken-and-egg problem of the Semantic Web.

The main objective of the Triplification challenge is to overcome the chicken-and-egg problem of the Semantic Web. How can there be a Semantic Web if there is no data? What is the incentive for exposing semantic data if there aren't any semantic applications consuming the data? The Triplification challenge is an opportunity for participants to expose existing applications and data sets as Linked Data on the web while competing for prizes. Several important datasets in the Linked Open Data cloud have been created through this challenge in the past two years. These include the Linked Movie Database, DBtune, and Linked Open Data Drug just to name a few. The Triplification Challenge has been taking place annually at the I-Semantics Conference in Graz, Vienna, since 2008.

This year we are presenting the third version of the Triplification challenge, which is exposing two new trends in linked data: applications that consume Linked Data and applications that use

open government data. The first trend is part of the open track, where we envision submission in three categories. The first category maintains the original essence of the challenge: exposing novel data sets as Linked Data and demonstrating potential benefit of use within applications. The second category receives submissions on novel generic mechanisms, approaches and methodologies that convert data to Linked Data and create links to other data sets. The new trend of consuming Linked Data is part of the third category where we expect submissions that showcase the benefit of Linked Data to end-users such as specialised search, augmentation of content, etc.

The second track is the open government track. We believe that it is very timely to present such a track due to the rise of government data being published on the web in the US, UK and further afield. Our strong requirement is that the government data be mashed up with at least one Linked Open Data Set. Any dataset produced by any government in the world qualifies if it can be related to environmental, cadastral, geographic, traffic, corporate spending or any of a huge variety of existing Linked Datasets.

The main criteria of the challenge is that the submissions need to demonstrate innovated use of the Linked Datasets and highlight the technical capabilities of the Semantic Web and show real benefit to end users. Participants will be competing for three 3000 euros cash prizes, and anyone can join in. There is more information on the site (<http://triplify.org/Challenge/2010>). 3,000 euros in prize money will be given to the most promising applications, newly published datasets and methodological approaches built upon Linked Data, but you need to get them in fast because the deadline for submissions is 14th June, 2010.

Many people ask what is the killer app of the Semantic Web and Linked Data? I usually respond by asking the following question: What was the killer app of the web? Was it the browser? Search? E-commerce? The web enabled new types of applications that people haven't even imagined. Linked Data is taking that a step further. Who knows, we may see a Linked Data killer app in this next Triplification Challenge. Stay tuned!

Juan Sequeda is a Graduate Student in the Computer Science department at the University of Texas at Austin. He has organised the Consuming Linked Data tutorial at ISWC2009 and WWW2010 and is a co-chair for the 2010 Triplification Challenge.

Nodalities Magazine SUBSCRIBE NOW



Nodalities Magazine is made available, free of charge, in print and online. If you wish to subscribe, please visit www.talis.com/nodalities or email nodalities-magazine@talismagazine.com.



Nodalities Magazine is a Talis Publication

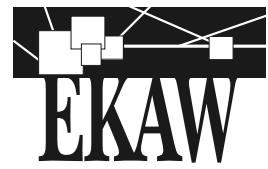
ISSN 1757-2592

Talis Group Limited
Knights Court, Solihull Parkway
Birmingham Business Park B37 7YB
United Kingdom
Telephone: +44 (0)870 400 5000
www.talis.com/platform
nodalities-magazine@talismagazine@talis.com



The views expressed in this magazine are those of the contributors for which Talis Group Limited accepts no responsibility. Readers should take appropriate professional advice before acting on any issue raised.

©2010 Talis Group Limited. Some rights reserved. Talis, the Talis logo, Nodalities and the Nodalities logo where shown are either registered trademarks or trademarks of Talis Group Limited or its licensors in the United Kingdom and/or other countries. Other companies and products mentioned may be the trademarks of their respective owners.



EKAW 2010

Lisbon, Portugal

October 11th–15th, 2010

17th International Conference on Knowledge Engineering and Knowledge Management **Knowledge Engineering and Knowledge Management by the Masses**

Conference Topics

- Knowledge Management
- Knowledge Engineering and Acquisition
- Knowledge in Use
- Social and Cognitive Aspects of Knowledge Engineering
- **Knowledge Management and
Engineering by the Masses** (special focus)

For more information and in order to register please consult
the website: <http://www.ontol.inesc-id.pt/ekaw2010>.

Register
now!!

Organizing Committee

- **General and PC chairs:**
Sofia Pinto (INESC-ID, IST, Lisbon)
Philipp Cimiano (CITEC, Universität Bielefeld)
- **Tutorial chair:**
Victoria Uren (University of Sheffield)
- **Workshop chair:**
Siegfried Handschuh (DERI Galway)
- **Demonstration and Poster chairs:**
Oscar Corcho (UPM, Madrid)
Johanna Völker (Universität Mannheim)

Confirmed Invited Speakers

- **Bernardo A. Huberman**
Senior Fellow and Director, Social Computing Lab
HP Labs
Palo Alto, USA
- **Tom Scott**
Executive Product Manager
BBC
United Kingdom
- **Enrico Motta**
Professor of Knowledge Technologies
Knowledge Media Institute
The Open University
Milton Keynes, United Kingdom

Conference Location

EKAW 2010 will take place at the **Gulbenkian Foundation** in Lisbon. The compound of Fundação Calouste Gulbenkian is a Portuguese private institution of public utility whose statutory aims are in the fields of arts, charity, education, and science. For more details you can have a look at <http://www.gulbenkian.pt>.