Nodalities

THE MAGAZINE OF THE SEMANTIC WEB

talis®

www.talis.com/nodalities

Issue 12

INSIDE

Read Previous Issues of Nodalities Magazine Online

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

twitter

Follow Us On Twitter
@nodalities

Open Government Data for fighting fires: when no traffic rules apply

By Bart Van Leeuwen

It is in the middle of the night, you are awake for 2 minutes, you and your colleagues are racing in a firetruck through the narrow streets of Amsterdam, with the 120dB horn screaming overhead. In this small and hectic environment it is your job to guide the driver to the location of the incident, your only help are old sun-bleached papers with driving directions typed out in a 10 point font. Since you are new on the job, there is no room for error, and the old guys with 20+ years of experience will not be too happy if you don't arrive on the scene on time.

The whole system of pre-written cards survived this long because all fire fighters did have at least some knowledge of the city layout. The fire department of Amsterdam traditionally consisted mainly of people born and raised in Amsterdam itself. But over the years that changed, so rookies turned up

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@talis.com.

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 (within the UK): 0870 400 5000
Telephone (outside the UK):+44 121 717 3500
www.talis.com/platform
nodalities-magazine@talis.com

Editor's Notes

This edition rounds out a dozen issues of Nodalities Magazine. We've covered stories from the particularly technical to the idealistically just-about possible. If this issue has a theme, I'd suggest it's probably practical. The world around the Semantic Web has changed, with technological advances in mobile technology, cloud computing and the rise and rise of Big Data. A quick scan of the catalog of past Nodalities shows that the Semantic World itself has been far from static itself.

In this issue, some rather practical things that Linked Data is good at solving are being put to use saving lives: quite literally, as Bart van Leeuwen explains in our cover story. Simple ideas joining up public data and GIS devices are helping the Amsterdam fire service get their equipment to the scenes of fires more quickly and safely.

It's also been a real pleasure to follow the story of the Guardian newspaper as they work with huge datasets, and explore how the applications of some semantics can help run a huge national daily newspaper. Martin Belam, an information architect there, tells us about their approach to Linked Data, and what it means at the Guardian.

The world of Linked Data faces challenges, and will doubtless evolve when practicality is stretched by difficulties. Talis' Leigh Dodds outlines some of these challenges and opportunities in his article here. Also supporting Linked Data research is the multi-organisational LATC Project which is introduced here. Elsewhere, Tim Hodson discusses a very practical approach to starting with Linked Data, and may also discuss eating an elephant.

Linked Data is also finding its place in large-scale retail, and we welcome Jay Myers from BestBuy here to talk about how publishing fairly straightforward metadata is helping with difficult sales items.

As always, if you have a Semantic Web story, are working on something interesting to do with data, or would like to find out more about Nodalities Magazine, drop me a line at zb@talis.com, or follow @nodalities on twitter.

-Zach Beauvais

Continued from front page.

By Bart Van Leeuwen, a firefighter in Amstardam and IT veteran.

in the centre of Amsterdam - literally for the first time - when they joined the fire department. This made it even harder for them to get a clue about the written directions, let alone getting a sense of direction.

With the rise of companies like TomTom, satellite navigation systems finally became affordable to a point that it would be viable to install them in all fire trucks. In the end, a PC-based solution was chosen because it allowed modifications to the route calculation. Remember, the fire trucks do not have to comply with any traffic regulations on their way to a serious incident, so standard systems would not have worked! But still, fire fighters who had great knowledge of Amsterdam's streets were able to beat any technical solution.

So why is that?

- Urban areas hinder real time GPS location.
- Routes are not always optimal.
- Navigation systems are largely designed for small cars, not 15-ton fire trucks.
- There are no alternative routes: if at some point you can not continue on the plotted route, you are lost.
- The instructions are hard to hear and see in the hectic and noisy environment of the fire truck.

This is exactly where the experience and good knowledge about the service area pays off. Satellite navigation systems are great at navigating you through an unknown city on the other side of the globe; but in your home town you will probably beat them because you know your way around.

I found that there are some digital receivers on the market which you can attach to your serial port and read out the messages real-time. That sounded very useful!

Thinking about something new

After several long, late-night discussions with some of the drivers at the fire station, we basically came to the conclusion that this satellite navigation system is more of an assistant than a real guide for us. One of the interesting features of the old card system was the 'well known points' principle: this gave the driver a good hint in which direction he has to go initially. By knowing 'Dam Square', he would immediately have an idea of the direction of the route the moment he hears it. We suddenly came to the conclusion that even just showing the address of where we need to go to on a map would be a great help.

This idea is not completely new. The Netherlands uses a system called P2000 which is a digital broadcasting network based on GSM technology. Pagers of volunteers and alarm systems in fire stations are organised in groups. Text messages are broadcasted to the groups to send out an alarm. Websites like http://monitor.livep2000.nl/ have been publishing this text information for quite some time, and they show the incident address in Google Maps. However, there are some serious limitations with this information source:

- Showing the map needs human interaction - we have no time for that.
- It is unstructured.
- It is slow: it can take up to 10 minutes for incidents to appear.
- We have no control over it, the website might be gone tomorrow.
- We cannot show extra information in the projection that might be of use for us.

So we noted down some very basic needs for a system that would really help us:

- Overview: we would like to see both the station and the incident location in one single view.
- No interaction: we do not want to interact with the system, it should just be a terminal showing the map.
- Reasonably fast: a new incident should show up in less than 15 seconds.

After doing some tests on a laptop, we figured out that the first two points are fairly easy to accomplish. But the last one was not easy to achieve: getting this data real-time seemed to be a real challenge since the websites deliberately put a delay in publication of the messages. So we turned to our own organisation and explained our ideas and why we need the messages real-time. The answer we received was simple and sobering:

"You guys don't need that!"

Welcome to the bureaucratic world! I had to put the idea in a drawer for a year or so until a new station chief appeared. In the meantime, I found that there are some digital receivers on the market which you can attach to your serial port and read out the messages real-time. That sounded very useful! My new station chief was willing to sponsor me £90 for one of those units. We were finally able to do some real testing!

The first prototype

Besides being a professional fire fighter, I have also owned an IT company for 15 years. One of the lessons learned in this time is that over-engineering and not conducting field tests are almost always fatal for the success of a project. With that in mind we started out very simply: we put an old PC on a desk in the garage. This PC had the receiver attached to it and special software installed allowing it to trigger an action when a certain group messages was received. We used the trigger to fire up a web browser which opened a page and showed the incident location and the fire station on a Google Maps screen. The address was extracted from the broadcast message by a decoder we wrote ourselves, and we were then able to embed it in the web page.

Within two weeks of usage, it was already clear that we never wanted to go to an incident without this system! After about two months of tweaking the message decoder, we reached a 99.9% accuracy of the exact location of the incidents on the map. We could even beat other fire stations in their own service area! So we proved our concept, now it was time for a proper implementation.

Enriching the data

After a few months, we got used to the system and new ideas and demands popped up automatically. First of all, other fire stations were really interested in our solution and wanted a screen in their own stations. Secondly, my own team came up with ideas on how to enhance the system; such as putting information about road construction on the maps. Although we already had this information, it was distributed on a 20-page fax which was pinned to the information pinboard as a reference for the drivers. It was assumed that the drivers would go through these faxes every two weeks when the new version arrived. I really did not feel like handcrafting all this information into a digital format used by our map, so I decided to figure out where this information actually comes from. After 20-30 phone calls in a week and continuously insisting

This PC had the receiver attached to it and special software installed allowing it to trigger an action when a certain group messages was received. We used the trigger to fire up a web browser which opened a page and showed the incident location and the fire station on a Google Maps screen.

on talking to someone with knowledge about the information source, I finally got introduced to 'Peter'. He is responsible for maintaining all the data about road construction work in the city of Amsterdam. All this information is plotted into a GIS-enabled computer system which can provide exports in a wealth of formats! We had some nice discussions and in the end we decided to go for the KML file format, as this would be an easy way to start testing.

The data is currently neither linked nor public but at least it is an open standard. We are now working together on opening this data up to the public in a more generic way so it can be used by the fire departments and anyone else. There is no final decision about what form the data will be eventually published in, and we might need some help on that; but at least the drive is there. It is not easy to explain the need and benefit for the organisation itself at management level and why they should open up information like this. But it is certain that the fire departments will get what we need and that is already quite a good start!

Towards Linked Data

While demanding the opening up of data from other organisations why not open up the data we acquired with the message decoder ourselves? The data is already available in the web in an unstructured form and all we do is enriching it by decoding the message itself.

The logical conclusion was to start generating RDF out of those messages. This would also directly solve some of the problems coming up with the distribution of the data to other fire stations or additional devices. During this process we also figured out that there is a wealth of information available inside our own organisation and this information could assist operational services of our fire department. Just generating RDF out of some Excel sheets already allowed us to link very basic information to incidents, like information about dispatched fire trucks etc. And this is just the beginning.

