Nodalities

THE MAGAZINE OF THE SEMANTIC WEB

talis®

www.talis.com/nodalities

Issue 9

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
www.twitter.com/nodalities

Open Government Data

By Dr. Dean Allemang

It has been about one year since President Obama issued a Memorandum on Transparency and Open Government[1]. That year has seen a flurry of activity around open government data covering all aspects of government operation, ranging from internal management of government information, to publication of government data to the general public. The Semantic Web standards, in particular RDF, SKOS and OWL, play a key role in all of these activities[2] [3].

Open Government Data

One of the highest profile efforts goes by the name data.gov, whose concept of operations (ConOps) was recently finalised and explicitly includes a role for RDF and OWL in the publication and sharing of data[4]. Data.gov makes available a large number (over a thousand at the time of writing this article, and increasing each day) of official datasets, on topics ranging over all aspects of government activity.

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 talis.com/nodalities or email nodalities-magazine@talis.com.

MEET US

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

FOLLOW US

Follow Nodalities Magazine on Twitter: @nodalities

Editor's Notes

Welcome to Nodalities' 9th issue and first of the new decade. The opening months of 2010 have already been busy ones for the Semantic Web community, with big-news events starting to match the trends we've been creating for the past few years. As an example, the dust-sheets came off the new data.gov.uk project here in the UK: proving that there is an appetite for large organisations and public bodies to begin taking an active role in publishing data. As a bonus, the UK has made it part of its remit to make the data as linked as possible.

Along with the notion of opening up data, or publishing it freely out on the open web, is the natural follow-up of reuse. What can be done with this information once it's published? What does it mean to make use of published data? There has been a lot of discussion around the clear licensing of data and the building of community standards around it. It is increasingly important for people to understand what they can do with the data they're looking for, and there are a series of articles contained in this edition which address it.

I hope 2010 is a prosperous and interesting year, and I also hope to see more exciting projects launched soon. Also, if you're a part of such a project, I'd love to hear from you for Nodalities. Drop me a line: zach.beauvais@talis.com.

-Best,
Zach Beauvais

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
talis.com/platform
nodalities-magazine@talis.com

Continued from front page.

It has been about one year since President Obama issued a Memorandum on Transparency and Open Government[1]. That year has seen a flurry of activity around open government data covering all aspects of government operation, ranging from internal management of government information, to publication of government data to the general public. The Semantic Web standards, in particular RDF, SKOS and OWL, play a key role in all of these activities.[2] [3]

While the open government initiative makes great strides, obstacles are still plentiful.

Open Government Data

One of the highest profile efforts goes by the name data.gov, whose concept of operations (ConOps) was recently finalised and explicitly includes a role for RDF and OWL in the publication and sharing of data[4]. Data.gov makes available a large number (over a thousand at the time of writing this article, and increasing each day) of official datasets, on topics ranging over all aspects of government activity.

Open government data enables what Kingsley Idehen calls the "citizen analyst" private citizens with effective access to public government data. Once exclusively the domain of dedicated watchdog organisations and investigative journalists, government data in particular fiscal data can now be accessed, organised and analysed by anyone with a bit of ingenuity, perseverance, and a particular topic to investigate. Each day, new applications are being added to the Open Government Innovations Gallery[7], empowering a wider variety of citizen analyst. But while the open government initiative makes great strides in this direction, obstacles are still plentiful.

The Data-gov Wiki at the Rensselaer Polytechnic Institute[8] provides a jump start to the cottage industry of citizen analysts by re-hosting the datasets from data.gov in RDF form. This removes one of the technical hurdles for utilising open government data. In the raw data.gov catalog, information appears in a wide variety of formats, including spreadsheets, XML and RSS feeds. While it is logically possible to render each of these as RDF in a straightforward way, the transformation is just one more step that the citizen analyst has to take. The Data-gov Wiki has already made that transformation, and boasts a dozen or so example apps of its own.

Ralph Hodgson and his colleagues at TopQuadrant have initiated the Open Ontology for eGovernment (oeGOV)[12]; a clearinghouse for models, vocabularies and ontologies for government. It recognizes the key role that shared models play in the effective utilization of linked open data. OeGOV includes models of government organizations, as well as listings of information about specific government institutions. It also includes a representation of the FEA in SKOS, a simple knowledge organization standard based on RDF and supported as a Recommendation by the W3C. oeGOV also serves as an open publication point for QUDT, a comprehensive model of quantities, units, dimensions and datatypes. The charter of oeGOV is to become a central index where high-value models, thesauri and vocabularies can be vetted and published, providing a reliable source of shared specifications, supporting effective use of linked open government data.

Data Interoperability

The Data.gov ConOps describes the publication of government data, but goes further to describe how that information can be linked together. It correctly identifies a key notion for wide-scale free exchange of information unambiguous description of key interchange concepts. The simplest and easiest way to manifest this sort of unambiguous description uses a tried-and-true technique a controlled vocabulary.

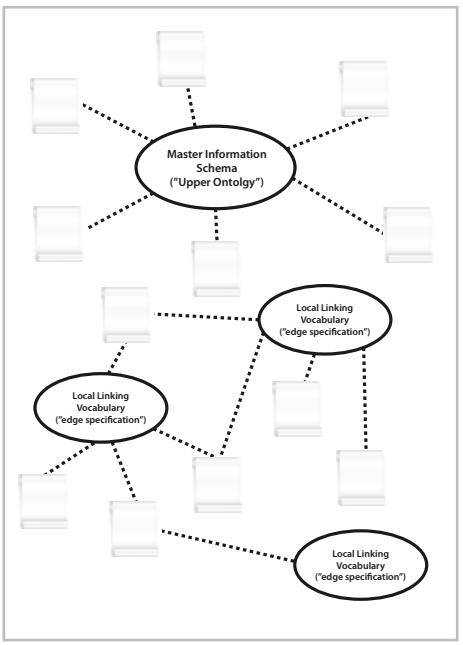
Many controlled vocabularies have been successfully deployed over the years. Of particular interest to government data is the Federal Enterprise Architecture Reference Model (FEA)[5], which provides, among other things, an unambiguous identification of lines of business that government agencies are involved in. The FEA has been in use for the past several years as a source of annotations for the Exhibit 300, a form that must be filed by every agency when undertaking a project that requires an IT budget. Even though most Exhibit 300 filings took place before the advent of data.gov, the discipline of using the controlled FEA vocabulary can help organise them today.

A very simple example of how the FEA vocabulary increases the analytic power of data was shown at the International Semantic Web Conference in October 2009[6]. The use of the FEA vocabulary makes it possible to ask questions like, "How much is each agency spending on a particular line of business?" Linking together all the Exhibit 300 data, and generalising them, is made possible because of the shared reference of the FEA vocabulary.

The utility of the FEA vocabulary goes beyond analysis of Exhibit 300 data. Many agencies use another facet of the FEA that refers to technology to describe their own internal information systems; this means that you can use the same approach to compare, say, vulnerability to a system-specific computer virus from one agency to the next.

Linked Data

Key to all linked data efforts is a global naming scheme. In order for two data sets to be linkable, even in principle, there must be some way for multiple data sets to make common references. This is where RDF as a foundational technology for linked data excels. The Web infrastructure already provides a global naming system in the URI. One of the most subtle contributions that an effort like the Data-gov Wiki[8] provides is a consistent mapping of public data sets to URIs.