With Linked Open Data the essence is that you do not have to move the complete organisation at once. With a good, tiny idea implemented well, there will be no discussions about the actual need or the benefits of open data anymore.

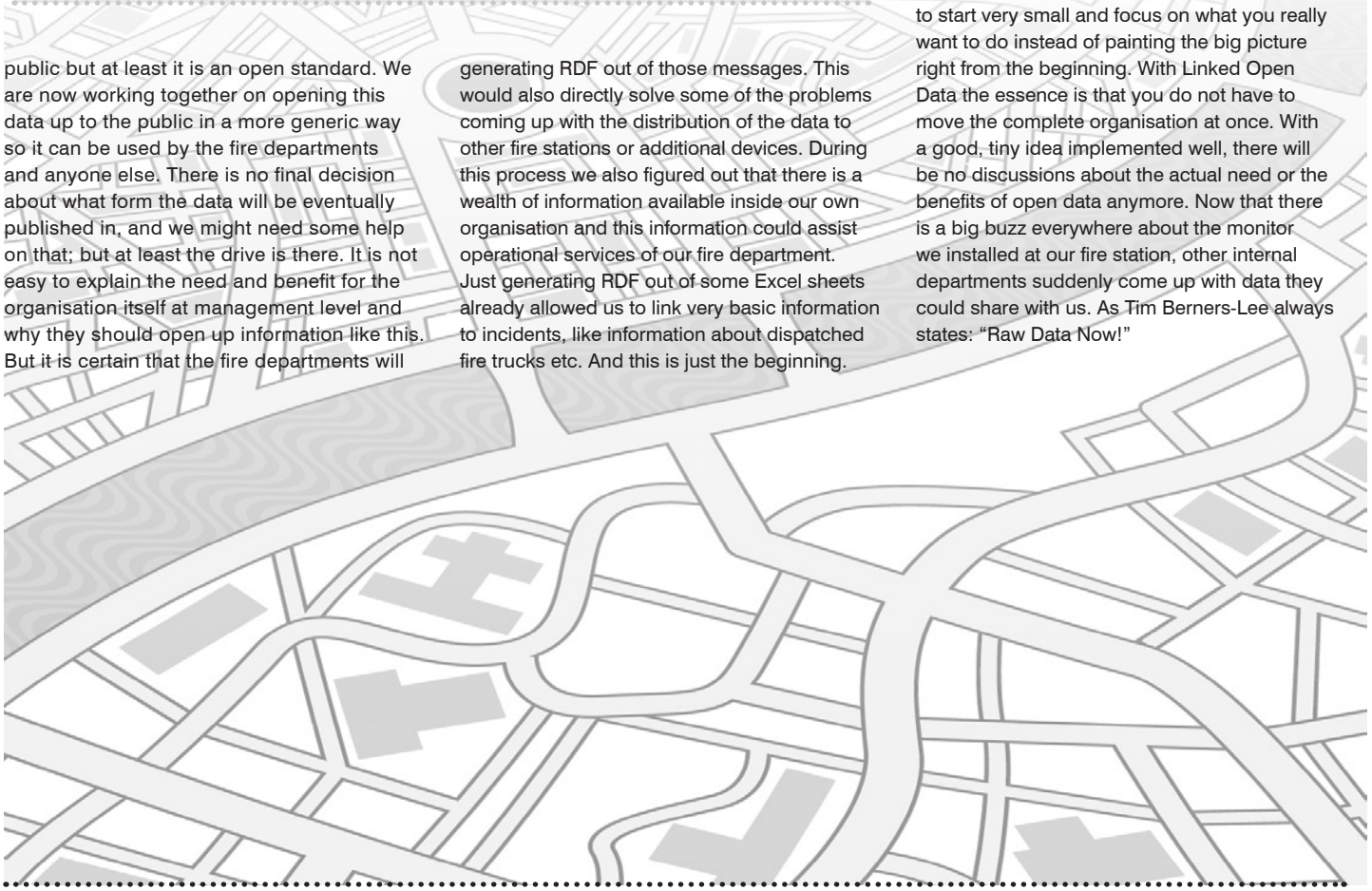
Again it was time to review what we did so far and build the base for the next iteration of our system. In a cooperation with netage.nl and netlabs.org, we started to develop an infrastructure on top of RDF which facilitates selecting and displaying information of interest in different situations.

Conclusion

Although the system is not developed up to its full potential, we already came to interesting and unexpected conclusions:

- Because the drivers know where they need to go from A to Z, they are calmer and more focussed on the traffic. This increases safety during driving a lot!
- The rookies start to get a better view of our service area and they improve their knowledge of the general city layout.
- By decoding the information from the messages, it is easier to determine what is exactly going on and where.

The biggest lesson learned is that in a slow-moving bureaucratic environment, it is better to start very small and focus on what you really want to do instead of painting the big picture right from the beginning. With Linked Open Data the essence is that you do not have to move the complete organisation at once. With a good, tiny idea implemented well, there will be no discussions about the actual need or the benefits of open data anymore. Now that there is a big buzz everywhere about the monitor we installed at our fire station, other internal departments suddenly come up with data they could share with us. As Tim Berners-Lee always states: "Raw Data Now!"



Introducing David Wood

By David Wood, Vice President of Engineering at Talis Inc.

Talis' new subsidiary, Talis Inc., was announced just one busy month ago. As Talis' new chief geek in North America it is high time for me to introduce myself to the Nodalities community.

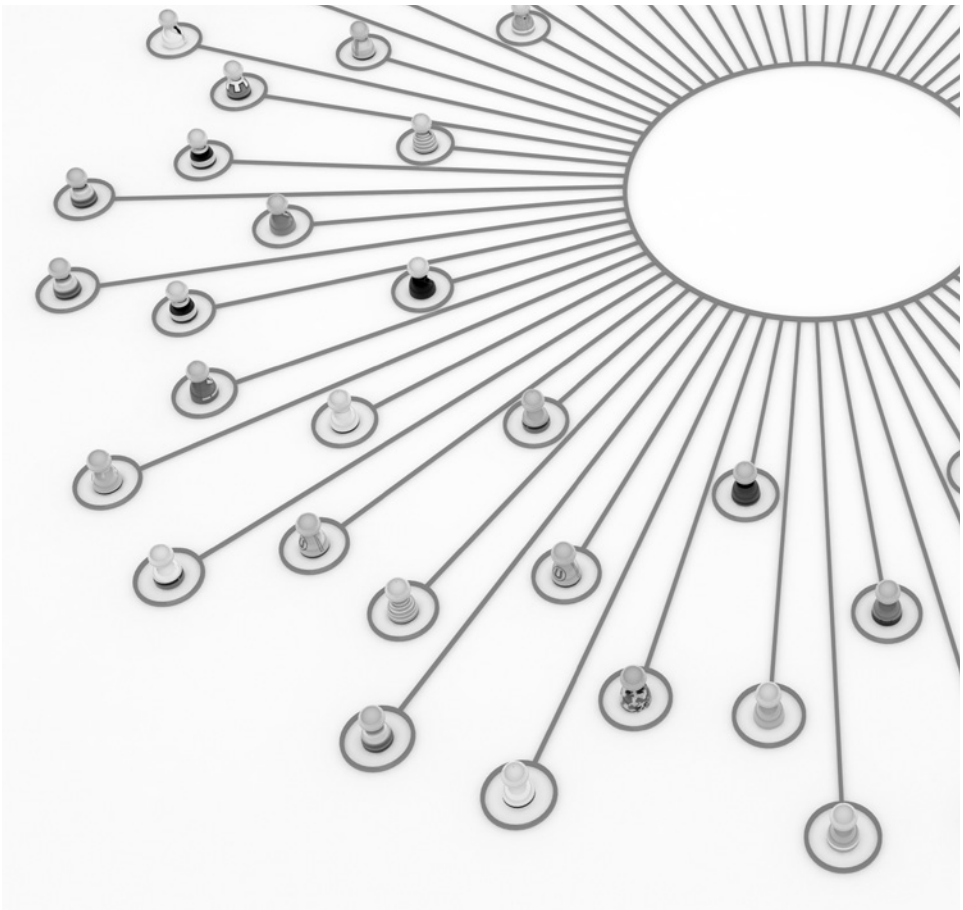
I am a software engineer with a long history working on the Web, developing Semantic Web and Linked Data infrastructure and building Web standards. At Talis, I will be helping to define our market entry into the North American market and contributing to the technical direction of the Talis offerings.

Talis is a wonderful company to work for. The environment is extremely collegial and pleasant, although also very productive. I look forward to contributing to both the company and its communities, and to bringing some of the Talis culture into North America.