A common misconception about linked data in general, which applies to government linked data in particular, is that there must be a single, master scheme by which concepts are mapped to URIs. As Tim Berners-Lee points out[9], this approach rarely works. A much more successful approach is to build a network of linked information. Creating a network of linked data requires much less commitment at the outset than a master data schema approach. Two datasets can express disagreement about an entity, while agreeing on a way to refer to the thing they are talking about.

Focusing on building a linked data web shifts the emphasis in data modelling from an attempt to find the 'right' representation of information structure, to a decentralised network of specifications (see figure). In the decentralised model, there is no center to the figure; every schema is on equal footing. Since no data specification is in the centre of the diagram, they are sometimes referred to as 'edge specifications'.[10]

In concrete terms, vocabularies like the FEA are exactly the sort of specification that play this role. Multiple data sets like those provided by Exhibit 300 can refer to terms in the FEA vocabulary. Some of those datasets might refer to other standard vocabularies ranging from very specific structures like the NIH Bricks[11] to comprehensive vocabularies like AGROVOC[12]. Specifications of this sort play a key enabling role in constructing a network of linked open data. There is no need for every data set to refer to the same linking vocabulary. But whenever two datasets happen to refer to the same one, there is opportunity for linkage. The utility of such controlled vocabularies has been known for centuries, but the Semantic Web technologies, especially the linked data capabilities of RDF, translate this value into a linked data web.

Linking Data Across Space and Time

The effective management of global identifiers provides advantages that go beyond enabling the citizen analyst. For example, The National Aeronautics and Space Administration (NASA) faces a unique challenge with its Constellation Program, a long-term plan to return to the moon and proceed onward to Mars. The Constellation Program requires that information be shared by diverse groups across the program, including engineers, technicians, operators and managers. Through the use of semantic web representations and controlled vocabularies, information can be reliably aggregated to inform decisions across diverse organisational groups and application domains[15].

In the Netherlands, the Ministry of Justice (MoJ) faces very different challenges. The Ministry manages legal records, such as simple records like birth, death and marriage, and more complex contentious records like criminal charges sentencing. Information needs to be exchanged between many parties in the Ministry. For these exchanges the MoJ has developed an ontology-driven Metadata Workbench. OWL models that comply with statutes of law are transformed into XML schemas for business documents that have to be exchanged between the various parties.

But as laws change regularly, and often in ways that can seem capricious or even irrational from the point of view of the record keeper, the solution has to be flexible and able to cope with changes to the ontology models. Documents kept in accordance with yesterdays regulation have to remain, as close as possible, consistent with today's documents. There is no hope of creating a single 'correct' specification and sticking to it flexibility is the very nature of a legal system. For this reason, the ministry of justice has chosen an RDF and OWL-based technology for managing the continually changing metadata and message design needs of legislation and regulation.

The effective management of global identifiers provides advantages that go beyond enabling the citizen analyst.

The Road Ahead

The Semantic Web has been identified as a key technology for open government data. Currently, little of the data.gov data is provided directly in RDF, but with data-gov.wiki and the emphasis in the ConOps on Linked Data, more and more data will be available in readily linkable form. This trend will allow more and more applications to exploit multiple data sets. But providing data in RDF is just the first step; the publication and adoption of common vocabularies, also in RDF, is a key enabler for linking government data. Even when vocabularies are provided by mandate (like the FEA), they also must be published as linked data. Look for more independent publications of this sort (like oeGOV), or institutional publication (just as the United Nations itself publishes AGROVOC in SKOS).

The availability of linked government data, along with applications like those already appearing at data-gov.wiki, has already begun to trigger a network effect, whereby it is productive to provide more linked data for existing apps, and more apps for existing data. This trend will continue, making the task of being a citizen activist more and more accessible a less technical audience.

Within the government sector itself, advances like those at the Dutch Ministry of Justice suggest that a linked data approach to managing government data can provide an effective way to deal with the flexibility and volatility of legislation. This, in turn, will allow government agencies to be more responsive to the will of the people.

Looking further ahead, we see that these efforts are just the starting point. New legislation mandates increasing visibility in every aspect of our economy, from carbon footprints to food safety[14]. The importance of open data in the government is here to stay, and the Semantic Web standards provide the technology to make it happen.

Dr. Dean Allemang is Chief Scientist at TopQuadrant and co-author of Semantic Web for the Working Ontologist.

1. Obama, Barack. Memorandum on Transparency and Open Government, http://www.whitehouse.gov/the_press_office/Transparency_and_Open_Government/
2. Wisnosky, Dennis. New Directions in Building Architectures in Support of Business Operations http://admin.actgov.org/KNOWLEDGEBANK/DOCUMENTSANDPRESENTATIONS/Pages/default.aspx?RootFolder=%2Fknowledgebank%2Fdocumentsandpresentations%2FDocuments%2FShared%20Interest%20Groups%2FEnterprise%20Architecture%20SIG&FolderCTID=&View={77884C43-97A1-4FA6-B65C-B52711DAEAEA}
3. Jackson, Joab. 13 resources for government Linked Data http://gcn.com/articles/2009/11/09/linked-government-data-sidebar-01-resources.aspx?sc_lang=en
4. Data.gov Concept of Operations. http://www.ideascale.com/userimages/sub-1/736312/ConOpsFinal.doc
5. Office of Management and Budget. Federal Enterprise Architecture Reference Model, http://www.whitehouse.gov/omb/e-gov/fea/
6. Jackson, Joab. 2 examples of how government data linking can work. http://gcn.com/Articles/2009/11/09/Linked-Government-Data-sidebar-Gvmt-Examples.aspx?Page=1
7. Open Government Innovations Gallery. http://www.whitehouse.gov/open/innovations
8. Rensselaer Polytechnic Institute. Data-gov Wiki. http://data-gov.tw.rpi.edu/wiki/The_Data-gov_Wiki
9. Berners-Lee, Tim. Putting Government Data online. http://www.w3.org/DesignIssues/GovData.html
10. Murphy, Rick. Linking Open Data: An Emerging Practice Area for the Semantic Web. http://phaneron.rickmurphy.org/?p=34
11. National Institute of Health. What is a "brick"? http://enterprisearchitecture.nih.gov/ArchLib/Guide/WhatIsBrick.htm
12. United Nations Food and Agriculture Organization. Agricultural Information Management Standards (AIMS) http://www.fao.org/agrovoc/
13. Open Ontology for eGovernment ("OeGOV") http://www.oeGOV.org/
14. Food Law Legislation Pending in the 111th Congress. http://www.cspinet.org/foodsafety/foodbills.html
15. TopQuadrant Helps NASA Launch the Constellation Project (press release) http://topquadrant.com/docs/pr/06_%2002_%2009%20_NASA.pdf

Open data and the law

By Jordan S. Hatcher, Open Data Commons

Open data, is becoming increasingly important. In the UK alone in the few short weeks we've had of 2010, we've had the announcement of the London DataStore and more work on www.data.gov.uk to come. This isn't to mention the issues that are at the heart for many Nodalities readers linked data and the legal aspects at the heart of the webs future. Recently I had the honour of partcipating in a half day workshop on the legal issues behind open data at the International Semantic Web Conference with Kaitlin Thaney of Science Commons and Tom Heath and Leigh Dodds of Talis.

Three key points when thinking about the legal aspects of open data came out of the workshop:

1. Legal rights can cover data & databases
2. Legal rights require legal solutions
3. Legal solutions for open data exist

We'll cover each in turn, but first:

Defining open data

Open data often gets referred to in two different contexts:

- Open-as-in-accessible the practical bit of being able to see and play around with the data.
- Open as in both legally and practically accessible often as defined under such standards as The Open Definition.

This article examines the legal side, and so uses the Open Definition, which defines open as:

- Allowing use, reuse, and redistribution
- Not allowing for discrimination on the type of use made (thus non-commercial restrictions would not be open)
- Allowing for restrictions such as attribution requirements or share alike (a/k/a copyleft or reciprocal licensing).

For more details, see www.opendefinition.org.

Legal rights do cover data and databases

Perhaps surprisingly for some (though perhaps not for Nodalities readers), data and databases are not a rights free area where no intellectual property rights apply. International trade agreements such as TRIPs, for example, requires that members of the World Trade Organisation, including Canada, the United States, and the UK, provide legal protection for databases[1]. Rights covering databases can include:

Copyright both for the selection and arrangement of the database contents and over the contents of the database itself (the data), though factual information will generally not be protected by copyright.

Database rights The European Union's Database Directive[2] requires member states to implement a sui generis database right covering the the extraction and re-utilisation of the contents of protected databases.

Contractcontractual obligations about what users can and can't do with a database and its contents can also be used to provide for protection.

Other rights rights such as trade secret and laws of unfair competition can also protect databases.

In addition, just as in many other areas, patents (whether software patents or not) and trademarks (over the database name, for example) can be an issue. Even further complicating the picture when discussing online open data, the above rights may have some broad harmonisation internationally, but differ greatly in the details inside each country. There's no copyright law of the internet and what will often matter most for online open data will be the law of one country (and that country might not be your own).

There's no copyright law of the internet and what will often matter most for online open data will be the law of one country (and that country might not be your own).

All of these rights produce a rights thicket protecting databases and data and can form a significant obstacle for the use and re-use of data. This is true for both the scientific community wishing to expand knowledge through use of others' data, for the internet and research community with aims to enable the semantic web, and for businesses that wish to incorporate open data into their business model.

The data/database distinction

Often when people talk about data they only mean factual information water boils at 100 and the like. But in the context of open data, the contents of a database can be anything, including:

- Mobile video
- Flickr images
- Birth dates
- ID numbers
- Haiku
- Tweets
- Facebook status updates

and anything that can be stored in a database can be data in this context. That's why it can be helpful to distinguish the two as the contents of a database and the database. This is the data/database distinction, and from a legal perspective it means:

- There may be legal rights specific to the contents of a database (the data).
- Any collection of contents (the database) may have legal rights covering that collection that are independent of the rights over individual contents within the collection.

So for example, if you had a database of still images and video of zombie movies from the 1960's and 1970's, you:

- May have rights, including for your selection and arrangement of the zombie films, over the database as a whole.
- You'd still have to clear the rights (such as copyright) over the images and video you included inside your database.

It's a bit like a layer cake, the bottom layer rights over the contents, the top layer rights over the database.

Factual information and the data/database distinction

Factual information as a general rule won't have copyright (though other legal rights may be an issue) the boiling point of water won't be copyrightable. But collections of factual information into a database can trigger copyright, and copyright can subsist over various elements of a database (such as field names, the schema, or tables). So while one layer may not have any copyright (facts in the data/contents layer), there may still be copyright and other legal rights such as the EU database right that cover a database of factual information.

As an aside, it's important to note that

privacy and data protection law may also be an issue when dealing with (factual) personal information, so there's more than just IP rights to think about when talking about data.

Legal rights require legal solutions

Legal rights mean getting lawyers involved, which means money (hourly billing) and risk (hauled off to court). For some, it may not be an issue, but if the goal of a database provider is to encourage use and reuse of their data, then addressing at least some of these legal issues will be necessary, especially for commercial users to feel comfortable. Say for example someone puts up a database on the web without any statement around the rights over that database. So what is a user of a such a database to do? Four options:

1. Ask for a license from the rightsholder(s) (likely the database provider)
2. Use the database under a legal exception, such as fair dealing/fair use in copyright.
3. Infringe (and take the risk of getting caught and paying damages)
4. Find an alternative.

Even finding out whether your use qualifies under an exception to copyright law often involves a lawyer, and asking for licenses (and drafting them) can be time consuming and cost prohibitive, and the laundry list of all the different legal issues from the previous section must be taken into consideration when opening data copyright, database rights, contract.

Even finding out whether your use qualifies under an exception to copyright law often involves a lawyer, and asking for licenses (and drafting them) can be time consuming and cost prohibitive

Wouldn't it be great if at least some of these were taken care of for users up front, and that solution at least tried to work worldwide? This is where open licensing comes in.

Open licensing

Open licensing addresses one of the four options set out above granting and asking for licenses and does it in a way that complies with the Open Definition. By making an open license a publicly usable resource, that open license (whether for software, content, or data) can then become a standard that encourages greater interoperability. To use a term from economics, having a public open license helps lower the transaction costs of licensing by making it as easy as possible (either as a licensor or as a licensee).

Open Data Commons

With the funding and support of Talis[3], we founded the Open Data Commons project in 2007 to provide legal tools for sharing data (and at www.opendatacommons.org). This project is now hosted by the Open Knowledge Foundation, a not-for-profit organisation promoting open knowledge. Fast forward to today, and we have two main legal tools available:

Open Database License (ODbL) this works solely on the database layer and requires both attribution and share alike. Equivalent in many ways to the Creative Commons BY-SA licenses.

Open Database License (ODbL) this works solely on the database layer and requires both attribution and share alike. Equivalent in many ways to the Creative Commons BY-SA licenses. The ODbL can be used with the Database Contents License (DbCL) to cover the rights (if any) over the contents of the database.

Public Domain Dedication and License (PDDL) this gives up all copyright and database rights over a database and its contents and so places a database into the public domain. This tool works with our Community Norms document to provide for a set of non-binding expectations for how the data should be treated (such as norms of citation and attribution).

Both tools were developed after long public comment periods and are available for use by the public. They're designed to be able to operate worldwide.

The Public Domain approach

In December 2007 Science Commons released their Protocol for Implementing Open Access Data[4] This protocol, written in the same style as a Request For Comment (RFC)[5], outlines a legal standard for open access to data based on three principles:

3.1 The protocol must promote legal predictability and certainty. 3.2 The protocol must be easy to use and understand. 3.3 The protocol must impose the lowest possible transaction costs on users.

The public domain approach, eliminating rights such as copyright and the EU database right, complies beautifully with all three of these principles. The public domain, unlike even minimal restrictions such as attribution and share-alike requirements, provides for the maximum amount of interoperability for the least amount of hassle. One no longer has to clear copyright permissions if there is no copyright to clear.

Science Commons has set a standard with its Open Access Data Protocol, and both the PDDL and Creative Commons' own CCZero tool comply with the protocol. Because the public domain is by definition interoperable, our PDDL tool works with the CCZero tools and even other public domain dedications. Science Commons set the bar high when they opted for a public domain approach, but the result provides for maximum interoperability and future-proofing in a way that other approaches don't provide.

How you can participate

The Open Knowledge Foundation works in a bottom-up, open and collaborative manner, and you can support either Open Data Commons as an individual project or OKF's work by joining one of our mailing lists and actively contributing. Or if you like what we do but don't have time to contribute, you can become a supporter or just drop us an email and let us know how we are doing.