Talis Inc's CEO, Bernadette Hyland, and I have started several companies together including Semantic Web startup Tucana Inc. (sold to Northrop Grumman Corporation in 2005) and, more recently, the SemWeb consultancy Zepheira. Much of our previous work has been released as Open Source Software, including the Mulgara Semantic Store, a popular Semantic Web database, the Persistent URL software and, most recently, a project aimed at making Semantic Web and Linked Data applications easier to create, called Callimachus.

The success of the Web is based on standards of the World Wide Web Consortium (W3C), so I have tried to help them whenever possible. I co-chaired the Semantic Web Best Practices and Deployment Working Group and, more recently, the RDF Next Steps workshop. In 2011, I'll be co-chairing a new working group aimed at updating the Resource Description Framework (RDF), the technical standard underlying the Semantic Web and Linked Data.

The growth of Linked Data has lead to some truly interesting applications. I've been working with many others to collect some of those use cases into book form. The goal is to help others replicate those early successes. The first book, Linking Enterprise Data has just been published and is available freely on the Web. It may also be purchased in ebook and printed form. A second book is to be entitled Linking Government Data. We are currently seeking contributions, so please contact me if you have a good story to tell about the use of Linked Data in government settings.



The growth of Linked Data has lead to some truly interesting applications. The goal is to help others replicate those early successes.

I occasionally teach computer science and mathematics courses at the University of Mary Washington in Fredericksburg, Virginia. Most recently I've taught Computer Networking and Introduction to Discrete Mathematics. It looks like I'll be teaching an upper division elective on Linked Data during the summer of 2011. I thoroughly enjoy working with university students. Theirs is a fascinating time of life, when they choose how they see themselves as individuals and what they will (at least initially) do for a living. They also help me keep a fresh perspective on our rapidly changing world. Seeing the Web through their eyes is really very different than seeing it through mine.

We, as a community, built the Web. We continue to build our community as we build the Web. I look forward to being on the journey with you.

"Linked Data" at the Guardian

By Martin Belam, an information architect at the Guardian

During October at Guardian News & Media we announced a change in our Open Platform Content API. For the first time, developers and users could query our database of over 1 million content items by using the common external identifiers of a MusicBrainz ID or an ISBN number. It is our first step into the world of 'Linked Data'.

The Open Platform Content API was launched as a beta in 2009, and earlier this year was launched as a commercial product [http://www.guardian.co.uk/open-platform/blog/open-for-business], allowing partners to re-use Guardian & Observer content in a variety of different ways. There is, for example, a Wordpress plugin that easily allows you to include Guardian content in your blog [http://www.guardian.co.uk/open-platform/news-feed-wordpress-plugin], and developers have built applications like a bespoke recipe search on top of the data [http://www.guardian.co.uk/open-platform/blog/recipe-search-microapp]. It is a unique proposition amongst news organisations on the web, and as well as the Content API itself, the Open Platform also includes publishing the source data behind Guardian journalism on the Data Store [http://www.guardian.co.uk/data-store], and providing a search engine for Government datasets from around the world. [http://www.guardian.co.uk/world-government-data]

For the first time, developers and users could query our database of over 1 million content items by using the common external identifiers of a MusicBrainz ID or an ISBN number.

Why linked data at Guardian News & Media?

The addition of linked data to the API is the culmination of a great deal of work behind the scenes to get the data prepared, and to work out the right way to make it available. Personally, I had been struck the first time I saw the linked open data cloud diagram that none of the bubbles represented any of the UK's traditional print news organisations. With our combined centuries of experience sifting, collating, organising and publishing information, it seemed to me that they should in fact be occupying a central position on that map. The principles of linked open data also chime with the over-riding principles we have about our web presence at Guardian.co.uk. We strive to be 'of the web', not just on the web. That means reaching out and embracing external services and data, and our intention is to have permanent, predictable URLs for all of our content.

The first challenge to implementing this was to pick stable and reliable external datasets that would form a permanent and meaningful relationship with our content. We decided that a focus on distinct cultural entities would work, and avoided the messiness of trying to decide whether a story was 'about' something, or whether it just 'mentioned' something. MusicBrainz IDs and ISBN numbers seemed like datatypes we could work with.

The domain model of our content already had a concept of an 'external reference' that can be added to a tag or a factbox or an article. We have previously used that to link articles to a page about a specific film, or to link a sports match report to game statistics provided by a third party like Opta. The obvious route was therefore to expose these 'external references' in our API.

MusicBrainz IDs

With MusicBrainz IDs, we did not attempt to tag all of our music story archive. There are around 42,000 music content items currently on our site, and to accurately add MusicBrainz IDs to them would be an arduous task. Fortunately, because of our domain model, we had a shortcut to tagging this content. All of the items in our database are given tags. These indicate the type of content (e.g. article, audio, video), the tone of content (e.g. news, comment, review, obituary), the contributor who produced the content, and keywords representing the subject the content is about. In the Music section, we have around 600 of the artists we write about most frequently who exist as keyword tags. The quickest route to adding MusicBrainz data was to add it to these artist keyword tags. The actual job of tagging

Fig. 1. MusicBrainz IDs API

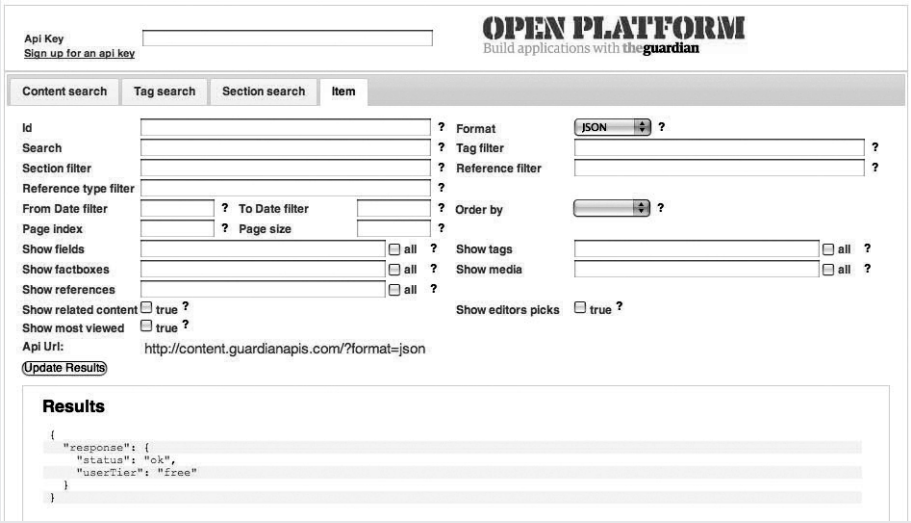
was achieved via the rather dull mechanism of filling in a Google Docs spreadsheet, although developer Daithi ó Crualaoich built a tool to help us. He came up with a quick browser-based hack that simultaneously put the same search string across our music tags and across MusicBrainz, and matched the outcome. A script then uploaded this to our database.

ISBN numbers

ISBN numbers were another obvious choice for us. The majority of our book reviews on the web feature a 'fact box', giving details of the publication and a corresponding link through to our book store to make a purchase. This 'fact box' frequently includes the ISBN number of the publication, and so exposing them as a search criteria was not a massive undertaking. Nevertheless, as with our music content, we do not have universal coverage. At the time of launch around 2,500 reviews out of a possible total of 17,000 had ISBNs attached to them. This is part of the production process now, and so all reviews going forward should have the ISBN added.

API query types

The Open Platform supports a range of ways to query this data, and you can find a guide at http://gu.com/p/2k6ay. Obviously you can query the API looking for a specific reference, so a query for reference=musicbrainz/05ec70a5-3858-4346-a649-fda0a297b8c1 will return content about Shirley Bassey. Additionally, you can get a list of content which has a MusicBrainz or ISBN attached to it, so reference-type=musicbrainz|isbn will give you content from the API which has a MusicBrainz OR an ISBN added to it. Adding the 'show-references' parameter will return a block in your API responses that includes MusicBrainz IDs or ISBN numbers for any item within the list.



If you've not used the Guardian's API before, you can get a feel for how it works by using our browser based API explorer at http://explorer.content.guardianapis.com/.

'Linked data' formats

It does seem that as soon as you put the words 'linked', 'open' and 'data' into the same sentence, you automatically invoke a debate about what formats are appropriate to use. At the present time we are making these persistent external IDs available alongside our content items in both XML and JSON formats. And yes, that does mean that we have steered away from RDFa and SPARQL.

From our point of view there is a clear reasoning behind this. We try to work in a lightweight and agile way, and providing the data in this format was the simplest way to meet our immediate requirements. We are trying to concentrate on making more metadata available. If we were to decide to invest in triple-stores and implement a SPARQL endpoint first, then I'd wager that we would still be waiting to dip our toe into the water.