OKF has been engaged in a number of open projects, including CKAN, Open Shakespeare, and data visualisations such as Where Does My Money Go? We regularly put on events such as OKCon (every Spring), and are always on the lookout for ways that we can promote open knowledge, such as by speaking at events. Please get in contact if there's anything you think we should be doing.

Jordan Hatcher (jordanhatcher.com) is a co-founder of Open Data Commons and, among other things, a lawyer, academic, and entrepreneur working on Intellectual Property and Internet law issues in the UK and worldwide. Currently he is Head of Research @ ipVA and on the Board of Directors of the Open Knowledge Foundation.

1. TRIPS Article 10 2 http://www.wto.org/english/docs_e/legal_e/27-trips_04_e.htm
2. Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases.
3. http://www.talis.com/
4. http://sciencecommons.org/projects/publishing/open-access-data-protocol/
5. For more on this process, see http://en.wikipedia.org/wiki/Request_for_Comments

Sharing data on the Web

By by Kaitlin Thaney, Program Manager of Science Commons

In the emerging data web, there have been multiple efforts working towards the same broad goal of data sharing (ie., the NeuroCommons, Linked Open Data, efforts of the World Wide Web Consortium), but are still unevenly distributed. Our understanding of the legal, social and technical issues is increasing, but is still at a very early stage.

This past fall at the International Semantic Web Conference in Chantilly, VA, USA, I joined three other leading minds to lead a tutorial examining some of the legal and social frameworks for sharing data in the emerging data web, focusing on an overview of the need for access, the social issues of applying Free-Libre/Open Source (FLOSS) licenses to data, and the approach we advocate at Creative Commons to help navigate this complex space converging on the public domain.

Lessons Learned

Creative Commons as an organisation works to make knowledge sharing easy, legal and scalable with applications in the culture space (music, text, film, art), education (open educational resources, virtual textbooks), and science (biological materials transfer, data sharing, Open Access, semantic web, patents). We maintain an integrated approach, and craft policy and legal tools to lower the barriers to knowledge sharing.

When it comes to data sharing, first and foremost, the information needs to be legally and technically accessible. The Open Access movement has increased awareness of this, using the Creative Commons licensing suite to unlock content, and has seen its share of qualified success. But what to do when the information you want to share and reuse falls outside the protections of copyright?

When it comes to data sharing, first and foremost, the information needs to be legally and technically accessible.

In short, it's complicated

This is the where the discussion of legal protections for data gets murky. Knowledge is not always copyrightable it may be easy to discern the rights associated with journal articles; but what about data, ontologies, annotations, or research statements described in triples?

The emergence, adoption, and use of the free-libre/open licensing regimes has allowed for remix and reuse of software code, music, film, educational resources and scientific research in a way that otherwise would be difficult to achieve.

The successes of these licensing approaches has caused a change in the social ethos of licensing, instead using a traditional all rights reserved model to make something more free, rather than less.

The successes of these licensing approaches has caused a change in the social ethos of licensing, instead using a traditional all rights reserved model to make something more free, rather than less.

But from our research, this approach is not ideal for data. The trend towards applying licenses, click-wrap agreements and other sorts of restrictions on scientific data is increasing, but with the undesired consequence of limiting the downstream use of this information, and even at times blocking interoperability. The costs are high, the terms are not always clear, nor the protections always legally sound, making it very difficult to scale for scientific uses. The result is a high barrier to entry to do meaningful analysis, annotation, search, etc. on the mass of data available currently that's continuing to grow exponentially, and integrating with the literature available.

We advocate an approach of converging on the public domain, and requesting behaviours often found in the various flavours of free and open licensing through norms not a legal construct. But first, let's take a look at some of the issues to be aware of and their social implications to furthering the goal of linked open data.

Attribution v. Citation

Under US Copyright law, "Copyright does not protect facts, ideas, systems, or methods of operation, although it may protect the way these things are expressed." Since facts are not covered by copyright, attribution a license obligation doesn't seem to apply to ideas or facts either, since those rights are conditional on compliance with terms of the license.

Socially, the scholarly concept of citation is fairly well understood credit where credit us due. It has long been viewed as an entrenched norm of good scientific practice.

But when it comes to the legalities of both terms and how to enact this behaviour, the devil is in the details. The two are actually rather different when it comes to enforceability and applications / ramifications in the digital world.

In a copyright license, the word "attribution" is a legal requirement, whereas citation evokes more of a club mentality and social practice. Citation in its sole form is not assured or enforceable in the same way, but that's not necessarily a downside. Ask yourself this, which one is more important legal enforcement or credit enforced through professional reputation? Attribution a relatively narrow legal term that can affect interoperability while at the same time possibly failing to provide what you really want? Or citation an entrenched scientific norm that asks for credit where credit is due.

Implications of FLOSS toggles and directives on data sharing

These issues emerge when instead of focusing on maximising interoperability of resources, one applies a property metaphor to data. And in the digital world, that tendency can have quite limiting ramifications to future use of the information, as technology continues to outpace the social components to data sharing.

...in the digital world, that tendency can have quite limiting ramifications to future use of the information, as technology continues to outpace the social components to data sharing.

Misunderstanding the legalities can lead to category errors on the social level, including unintentional infringement or on the other side of the spectrum, choosing not to use the resource for fear of infringement. The intentions are often good believing that applying a less-restrictive copyright license is ensuring the data

can be freely shared, reused, and built upon. But without existing precedent or involving a legal team, these issues make for a problematic area to navigate, creating additional confusion and burdens for the users, as well as data providers.

Let's look at a few examples to gain a better understanding.

Non-Commercial: When used in the context of data, what is a commercial use of the data web? Is it the extraction of a subset, a query that may touch on the data set, hyperlinking?

Attribution: As detailed above, the definitions of attribution and citation are often conflated. Attribution speaks to the legal requirement triggered by the use of the work. But in the case of linked open data, if one were to run a query involving 30,000 data sources (something that is happening every day at an ever decreasing cost), would they then be required to attribute the contributors for all 30,000 databases? You can see how this unintended consequence of attribution stacking could impose a very daunting task for the researcher.

Share-Alike: This toggle specifies that any derivative product be relicensed under the same terms. In the example above of running a large query, all it would take would be one database licensed with a share-alike provision for the entire derivate work to then be under the same terms and no other license. This leads to compatibility issues.

There are other external mechanisms and limitations imposed by various jurisdictions and countries that can have a profound effect on data-sharing, especially in terms of international data sharing efforts. These include the sui generis database directive in the European Union, Crown Copyright, "sweat of the brow" and "industrious collection" limitations, trade secrets and unfair competition laws, adding another dimension of complexity to an already complex arena.

After convening a series of meetings, roundtables and other discussions with members of the scientific community, the need emerged for a legally accurate and simple solution, that reduced and/or eliminated the need for one to make the distinction of what's protected. The conflict between understanding the legal issues and complexities can best be resolved by a two-fold approach:

1. a reconstruction of the public domain and
2. the use of scientific norms to request behaviour through a non-license means.

Converging on the Public Domain (+ Norms)

We believe that the public domain is the best means to achieve maximum interoperability of data with the lowest imposed burdens on the user. This can be achieved through the use of a legal tool either the Creative Commons CC0 Waiver or the Public Domain Dedication and License (PDDL) waiving all intellectual property rights asserting that the provider makes no claims on the data. These tools put the work as closely into the public domain as possible.