Moreover, it would be wrong to commit our editorial production colleagues to tagging up all our content with this extra layer of semantic data, if we can't show the benefits. It is my hope that by incrementally releasing extra layers of linked data through our API, in a simple way, we can see what works and what doesn't, and what types of data interest people and inspire them to develop applications using the data

As I've personally argued before, particularly in response to Tom Coates' recent call for "Death to the Seamntic Web", I'm entirely agnostic about formats myself. What I think is most important is that we provide consistent, RESTful, predictable, persistent hooks into Guardian.co.uk content, in as many ways as possible, with the right licence for re-use. [http://www.currybet.net/cbet_blog/2010/09/dconstruct-tom-coates-semantic-web-must-die.php]

What I think is most important is that we provide consistent, RESTful, predictable, persistent hooks into Guardian.co.uk content, in as many ways as possible, with the right licence for re-use.

What next?

We are now evaluating where else we can add value to our API with joins to external datasets. Again we will aim to be pragmatic - tagging the most amount of data with the least amount of effort. And we also want to listen to the linked data community - what are the data joins that would be most useful to external developers.

Linked Data Around the Clock

By Lin Clark and Michael Hausenblas, DERI Galway

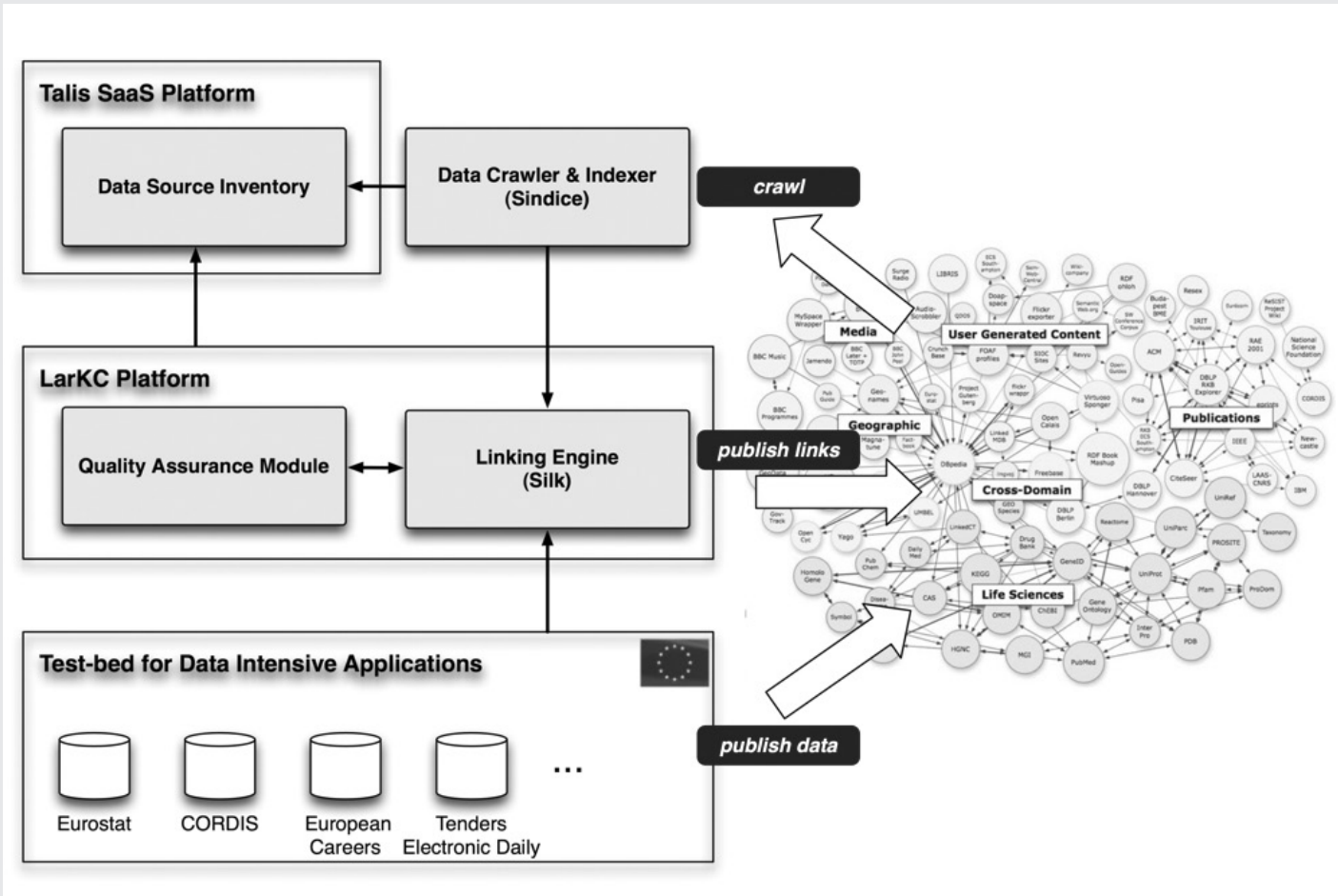
In this, the Petabyte Age, technologists have a growing obsession with data & big data. But data isn't just the province of trained specialists anymore. Data is changing the way scientists research and the way that journalists investigate; the way government officials report their progress and the way citizens participate in their own governance. The challenge that all of these accidental technologists face is how to surface data and bring data together in meaningful ways. As Google's chief economist Hal Varian has said, the scarce factor is no longer the data, which is essentially free and ubiquitous, but now the scarce factor is the ability to understand that data and extract value from it. The emerging Web of Linked Data is the largest source of this data multi-domain, real-world and real-time data that currently exists. As data integration and information quality assessment increasingly depends on the availability of large amounts of real-world data, these new technologists are going to need to find ways to connect to the Linked Open Data (LOD) cloud. With the explosive growth of the LOD cloud, which has doubled in size every 10 months since 2007, utilising this global data space in a real-world setup has proved challenging; the amount and quality of the links between LOD sources remains sparse and there is not a well-documented and cohesive set of tools that enables individuals and organisations to easily produce and consume Linked Data. A new project aims to change this, making it easier to connect to the LOD cloud by offering support to data owners, Web developers who build applications with Linked Data, and small and medium enterprises that want to benefit from the lightweight data integration possibilities of Linked Data.

...there is not a well-documented and cohesive set of tools that enables individuals and organisations to easily produce and consume Linked Data. A new project aims to change this, making it easier to connect to the LOD cloud by offering support to data owners, Web developers who build applications with Linked Data, and small and medium enterprises that want to benefit from the lightweight data integration possibilities of Linked Data.

LATC to the Rescue

The new LOD Around-the-Clock (LATC) project kicked off on September 13-14, 2010 at the Digital Enterprise Research Institute in Galway, Ireland. LATC brings together a team of Linked Data researchers and practitioners from DERI (National University of Ireland Galway), Vrije Universiteit Amsterdam, Freie Universitat Berlin, Institut for Angewandte Informatik, and Talis. This team will support the production and consumption of Linked Data by providing:

A recommended tools library for publishing and consuming Linked Data, supplementing



documentation for the tools, and free implementation support for large-scale data publishers and consumers. Tools include the D2R Server for publishing relational databases on the Semantic Web, the Drupal CMS and related publishing and consupmtion tools, and others.

A 24/7 interlinking platform (see Fig. 1) that acquires new data and creates links between existing datasets in the LOD cloud.

Publication of new large-scale LOD datasets with data from governmental departments and other organizations. The focus will be on EU level datasets such as CORDIS, the European Patent Office, and Eurostat.

In addition to the core team, a large Advisory Committee with more than 30 members will participate in the LATC activities and connect the Linked Data community to LATC's recommended tools library and support services. Organizations on the Advisory Committee are entitled to support from the project and thus will be in a position to give feedback to improve the support services.

The Advisory Committee includes governmental organisations such as the UK Office of Public Sector Information and the European Environment Agency; researchers and practitioners such as the University of Manchester, University of Economics Prague, Vulcan Inc., CTIC Technological Center, the Open Knowledge Foundation; and standardisation bodies, including W3C (Tim Berners-Lee). The LATC partners will also liaison with other EC projects and related activities, including LOD2, PlanetData, SEALS, datalift.org, Semic.EU, OKFN, and the Pedantic Web group. LATC organises and supports a number of community events, including tutorials at the International Semantic Web Conference 2010 in Shanghai, China, as well as the Open Government Data Camp, London. LATC is a Support Action funded under the European Commission FP7 ICT Work Programme, within the Intelligent Information Management objective (ICT-2009.4.3).

Linked Data: Say What You See

By Tim Hodson, Consultant at Talis

I'm amazed at how little technical knowledge you actually need to be able to describe something in a form that can be used as Linked Data. Really?

Well, yes.

The oft used: "Q: How do you eat an elephant? A: one spoonful at a time" aptly describes both the enormity of the task and the simplicity of working out how to actually begin describing some 'thing'.