It calls for data providers to waive all rights necessary for data extraction and re-use (ie., copyright, sui generis database rights, claims of unfair competition, implied contracts). It also requires the provider place no additional obligations such as copyleft or share-alike on the information, which could limit downstream use, as discussed above.

Science Commons also crafted the Protocol for Implementing Open Access Data a protocol for evaluating database terms of use, in hopes of providing a unified framework for users to evaluate if any given database may be integrated with any other database.

The Protocol recommends one request behaviour, such as citation, through norms and terms of use rather than as a legal requirement based on copyright or contracts.

We are aware that different disciplines and jurisdictions call for different approaches, and this is not always a one-size-fits-all solution. With requesting behaviour through norms and terms of use rather than a legal construct, various scientific disciplines have the ability to develop their own norms for citation, allowing for legal certainty without constraining one community to the norms of another.

Final Thoughts

In the early days of the World Wide Web, there weren't many free-libre licenses available, and after a debate over using GPL for the original web code, CERN chose to put it into the public domain. Getting the law out of the way was key to allow for network effects, and to the success of the Web.

Converge on the public domain and ensure the freedom to integrate. It's the most scalable solution.

http://sciencecommons.org/about/whoweare/thaney/

http://neurocommons.org http://linkeddata.org http://www.w3.org

http://www.copyright.gov/help/faq/faq-general.htm

Conference Report: ESTC2009

By Lejla Ibralic Halilovic, STI International

For the third consecutive year, the European Semantic Technology Conference (ESTC) took place in Vienna, Austria. ESTC—organised and hosted by STI International—Attracted over 240 participants from 25 different countries to meet up, focusing on the growth of ICT semantic technology markets. The two-day conference took place at the modern Hotel Le Meridien, offered not only an interesting program, but also excellent networking opportunities for all participants.

The opening keynote entitled "Visions and Strategic Directions for Semantic Technologies" was delivered by Dr. Michael Brodie from Verizon Communications and Dr. Mark Greaves from Vulcan Incorporations. The first day continued with presentations from sessions focused on business applications, case studies and technological innovations. The afternoon X-Media workshop focused on knowledge management drew quite a crowd with presentations from Rolls Royce, FIAT, University of Sheffield, University of Koblenz and Quinary spa.

The highlight of the first conference day was the Innovation Seed Camp sponsored by AWS—the Austrian promotional bank.

The highlight of the first conference day was the Innovation Seed Camp sponsored by AWS—the Austrian promotional bank. 15 startup companies pitched their business ideas to the audience and afterwards had the possibility to get feedback from investors, experienced entrepreneurs and experts within the field of semantic technologies. The innovation seed camp was closed with an expert panel on the topic "How to gain venture capital investments."

The first day came to a close with organised matchmaking sessions, whereby the participants held blitz business-to-business meetings with previously selected partners, supported by the innovative B2Match matchmaking tool.

In the evening the participants moved to the Palmhouse restaurant across the street for the conference dinner, complete with buffet and live music. The winners of the Innovation Seed Camp were announced during the evening. First prize went to Felix Van de Maele from Collibra (http://www.collibra.com). Karin Haager from Flimmit GmbH (http://www.flimmit.com/) was awarded with the second prize, and Leo Sauermann from gnowsis.com (http://www.gnowsis.com) received the third prize in the contest. Susie Stephens from Johnson and Johnson opened up the second day with a keynote talk about usability of semantic technologies within the life sciences. Stephen's talk focused on how companies can benefit from Semantic Web technologies, referring to the technologies which are currently being used within Johnson and Johnson.

The subsequent panel discussion, sponsored by the NeOn research project, covered the related topic: "Next steps in intelligent information management in the Pharmaceutical and eHealth domains." Representatives from AstraZeneca, Ontotext and the Spanish health systems discussed the status quo and promising trends within these domains and others.

In parallel to further research presentations, the afternoon of the second conference day included a third panel discussion entitled "Linked Data in the Enterprises." as well as an event organised by the CELTIC Cluster with high-quality presentations from large telecommunication companies. The opening talk of the CELTIC Cluster event was given by the director of EURESCOM, David Kennedy, followed by presentations from Riccardo Pascotto from Deutsche Telekom and José Jiminéz from TID. After the break, the CELTIC event continued with presentations about on-going projects, with representatives from Telecom France, Salzburg Research, Nokia Siemens Networks, University of Vienna, Telefonica and Phillips among others.
The conference concluded with a second matchmaking session, attracting even more participants than the first day and proving to be a beneficial addition to the 2009 program.

In support of business-to-business networking becoming a core theme to this year's program, companies were provided an open demo space for spontaneous business demonstration. Participants who were interested in presenting their business idea to the target audience had the possibility to reserve a presentation slot. A total of 10 demo slots were assigned to different companies throughout the duration of the conference.

Those who were not able to attend the conference onsite had the possibility to follow the program on twitter and facebook. The entire conference was filmed by videolectures.com and the recorded presentations are now available online at: http://videolectures.net/estc09_vienna

Summing up, this year's ESTC showcased innovative ideas and industrial products already bringing semantic solutions to ICT markets, providing key players with new, effective networking opportunities. On behalf of the organisers, many thanks to our supporters and sponsors; we are looking forward with great anticipation to ESTC2010.

For visual impressions of ESTC2009, check out the online gallery:

http://www.estc2009.com/impressions
Details about the ESTC2010, which will again take place in the winter of 2010 in Vienna, Austria, will soon be available on the STI International website: http://www.sti2.org

Lejla Ibralic Halilovic is an Event and Conference Manager at STI International.

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

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)

Important Dates

- **Submission**: 19th of March 2010
- **Notification**: 14th of May 2010
- **Camera Ready**: 11th of June 2010

More Information

For more information please consult the website:
http://www.ontol.inesc-id.pt/ekaw2010.

Topics & Paper Types

EKAW 2010 calls for papers on the following general topics:

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

Further, EKAW 2010 will accept different paper types which will all have the same status in the proceedings:

- Standard research papers
- **In-Use papers**
- **Problem analysis papers**
- **Validation papers**

New!

Keeping up with a LOD of changes

By Michael Hausenblas and Juergen Umbrich, DERI

Linked Open Data has gained tremendous momentum over the past three years. To date, over 100 large-scale datasets such as BBC Music and, just recently, the New York Times dataset have been published, providing several billion RDF triples interlinked by roughly 160 million RDF links.

Many of these datasets, such as DBpedia, have been available from the very beginning of the Linked Open Data movement and undergo changes on the schema and instance level new resources are added, old resources are removed, new links are set to other datasets, and old links are removed as the target has vanished (the 404 issue). These changes may sometimes be planned for example, if new data is available from the origin data source (such as for Eurostat, Semantic Web Dog Food) or stuff gets fixed to increase quality. But the data source may be inherently dynamic think news feeds stemming from Twitter, for example.

The key question is how does a consumer of the dataset, such as another dataset linking to it or an application using it, learn:

Linked Open Data has gained tremendous momentum over the past three years.

1. That something has changed, and
2. What has changed.

The former requires the dataset consumer to either crawl the dataset or receive a notification whenever anything changes. The latter requires a way to describe the actual changes (insert, remove, update) to a dataset. We will consider these aspects of linked datasets as dataset dynamics[1] for this article.

Where all started: Linked Data camp 2009 in Vienna

To start to address issues around dataset dynamics, we kicked-off a small session to discuss these dynamics at last year's Linked Data camp (LDC09) in Vienna[2]. The authors of this article participated remotely. One tangible result of the hacking session was the dataset dynamics demo[3], which is briefly summarised in the following (Fig. 1).

We assume a publisher who has a dataset (upper left corner) and who provides a voiD[4] file (to the upper right) that describes the dataset. We also assume a consumer of this dataset, who needs to keep track of the changes (in the lower half of Fig. 1). The following workflow ensues:

1. The publisher uses the dady (dataset dynamics) vocabulary[5], a voiD extension, to announce the notification mechanism (a slightly extended Atom feed).
2. The consumer invokes the notification mechanism (that is, subscribes to the feed and observes it).
3. For every change event in the feed, the consumer decides if the change is relevant and performs actual updates to its local (RDF) data store.

We learned two lessons from the LDC09 demo. On the one hand, the demo highlighted the benefits of a loosely coupled, highly distributed

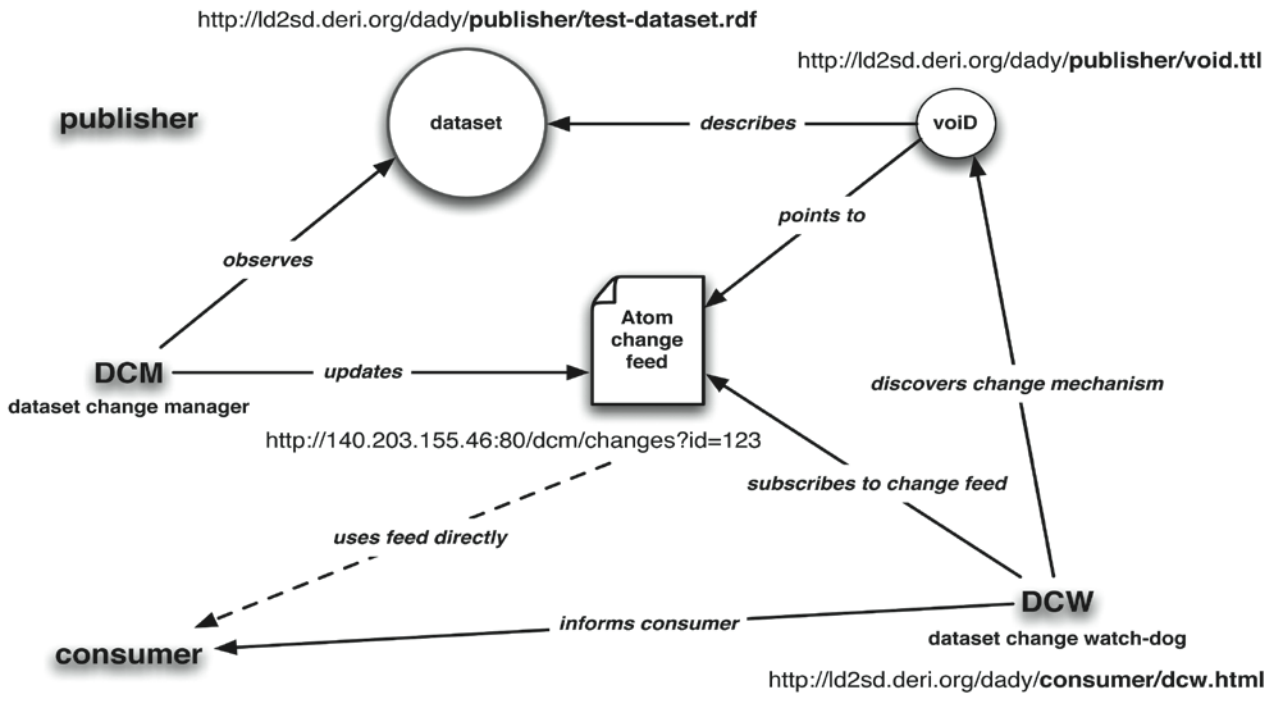


Fig. 1. LDC09 Linked Dataset Dynamics demo.

system. On the other, it revealed where the deeper issues lie: though this setup presumably scales nicely, the burden is still on the consumer to figure out the change semantics.

Summing up, the LDC09 session gave us some interesting insights but also showed us quite plainly that we need to invest much more to provide a truly generic and powerful solution.

We learned two lessons from the LDC09 demo. On the one hand, the demo highlighted the benefits of a loosely coupled, highly distributed system. On the other, it revealed where the deeper issues lie: though this setup presumably scales nicely, the burden is still on the consumer to figure out the change semantics.

Panta rhei …

As a direct result of the LDC09 session, the participants have founded a group[6] that has the goal to "develop a set of technologies to enable Linked Data consumers … to keep track of changes in a Linked Open Dataset." We are currently collecting use cases and identifying potential technologies that can be used. Where appropriate, we will build on existing tools such as Talis' changeset[7] or the DSNotify framework[8].

The problem concerning dynamics may have been identified some 2500 years ago, when Heraclitus coined the phrase panta rhei: "everything flows." However, it has yet to be demonstrated how one can best keep up with a excuse the pun LOD of changes on a global scale. If you're interested in contributing to this effort, you're more than welcome to sign-up for the Dataset Dynamics group, let us know what you think is an important use case as well, or to suggest worthwhile technologies such as protocols or vocabulary.

Michael is a Web of Data researcher, coordinating DERI's Linked Data ResearchCentre, see http://linkeddata.deri.ie/ Jurgen is a second-year PhD student researching processing and querying Linked Data.

1. http://esw.w3.org/topic/DatasetDynamics
2. http://www.linkeddatacamp.org/wiki/LinkedDataCampVienna2009/DatasetDynamics
3. http://code.google.com/p/dady/wiki/Demos
4. http://semanticweb.org/wiki/VoiD
5. http://purl.org/NET/dady#
6. http://groups.google.com/group/dataset-dynamics
7. http://n2.talis.com/wiki/Changeset_Protocol
8. http://dsnotify.org/

Rights Statements on the Web of Data

By Leigh Dodds

Why do we publish open data? It's to allow other people to reuse it; to take it and do creative and innovative things with it. We open data because we want it to be used outside the confines of our own projects, applications and organisations. To achieve that aim we need to do lots of ground work particularly around adopting open formats, and also in ensuring that data is both discoverable and wrapped in services that make it useful. But we also need to clearly communicate our basic intention: that the data is available for reuse. And we need to be clear on what forms of reuse we expect or want to support.

Within the open data movement in general, and the Linked Open Data movement specifically, we're building a commons. An open environment that contains data from a wide variety of different sources that can be meshed together and re-used in a number of powerful ways. We're building the foundations for the next wave of innovative web applications. Ensuring the stability of those foundations involves addressing any rights or legal issues that are going to impact the community of users in the future, and impede innovation and progress.

We're building the foundations for the next wave of innovative web applications.

In the open source and open standards worlds a lot of attention is rightly paid to legal issues, e.g. around patent rights, that may effect the spread of open standards and software. The equivalent area for open data is understanding the rights that are associated with a dataset. Clear, explicit rights statements are a means to achieve that. Very often though, this is a difficult thing to achieve.

In many cases the people involved in the process of opening up data do not have a legal background or framework that can be used to understand the issues of that may impact their efforts. At ISWC 2009, Jordan Hatcher, Kaitlin Thaney, Tom Heath and myself ran a workshop on Legal and Social issues facing data sharing on the web. The goal of that workshop was to help increase understanding in the community about the importance of open data licensing. In this issue of Nodalities each of us has contributed an article on one aspect of that discussion.