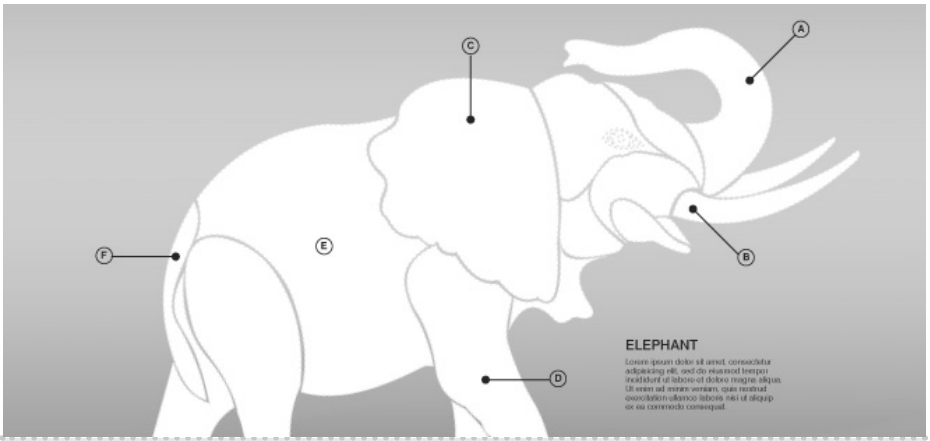
- This thing is a type of mammal.
- This thing is grey.
- This thing has four legs.
- This thing has a prehensile proboscis
- This thing is called 'elephant' in english

Very simple statements that allow us to start describing an elephant. You can apply the same principles as expounded by Roy Walker on Catchphrase, "Say what you see". (The implication also being the reverse in that you should not say what you don't see.) This is the basis of the open-world principle: the idea that the things that you don't know are just things that you haven't found out yet. Whereas the closed world principle asserts that anything I don't know must be false.

In the first of our three statements we are simply starting with a general assertion that the thing we are talking about is a type of mammal. We then go on to describe some other properties of the thing. Each statement is a standalone spoonful; an assertion that something is true.

You see how easy it is to describe something? The next stage is to codify that description in a standard, machine readable way. This is where it does get more complicated, and we start to talk about using the Resource Description Framework (RDF) as the standard machine-readable form for our statements.

In the world of RDF there is an important maxim that eloquently sums up what you can do with RDF. RDF allows anyone (including you) to describe anything about anything. Invariably this will throw up issues of truth and trust. But because RDF is being used as part of the existing web of documents those issues of truth and trust are not that different to those found in traditional web publishing.



Linked Data is not just about mapping the data from one format to another. It is about remembering that the data you have is just a way to say a whole bunch of stuff about some thing. Your data describes the attributes of real things, and it is those real things that are linked by context or association in the 'real world'.

Have you decided to take the plunge and start describing your elephant one sharp spoonful at a time? One of the first questions we get asked is "What ontology should I use?". Our answer is that there is no single ontology that fits any given elephant.

If three blind men were to be taken to an elephant, the first man might describe an animal with a long nose; the second might describe an animal with big ears, and the third might describe an animal with thick legs. All three would probably agree that the animal had rough skin and was taller than they were.

What are we demonstrating here? In RDF, things (like elephants) are described using properties. If the properties used can be shared in descriptions of similar things by different people, then there is a common understanding and a link between those things. Where one blind man uses words to describe characteristics that the other blind men also use, there is a common understanding.

Lets have a real example, there is a property known as foaf:name which is defined in the Friend of a Friend vocabulary as being the 'name of a thing'. Because this property is defined clearly, and without constraint, you can use it to describe any thing. Many people use foaf:name when they are describing their own things. Schemapedia allows you to search for properties that have already been defined by other people, so that you can re-use them in your model.

But our best advice is to get away from the computer and talk about the modelling of your data with other people who are interested in the same things. You can share your common understanding, share your expertise in the domain and let other people who are familiar with the nuances of RDF help you to refine your model.

Linked Data is not just about mapping the data from one format to another. It is about remembering that the data you have is just a way to say a whole bunch of stuff about some thing. Your data describes the attributes of real things, and it is those real things that are linked by context or association in the 'real world'.

In many places within this text, if you substitute the word elephant with dataset, you will suddenly find that you can describe your data as easily as you can describe elephants. Except that datasets don't have long noses, just long tails.

Talis work closely with people like the BBC, Data.gov.uk and The Open University, who are interested in seeing their data reach new audiences by being transformed into Linked Data. Talis offer rent-a-brain consultancy to help you make the transition to using Linked Data technologies.

Sponsorship of SSSW11

By Sarah Foster

Talis are delighted to be one of the sponsors of the 8th European Summer School on Ontological Engineering and the Semantic Web (SSSW 2011). http://sssw09.org/2011/index.php There will be more about this in coming issues but just to start this off...

We are sponsoring it for a very simple reason. The mix of theoretical, practical and collaboration skills used by all the students involved from across Europe directly corresponds to how we work at Talis. It's an environment of support and challenge, contribution and connection that has proved beneficial for all involved over the years and Talis is proud to be able contribute and participate to further the aims of the community at large and hopefully encourage continued connection too.

Talis is a small and ambitious company of likeminded motivated people. A phrase we often use here is Human Scale. Culturally what we mean by that is we like working closely with people who we all know, whether as employees of Talis or more likely over time collaborating as partners in joint endeavours. We want to grow our company and contribute to the communities we belong to, we know that it is by enabling and fostering relationships with others driven by the same passion to collaborate, learn and share that we can build on the ambitions we have for ourselves and for the communities we belong too. One particular aspect of the summer school is this same notion of social connectedness, a personal network of trusted relationships that challenge and enhance the experience for everyone.

The Plings Project

By Zach Beauvais, Talis' Platform Evangelist

It's always good to work on projects that aim to make a difference and to contribute something: you could say we look for projects with some substance to them. So, it's been fun to work with social research company, Substance on their Plings project. If you'll forgive the opening pun, I'll explain a bit.

The Plings project aims to gather together the best available information about "positive activities" for young people: PLaces to go + thINGS to do = PLINGS. Substance describes Plings as "a search and discovery tool that helps people to find accurate and trusted sources of information about positive activities for teenagers". So, I can look for Plings around Talis' offices, and find out about football coaching, cafes, dance and musical projects: all happening within a set radius of my postcode. It's a versatile tool, letting the searcher facet their results and customise the display, and it also ties in with social networks (check out the fantastically-named "boredometer" for example), and devices.

Feeding the Plings site is a dataset comprising two main parts:

Data on the actual activities: places to go, things to do

Data on feedback relating to the activities: "Plingback"

Substance uses various methods to collect the first dataset, routing it through their own API. This lets them use data from many different formats and shapes: from local authorities, third sector and community groups and the private sector. For Plingbacks, though, Talis has been working with Substance to create an infrastructure that can be used to generate data in RDF which Talis hosts through it's Managed Service. There is a bit more about the Plingbacks app on appspot for more detail, too.

It's a versatile tool, letting the searcher facet their results and customise the display, and it also ties in with social networks (check out the fantastically-named "boredometer" for example.), and devices.

In short, the Linked Data approach enables Substance to have multiple Plingback widgets that can be presented through multiple channels. Because they all share the same API and data structure, they can use the Talis Platform to query and visualise the data dynamically.

Substance's Steven Flower also told me a bit about a related project building on the back of Plingbacks and the Talis Platform called Plingalytics: a sort of dashboard enabling local authorities and stakeholders to get a very useful view of the Plings datasets. It will let them answer questions like: "How many Plings do we have on a Friday night?" or simply: "What's hot? What's not?"

This ties in with another side of Plings, which works with local authorities to "fulfil their statutory duty to publicise and keep up to date comprehensive and accurate information on positive activities for young people and to make it accessible," according to Substance's site.