It is important to understand that there will never be a single off-the-shelf solution to open licensing. There will be a range of different approaches that are tailored to the needs of a particular community or type of publisher. In some communities, putting all data into the public domain is likely to become the norm; if it's not already. In others very little data may be opened up, and even then its use may be restricted in different ways. It's therefore important to understand what legal tools are suitable in which context. There are a growing number of these tools available, and in his article Jordan Hatcher introduces several of them.

It is important to understand that there will never be a single off-the-shelf solution to open licensing. There will be a range of different approaches that are tailored to the needs of a particular community or type of publisher.

Social norms, e.g. around attribution and citation, are another area in which communities will differ. In her role at the Science Commons project, Kaitlin Thaney has been working closely with the scientific community on issues relating to open data publishing. The norms for strong attribution and sharing used within that scholarly community set a high standard to which those of us working on Linked Open Data should aspire.

We also need to be publishing rights statements in both human and machine-readable formats. To achieve that we need to understand the range of different ways in which data is published and accessed on the Linked Data web, and ensure that data delivered through each mechanism can be properly linked to, or annotated with a rights statement. In his contribution to this themed issue, Tom Heath reviews some of the existing tools available for annotating datasets with machine-readable rights statements.

The norms for strong attribution and sharing used within that scholarly community set a high standard to which those of us working on Linked Open Data should aspire.

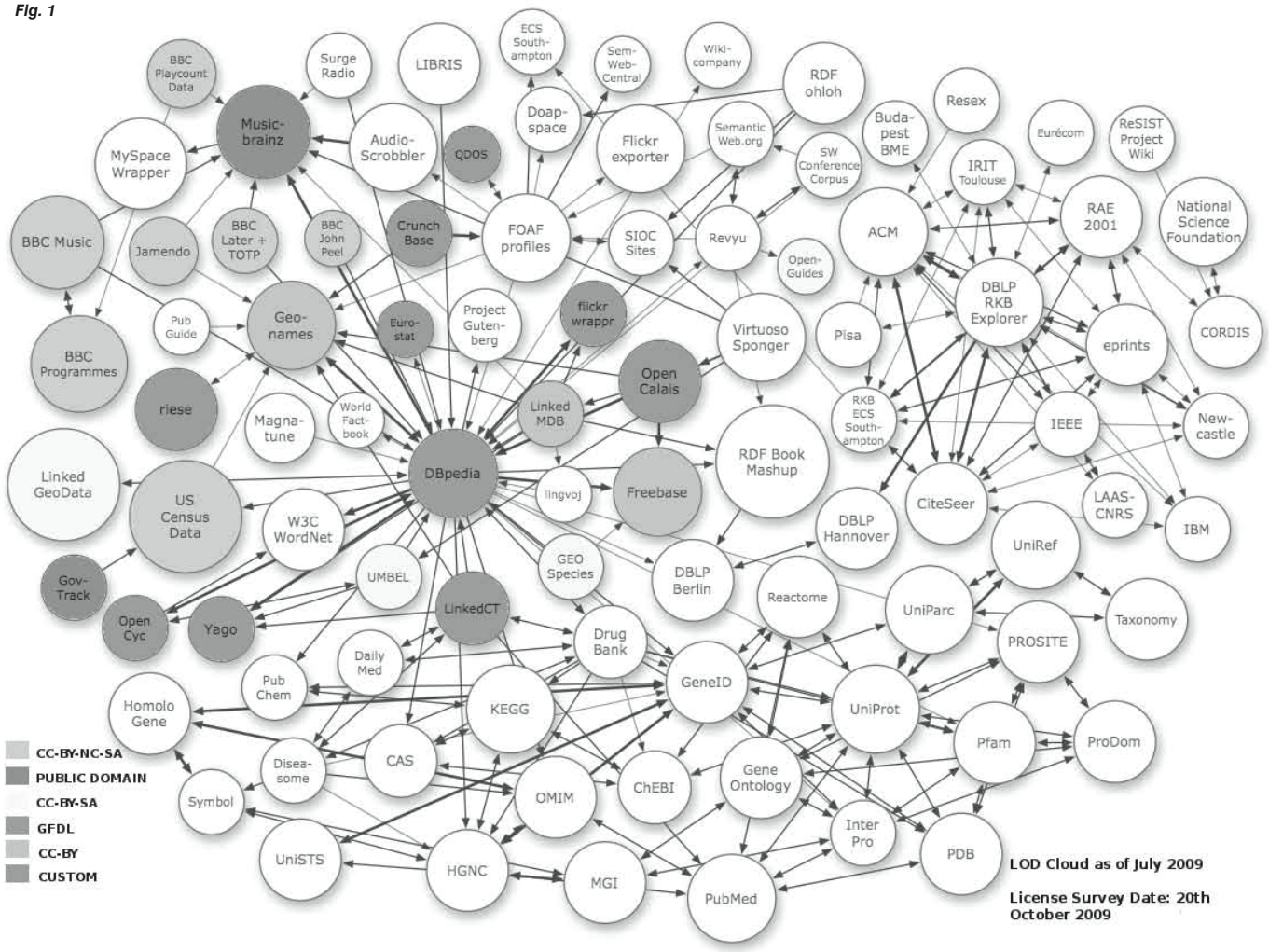
Linked Data also offers its own challenges. Data exposed through a Web 2.0 API is typically done after an explicit agreement has taken place between the producer and the consumer: to access the API you will need to have signed up to gain an API key and agreed to terms and conditions. There is nothing like this in use in Linked Data today; anyone can access any data at any URI. And a Linked Data browser may be simultaneously showing and aggregating data from a number of different sources. We need to think carefully about how usage terms apply in these different contexts, and how those terms can be made explicit to the end user of an application.

As part of the background research for our ISWC tutorial I took a look at the state of licensing across the Linked Data cloud. For every dataset in the March 2009 version of the Linked Open Data cloud diagram I tried to discover two things. Firstly, if any rights statements about reuse had been made available and, secondly, what kind of license was being used (if any). You can see a summary of the results in figure 1. The diagram is a colour-coded version of the original, with each colour representing a different kind of license.

There are a number of conclusions that can be drawn from this analysis:

- The majority of datasets do not have clear licensing terms associated with them. It is therefore unclear as to how reusable the data really is. How do we improve this?
- Creative Commons licenses are well represented. These licenses are certainly easy to understand, and are increasingly common. But are they the best tool for the job?
- Attribution is a strong theme across all licensing schemes. Clearly attribution is an important part of data publishing. How do we best carry out attribution across an increasingly interwoven and interdependent series of datasets?

Fig. 1



Consider these questions as you read the rest of the articles in this issue of Nodalities, and as you return to your own open data practices think how you can start putting this advice to good use in building a strong foundation for the future of the open web.

(http://www.opendatacommons.org/events/iswc-2009-legal-social-sharing-data-web)

Applying Licenses and Waivers to Linked Data

By Tom Heath

The preceding articles in this issue have outlined the key conceptual and legal issues related to licensing data and content, or waiving one's rights to these. Leigh's article has also highlighted the current lack of explicit rights statements among datasets in the Linked Data cloud. This is surprising given the community's roots in Linking Open Data, a project wholly dependent in its early days on openly-licensed data.