It's an exciting project to be working on, and I'm very interested in the way it ties in local government, young people, and activities through a very positive use of the Web. The fact that they're using Linked Data to back the interrelated data makes a lot of sense, and

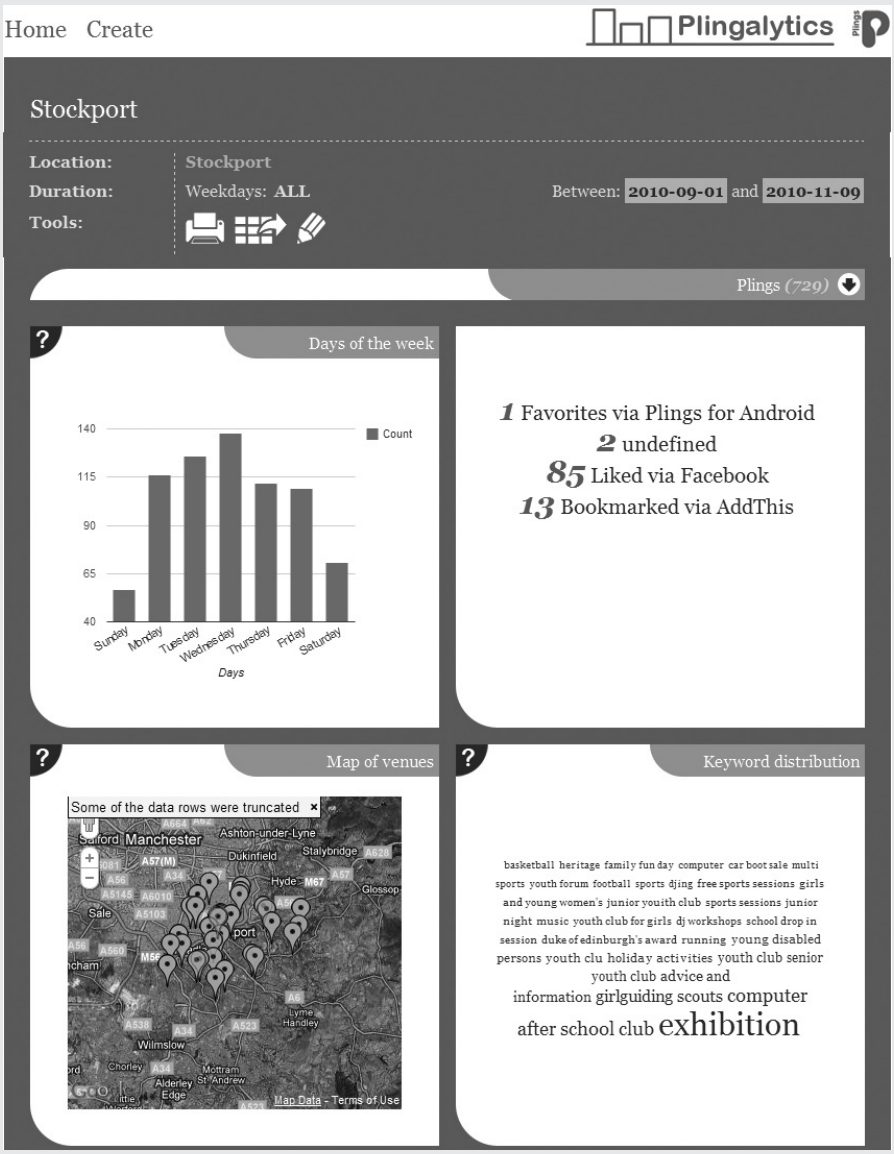
Plings is able to make use of Talis' Connected Commons scheme for some of its data, meaning that not only can this information be managed free of charge, but it's available on an open data licence.

we've been working together for a long time pulling together Linked Data opportunities and matching them with solutions. Alongside looking after the Plingback dataset on the Platform, our consulting team has worked with Substance to model and convert their data to RDF. In addition, and because of the open nature of the data Substance is working producing, Plings is able to make use of Talis' Connected Commons scheme for some of its data, meaning that not only can this information be managed free of charge, but it's available on an open data licence.

Steven Flower said: "We are very excited about this. From a technical point of view, the opportunity to build this upon Linked Data sets is also interesting. Hence, we have chosen to work with Talis for the infrastructure, knowledge, support and enthusiasm that they bring.

We have had the support of Talis since early days of Plings, so it's good to continue."

More information on the Plings project from Substance can be found on their plings.net.



Challenges and Opportunities for Linked Data

By Leigh Dodds

Recently I gave a short talk at Online Information 2010 titled "Challenges and Opportunities for Linked Data". The presentation highlighted what I saw as the main challenges that face us as we grow the web of data, and highlighted some opportunities for organisations that want to get involved. I believe there will be video from the various presentations online at some point, but I wanted to post a transcript of what I said (or had planned to say!). The slides are up on slideshare if you're interested, although they're largely just transitions to highlight my main themes.

Introduction

2010 has certainly been the year of Linked Data. I've been working with RDF and Semantic Web technologies for about 10 years now, and its clear that the last 12 months have been one of the critical growth points for Linked Data and the Semantic Web as a whole. There has been more debate, engagement, and publication of data than ever before. This is in no small part due to the fantastic work that has taken place at data.gov.uk. The project has not only championed the approach but also lead the way as an exemplar for how to do this stuff really well. The adoption of RDFa by Facebook, Google and others has also created a much needed feedback loop that is driving the publication of more structured data. But as the technology grows we're starting to experience growing pains which are presenting challenges for further growth and adoption. I think we're also getting a sense of the opportunities that may arise from the web of data. I picked out three key challenges to review in the presentation.

2010 has certainly been the year of Linked Data. There has been more debate, engagement, and publication of data than ever before.

Craft

The first of these relates to what I'd call "the craft" of Linked Data. To date the growth of the Linked Data cloud has largely been driven by skilled artisans — from academia and a small number of commercial organisations — who know how to work with the technology, how to use and manipulate the data that is already available, and how to get things online and linked together in a way that achieves the 5 star approach. To scale beyond the initial Linked Data community we need to move from an artisan lead approach and enable "journeyman" developers to achieve the same things. There are several facets to this skills transfer. Tooling is clearly one important area. It's a truism that Linked Data tools aren't as polished as they might be. After all it's still a relatively new technology area. The majority of Linked Data artisans have been happy enough either to make their own tools or to work with a disparate selection of tools to get the job done. But there is still a lot more work to do in creating a more integrated toolkit that journeyman developers can reach into to help them quickly and easily publish data.

To be fair though, I think we've needed these past few years of publishing and experimentation to really highlight what those basic tools might be. The other aspect of craft is education and training. There's still a relatively small community with deep skills in this area, so thought has to be given to the ability to transition wider. Having helped train and advise a number of team and organisations over the past few years, most recently as part of our consulting work at Talis, its clear that there's a journey or apprenticeship that many teams and organisations undertake as they begin to experiment and gain experience with the technology. Within the Linked Data community we need to prioritise the work on these tools and services to make it easier for others. We also need to devote additional work to help nuture or define more standard vocabularies for publishing specific types of data. In my opinion this is the real challenging work: it's not as fun or exciting as publishing the next new dataset or exemplar, but it's absolutely necessary to push things to the next level. It's going to take real commitment from all of us. In my mind there is no better way to help pass on the skills of the initial artisan community than by encoding that knowledge in the form of tools, vocabularies, best practices and design patterns.

Fuelling Applications

Linked Data isn't being used as much as it could or should be. Why is this? I think there are two reasons. The first relates to my previous point about enabling the "journeyman" developer. Right now it takes a certain amount of skill to get the most from Linked Data and SPARQL. This presents a road-block for developers who may be interested in using some of the available data. It may even stop them looking at all. To solve this we must be ready to meet people half-way. Publish simple JSON formats alongside the RDF. Use the Linked Data API created for data.gov.uk to provide simple RESTful APIs into your RDF data. Choice opens up more integration opportunities as well as encouraging engagement. The power of SPARQL and other tools is fantastic, but that power is not needed by every developer in every application. Be inclusive when opening up data.

A potentially larger issue is that much of the data available as Linked Data is either static, irregularly updated, or already available in other more accessible formats and APIs. This isn't true across the cloud as a whole, but timeliness is an issue in many areas. It's a consequence of the early boot-strapping process which emphasised conversions of available data dumps, and the wrapping of existing APIs and services. As a boot-strapping process that has been fantastic. But it's not driving engagement: why use data if you can get it somewhere else easier, and in a more up to date form, using tools that you're already familiar with? I also think that this is contributes to the reason why it has been difficult to show the power of Linked Data: many of the demonstration apps could easily have been built with other APIs. I think this could be on the cusp of changing as there is now a critical mass of information available to do some powerful queries, and an increasing amount of data is now becoming primarily available as Linked Data. The challenge we face is changing the nature of the Linked Data cloud from what is a largely static and slow moving environment to one that is much more lively and real-time.

Sustainability

The third challenge I highlighed was sustainability. It's easy to look at the Linked Data diagram and think: "Well, those bits are done, all we need to do is look how to grow the diagram. We just need to add more

data". I think that's a natural but unfortunately misleading viewpoint: we need to look carefully at our foundations. Not all of these sources are on infrastructure that could support real, high volume usage. And few of the datasets are clearly licensed. I've personally encountered a number of occasions where some significant datasets are offline or unavailable. So we need to be realistic about whether people can build a stable, commercial application against the web of data as it exists today. Again to solve this, we need an increasing number of primary sources, making high quality data available on a regular and timely basis, backed by the ability or commitment to deliver those services at the achallenge isn't unique to Linked Data. It's largely true of the web as a whole; after all not every web site or application is intended to scale to high volume usage. But we're now talking about a potentially much deeper integration between different applications.

We can see the same issues occuring around APIs and data access in general. In recent months there have been a number of stories of developers scrabbling to adapt as APIs get changed, taken down, restricted or re-licensed leaving them high and dry. To me the beauty of Linked Data, and RDF specifically, in this regard is that it is so much more portable than any other format. This means that we can easily replicate data to share the load of providing access. With Linked Data we have the option of federating or sharing data across the web. (One of the reasons we started the Talis Connected Commons scheme was to help create sustainability around Public Domain datasets.) The portability of RDF also makes it easier for a range of organisations to offer scaleable value-added services over the same datasets. For the first time we can decouple the curation of data from the delivery of services over that data. So those are my three challenges. I think these are largely point in time issues, but we're going to have to work at them to move forward. What about the opportunities?

Become a Hub

One of the interesting properties of the Linked Data cloud diagram is how it clearly illustrates the emergence of a number of hubs — like Dbpedia — that form the focal points for links from a number of different datasets. If you look closely you can also see that there are emerging hubs within specific subject domains. I wonder whether the hubs that we see today will continue to play such a key role as the web of data evolves? My feeling is that in a few years time the picture and connectivity is going to be quite different. Particularly if we continue to see engagement from government and other sectors. There is clearly an opportunity here for organisations who are already key enablers within a particular sector to become a linking hub on the web of data. If you poke around in any industry, its not hard to find organisations who act as the "switchboard" for that particular sector. Either because they manage some key identifiers for the sector as a whole, or because their identifiers and systems have become de facto standards for achieving interoperability. It would be a natural step for those organisations to carry that role forward to the web of data, retaining that key position. Clearly not everyone can be a significant hub like Dbpedia. But every organisation can act as a hub for its community of customers, partners and users. The reasons and benefits for doing this are well documented: opening up data can drive new business, innovation, and traffic. Success on the web involves giving your organisation the greatest possible surface area and points of attachment. Linked Data is an excellent way to achieve this as to emphasises the right forms of web integration.

Turn Identifiers into Channels

Linked Data requires you to assign URLs to identify things: people, places, events, whatever. Generally we tend to focus on how that is an important step to publishing data: concentrating on the mechanics of what makes a good, stable identifier and highlighting how this becomes a key way for other people to find your data. What this misses is that those identifiers can also become channels, or hooks, for your organisation to find other people's data. Once you have published Linked Data and it becomes linked to by other datasets all of that external data annotates and enriches your own, providing valuable and useful context. Linking data creates network effects, and everyone in the network benefits. That includes you. The external data is easily accessible through link discovery so it becomes much easier to find, aggregate and analyse it for a variety of purposes. That might be to drive new product features, or to simply power business intelligence and analysis within the enterprise. I tend to think of it as being able to fish the web of data for useful context. Your URIs are the hooks. Your data is the bait. I stopped to draw a parallel here with some comments made by Dion Hinchliffe in his opening keynote. Hinchcliffe pointed to the rise of a number of startups and tools supporting analysis of data collected from the open web, perhaps mixed with data from internal enterprise systems. The end results of that analysis is new data and insights that will need to be integrated into an organisations core systems, especially if the intent is to drive more than just management reports. My prediction was that over the next 12-24 months we'll begin seeing this type of third-party organisations not just offering SaaS access to analysis systems, but direct insights that are already integrated into a customer's data via the public identifiers its sharing as Linked Data. This has huge potential value and can completely change the costs and approach to data integration. The time scales may be completely off. But there's a real opportunity there in my opinion, particularly for organisations that do market and social media analysis.

Data as a Service

It's been said before but its worth repeating: Linked Data isn't necessarily Open Data. The technology is not at odds with exploring business models around data services or access. The "Data as a Service" (DaaS) idea is gaining momentum in a number of different areas with an increasing number of commercial APIs coming online. We should also soon be seeing commercially available services directly powered by open data sources or through mining those sources. There are a number of different business models that can be wrapped around data access, ranging from charging for the data itself, through cost recovery for service provision — something that may be relevant for long term usage of government sources — or just charging for delivering reliable, high performance services over open data. There are good reasons why developers may want to pay for reliable services. Clearly open, sponsored access to data and services will remain an important part of the ecosystem. In fact some level of open data is required to drive the network effects we are seeing around Linked Data: the identifiers and some key metadata needs to be open and remain open; but additional "depth" could be available at a premium.

Summing up

I had no big conclusions to draw from my talk as my goal was to highlight the challenges and opportunities ahead. Clearly I could have chosen a different mix but drawing on my recent experiences engaging with a wide range of different organisations these are the issues and opportunities I've most commonly encountered and discussed. Do you have a different perspective? Perhaps some ideas about how to face these challenges, or a different view of the immediate opportunities? If so, I'd love to hear from you on leigh.dodds@talis.com.

Best Buy and the Semantic Web

Jay Myers talks to Zach Beauvais for a Nodalities Podcast.

Zach Beauvais: Hi, welcome to this Nodalities podcast. Today I am speaking with Jay Myers from Best Buy. Jay, can you tell us a bit about what you do at Best Buy?

Jay Myers: Well, my official title is Lead Web Development Engineer. So I work on the bestbuy.com platform and other related initiatives. We have a host of different companies that we have under the Best Buy umbrella that we work with. So, providing guidance to business folks, to get good websites out and get good web experiences out on the web under our brand, and also guiding people towards newer technologies and doing a little bit of thinking about what technology and the web means for retail.

Zach: OK. That's quite interesting because the reason I wanted to talk to you today is because I've heard a lot about the kind of things that Best Buy is doing, and it seems fairly new to be doing Linked Data in retail.

Jay: Well, Linked Data means a lot of things to retail. So as large retailers, we have a huge opportunity to be a source that informs customers and consumers to make intelligent decisions about what they're purchasing. And consumers now have access to billions of products online. And it can be very confusing to find the exact thing that you're looking for if you don't know exactly what you're looking for.

So data is really about keeping people informed and having them make smart decisions that fit what they are looking for. Additionally it's even about knowing where to purchase that stuff.

So we have a lot of really basic valuable data about stores, about store locations, things that people should really know about.

Additionally, as a retailer we have a huge amount of product. At any given time on bestbuy.com you can see almost 500,000 SKUs.

So data, I think has enabled to help us to make more sense of products and develop relationships between products that ultimately benefit our business and benefit consumers.

Zach: OK, and one of the things that you've been looking into — I know because I've cheated a bit, and Googled you — so I know that you are involved in working with RDF and linking up these data that you are talking about. So, what does taking that sort of linked approach bring out?

Jay: So my focus is on linking products together. As retailers often times we don't do a great job of making relationships between products that make sense to people. Obviously, it makes sense for us to present a primary product — something like a TV or a larger ticket item — and then link associated products to it, such as an HDMI cable or a Blue Ray player. And, I believe that Linked Data actually provides the facility to be able to make those connections between products.

Again, that can benefit our consumers by getting this complete package they're looking for without having to go to the store a bunch of times or having to shop multiple times on the website. But it also benefits our business.

Being able to describe what a product is on a web page and being able to have that findable is something that I think Semantic Web can definitely play a part in.

Zach: OK. So you've got a couple of things. You've got the data that you already own from being the people who have the thing you're selling, and you're using that for recommendations?

Jay: We're using that for a number of different things: not only for recommendations but, again, for building product relationships.

Zach: OK.

Jay: We're just basically starting that. So we have amassed a huge amount of data around our products. Obviously when you have 500,000 active products every day, there is a huge amount of data to go through. So we're starting to kind of look into Linked Data and the Semantic Web and do more analysis around products and what works with what. And just getting our feet wet. I can recognise a lot of possibilities using this data in a lot of different ways to benefit our business, and our clients and our customers.

Zach: OK. It's quite exploratory then?

Jay: Mm-hmm. Yeah up to this point it is. My goal personally is to use these technologies to start solving real world business problems. Because exploration is great but unless you can prove some value to these technologies, these ideas tend to die pretty quickly.

Zach: So, have you got in mind a use case when you talk about solving real world business problems? What kind of business problems spring to mind?

Jay: Sure. So, as retailers and as an industry, we actually see a good number of products get returned. We call them open-box products or clearance products or return products or whatever. But they are a detriment to our business because they are expensive. So, let's say somebody goes and buys a laptop; takes it home, doesn't like it, and returns it. There is nothing wrong with that product. But we as retailers have to mark down that laptop and we lose margin on those products and those products already have a very limited margin.

Zach: OK.

Jay: But we don't do a good job displaying those products on the web and making them findable and accessible to people. And that's where the Semantic Web comes in. Being able to describe what a product is on a web page and being able to have that findable is something that I think Semantic Web can definitely play a part in. And one of the things that I've been working with — and a lot of other people have been pioneering — is not only RDF but RDFa, so embedding RDF in HTML documents.

What we're trying to solve is that all of our stores have these open-box or clearance products. The interesting thing is that they're almost hidden. They're "siloed" per store, so some stores can have up to 200 or 300 open-box products, and they're not very visible on the web.

What we've done is actually taken Semantic technologies like RDFa and marked up these products, making them more visible for machines to consume, and then also visible for humans to find. We've definitely seen a positive reaction.

The products are definitely more visible on the web and I think this is an interesting model to make customers aware of them. I think that's really going to benefit our business.

Zach: OK, this mark-up is actually on the public site, so you're actually publishing this. It's not just something that you've marked up for your internal processes?

and services on the web. GoodRelations is supported by Yahoo, which indexes or uses this data in their search results. Google Rich Snippets has support for some RDFa, not specifically GoodRelations, but we're hoping that they increase some support for the more popular ontologies.

Then you look at Facebook and their Open Graph. They're not necessarily a search company, but they're definitely interested in data. That's something that we see as an up and coming thing, where they'll map some of their Open Graph activity to products.

If we can put a "like" button on all of our product pages and have those product pages marked up with data in RDFa using

Every U.S. store has a blog and we give access to a couple of employees per store. They're able to publish things like events and announcements.

Jay: Right. Every U.S. store has a blog and we give access to a couple of employees per store. They're able to publish things like events and announcements. The other thing that they're able to do is to post their own open-box products to the web through this blog interface. It's really easy to do. We've enabled our employees to be able to start publishing links to them and stick data out on the web, which I think is a cool thing and an exciting thing for me. We've made these techniques accessible so you don't necessarily have to get all of the concepts.

Zach: Obviously I can see immediate benefits in publishing human readable information from each store. I can presumably go to a Best Buy, find my local blog and they can say: "Hey, look. We've got these laptops that are all deeply discounted."

Jay: Right.

Zach: Providing the machine readable stuff: that sounds like you want people to actually build something that searches for this information, or are you planning on building it?

Jay: Well, I think there is a good amount of interest from the major search engines already. If you look at some initiatives from a couple of the big players like Yahoo with their Search Monkey; they actually actively support some of the open vocabularies that we use. Best Buy has been using a vocabulary called GoodRelations, which describes products GoodRelations, and then also making that available to humans; the Open Graph project, or Facebook can come and grab it and have a very rich dataset up that's easily accessible and be able to do with it what they want. Hopefully we'll see some business benefit from that.

Zach: All sorts of ideas come to mind when you start talking about that. I'm just imagining that I'm a particular demographic. I'm after this kind of product. Oh, blimey, Facebook knows this and it's saying, "Best Buy happens to have the thing that you're looking for."

Jay: Or taking social shopping to the next level, right? If my friend has liked a specific product and it has these product attributes, then you can do the social linking through Facebook by having open and detailed data about what you're purchasing and what you're liking. You can start making connections that you may have never been able to before.

Zach: OK. I mentioned earlier that my immediate thought was that you've got two sides to this. You've got the data bit that you've already got hold of: the data that you've created as a business and as a retailer. You're also talking about getting a bit more: looking at things like cart sizes and shopping basket contents. Are you looking at aggregating data from people who are shopping as well?

Jay: Sure. I see it in two different ways. I see, with the data we have right now, using that internally to power different applications whether they be customer-facing or not. Then secondly, I see by publishing data openly, we're giving access to other entities — whether they be search engines, Facebook, places like that. We want to be as open with our data and other companies really should be thinking about having their data be accessible on the web. That's what we're driving for.

Zach: Is there any kind of concern about that? If you're publishing all of this data, are you worried about competitors using it against you, or is that an old argument?

Jay: Yeah. I don't know if there is a concern. Honestly, I think competitors will be able to get that data regardless. I don't think we're publishing anything that's new, or anything that would be detrimental to the business. What we are doing is making it more accessible. So, I'm sure you've heard of price aggregation engines like FatWallet, and places like that, that go and scrape our site. They'll find a way, regardless, to get this data. We're just presenting it in a way that's more accessible and letting people use it, which I think makes sense.

Zach: Cool. So, what are the next steps for you and your team then?

Jay: So next steps… we've done some tests and some projects — some smaller projects — with RDFa to try and prove that it is beneficial. And we've seen an increase as far as SEO is concerned, so there are lot of different ties to Linked and Semantic data across the web. Things like findability and accessibility and searchability are really important to us as a business: that we have our products found.

I'm hoping to spread that to our main production site, so start marking up all of our product detail pages with RDFa's and GoodRelations, and start making this markup actually part of something we do every day

So, not really having a specific project in mind to start doing Semantics, but making it something that our developers do every day and are always cognisant of in every project we work on. The first step would be to start marking-up things should be (like product details) in RDFa and make those accessible externally.

Zach: OK.

Jay: As far as internally is concerned, my goal and vision would be to start looking at how we can use all this data we produce on a daily basis. Wo we have hundreds of thousands of employees who are all kind of little data sources. And then we have customer input. We

have product input. How can we start to use this data semantically to start getting insights from it. Because there are messages in there that I think that we could be missing.

So, if we can somehow harvest this data and use it for business benefit and use it for customer benefit, I think there are huge amount of efficiencies that can be gained from doing that: from having tools that do analysis on all the data that we produce on a daily basis.

Those are a couple of the things both internally and externally that we hope to work on in the next six months to a year.

Zach: OK, brilliant. So is there anything that you'd like to add to this for the Nodalities community?

Jay: Definitely. So there is a huge amount of momentum around semantics, and with the recent Semantic Technologies conference in San Francisco, there's a lot of talk about RDFa. That's one of the things that I'm pretty passionate about: RDFa and microdata, and some of these other markup techniques allow everyday web developers to start marking up their HTML with very specific data in mind. So using the ontologies and vocabularies that have been established or creating new ones, to make things more machine readable and also to make them customer and human friendly, something we've grown accustomed to on the web.

There's a huge amount of momentum making the web accessible to everyday developers, and that's what I hope to focus on over the next year is getting developers more in tune with what they can start doing.

So you don't have to have a huge enterprise software package that publishes Semantics. You can start on a very basic level publishing data in RDFa.

And one of the things that I'll make a call out to people who are interested in Semantic Technology is we have a number of people who are submitting proposals for the South by Southwest Interactive Conference that's coming up in March initially.

A portion of the way that they determine who's going to be presenting at SxSW is actually votes from the general public, so we have several people that are proposing different topics to speak on.

If you're interested in hearing more and making things more accessible to everyday web developers and start publishing semantics through HTML in a very open and easy way, we'd love to have your support in getting some votes for our proposals so we can start to teach this message to the interactive people that typically would attend South by Southwest.

And so hopefully we can start spreading this message, because there's a huge amount of benefit to us as consumers, to us as just people, and there's a huge amount of information out there that does need some structure and should be structured so we can start gaining insights and start seeing efficiencies, especially in retail, but also in everyday life.

Zach: Cool. So good for Best Buy, good for the development community, and better news for consumers. That sounds like sort of a triple win, doesn't it?

Jay: Yeah, it does. That's what I'm interested in, and I think everybody benefits from being more informed, and that's what I think Semantic Web and Linked Data is really all about.

Zach: Brilliant. Well, Jay Myers, thank you very much for your time.

Jay: Thank you.

Nodalities Magazine is a Talis Publication

ISSN 1757-2592

Talis Group Limited
Knights Court, Solihull Parkway
Birmingham Business Park B37 7YB
United Kingdom
Telephone (within the UK): 0870 400 5000
Telephone (outside the UK): +44 121 717 3500
www.talis.com/platform
nodalities-magazine@talis.com

WIMS'11

International Conference on

Web Intelligence, Mining and Semantics

25 - 27 May 2011
Sogndal, Norway

Attend a new conference concerned with intelligent approaches to transform the World Wide Web into a global reasoning and semantic-driven computing machine

Program Chair

Rajendra Akerkar, *Vestlandsforsking, Norway*

Advisory Committee

Amit Sheth, *Ohio Center of Excellence on Knowledge-enabled Computing, USA*
Frank van Harmelen, *Vrije Universiteit Amsterdam, NL*
Grigoris Antoniou, *FORTH- Institute of Computer Science, Greece*
Guus Schreiber, *Vrije University Amsterdam, NL*
Harold Boley, *Semantic Web Laboratory, IIT, NRC, Fredericton, Canada*
James Hendler, *Rensselaer Polytechnic Institute, Troy, NY*
Takahira Yamaguchi, *Keio University, Japan*

Keynote Speakers

James Hendler
Marko Grobelnik
Peter Mika
Sören Auer

For more information, visit: wims.vestforsk.no

Organised by:
Western Norway Research Institute
VESTLANDSFORSKING