My hunch is that the primary reason for this discrepancy is a lack of clarity among Linked Data publishers, not just about the legal aspects of data licensing, but about how to express this information as Linked Data itself. In this article I hope to provide some introductory guidance to publishers of Linked Data on how they can make explicit rights statements about their data. By making explicit rights statements, Linked Data publishers are not constraining how others use their data, but actually freeing them to use it as intended, without fear of negative repercussions due to confusion about rights to use the data. The examples I give below are written in Turtle, which is one of the ways of writing RDF, itself the language of Linked Data. Whatever your degree of technical experience, I hope they will help illustrate the principles discussed in this issue of Nodalities.

```
<http://creativecommons.org/licenses/by-sa/3.0/>
 cc:permits
      cc:DerivativeWorks ,
      cc:Distribution ,
      cc:Reproduction ;
 cc:requires
      cc:Attribution ,
      cc:Notice ,
      cc:ShareAlike ;
 dc:creator <http://creativecommons.org/> ;
 dc:title "Attribution-Share Alike 3.0 Unported" ;
 xhv:alternate
      <http://creativecommons.org/licenses/by-sa/3.0/deed.af> ,
      ...
      <http://creativecommons.org/licenses/by-sa/3.0/rdf> .
```

Figure 1. Creative Commons Attribution-Share Alike license expressed in RDF

Understanding your Content/Data

The first requirement when applying licenses and waivers to a set of content or data is to understand the nature of the material with respect to copyright law. Does it constitute a creative work, such as a blog post or a restaurant review, that falls within the scope of copyright law? Or does it consist of plain facts, such as the latitude and longitude of the same restaurant, over which no-one can claim copyright? The answer to this question will determine which of the examples below are relevant to a particular scenario. In some cases, data that we think of as comprising a single dataset will comprise of copyrightable and non-copyrightable material, and the examples below show how to address this by applying different rights statements to subsets of the data.

```
@prefix rev: <http://purl.org/stuff/rev#> .
@prefix cc: <http://creativecommons.org/ns#> .
<http://revyu.com/reviews/
fa993986b4c04ef89c2d07ff675aa6894cb3f615>
 rdf:type <http://purl.org/stuff/rev#Review> ;
 rdfs:label "Review of The Little Owl, Solihull, Birmingham" ;
 rev:reviewer <http://revyu.com/people/tom> ;
 rev:rating "3" ; rev:text "Not bad for a pub next to the
business park..." ;
 ...
 cc:license <http://creativecommons.org/licenses/by-sa/3.0/>
```

Figure 2. Applying a Creative Commons license to a review of a pub

Copyrightable Material

If your dataset comprises a number of creative works that fall within the scope of copyright law, the Creative Commons provides a range of licenses expressed as human-readable HTML documents and in machine-readable RDF.

Figure 1. shows how the Creative Commons Attribution-Share Alike license is expressed in RDF. Specifically, the statements say that the license, identified by the URI http://creativecommons.org/licenses/by-sa/3.0/ permits reproduction and distribution of the work, as well as the creation of derivative works. In return, the creator of the work must be

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix void: <http://rdfs.org/ns/void#> .
<http://revyu.com/datasets/things>
 rdf:type void:Dataset ;
 dcterms:title "Revyu Things" ; dcterms:description "All the Things in the
 Revyu dataset" ;
 foaf:homepage <http://revyu.com/things/> ;
 void:exampleResource
   <http://revyu.com/things/the-little-owl-solihull-birmingham>
```

Figure 3. voiD description of the Revyu Things dataset

attributed, the licensing terms of the work must be made clear, and derivative works must be licensed under the same or similar terms. The statements at the end of this snippet of code point to versions of the license in alternative languages.

The specific Creative Commons license a data publisher chooses will depend on personal preference and whether there are additional considerations related to derivative works, for example. Once a license has been chosen, and it's corresponding URI identified from the Creative Commons site, this can be applied to the relevant material using a simple RDF statement.

Figure 2. shows a review of a pub near the Talis offices, expressed in RDF, from the Linked Data reviewing and rating site Revyu.com. The review itself is identified by the URI http://revyu.com/reviews/fa993986b4c04ef89c2d07ff675aa6894cb3f615 and as this review constitutes a creative work the CC Attribution-ShareAlike license can be applied by stating that the relationship cc:license exists between the review (identified by its URI) and the license identified by http://creativecommons.org/licenses/by-sa/3.0/

Non-copyrightable Material

The situation with non-copyrightable data is a little more complex, as these tend to come in smaller units (i.e. a single fact) than creative works, which may consist of several chunks of content grouped together. The solution is to apply the same principle to data by grouping individual facts into meaningful chunks and applying rights statements to these. This grouping process is easily achieved using an RDF vocabulary called voiD, which stands for Vocabulary of Interlinked Datasets.

According to the its Web site, voiD is "a vocabulary and a set of instructions that supports the discovery and usage of linked datasets. A linked dataset is a collection of data, published and maintained by a single provider, available as RDF on the Web." In a nutshell we can think of voiD as a set of terms for describing

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix void: <http://rdfs.org/ns/void#> .
@prefix wv: <http://vocab.org/waiver/terms/> .
<http://revyu.com/datasets/things>
 rdf:type void:Dataset ;
 dcterms:title "Revyu Things" ;
 dcterms:description "All the Things in the Revyu dataset" ;
 ...
 wv:waiver
      <http://www.opendatacommons.org/odc-public-domain-
dedication-and-licence/> ;
wv:norms
      <http://www.opendatacommons.org/norms/odc-by-sa/> .
```

Figure 4. Waiver and Norms statements added to the Revyu Things voiD description

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix void: <http://rdfs.org/ns/void#> .
@prefix cc: <http://creativecommons.org/ns#> .
@prefix wv: <http://vocab.org/waiver/terms/> .

<http://revyu.com/datasets/all>
 rdf:type void:Dataset ;
 dcterms:title "Revyu Data" ;
 void:subset <http://revyu.com/datasets/things> ;
.void:subset <http://revyu.com/datasets/reviews> .

<http://revyu.com/datasets/things>
 wv:waiver <http://www.opendatacommons.org/odc-public-domai...> ;
 wv:norms <http://www.opendatacommons.org/norms/odc-by-sa/> .

<http://revyu.com/datasets/reviews>
 cc:license <http://creativecommons.org/licenses/by-sa/3.0/> .
```

Figure 5. Applying licenses and waivers to subsets of data

aspects of data sets. Figure 3. shows a simple voiD description of the Revyu Things dataset, a grouping of all the factual data about things reviewed in the Revyu.com site, such as names, latitude/longitude and homepage addresses. The Revyu Things dataset is identified by the URI http://revyu.com/datasets/things and includes, for example, the resource identified by the URI http://revyu.com/things/the-little-owl-solihull-birmingham

Having grouped items and their associated data together into a void:Dataset, it is trivial to follow patterns similar to those used for applying Creative Commons licenses. Taking a lead from the public domain approach described in Jordan's article, the example in Figure 4. below shows the application of the PDDL waiver, identified by the URI http://www.opendatacommons.org/odc-public-domain-dedication-and-licence/, to the Revyu Things dataset. The CC0 public domain dedication, http://creativecommons.org/publicdomain/zero/1.0/ could also be used in this situation. To indicate the desire for attribution and licensing of all derivative works under the same or similar terms, a norms statement is also added, pointing to the norms definition at http://www.opendatacommons.org/norms/odc-by-sa/.

Tying it all together

Taking each of these individual parts, it is again relatively trivial to tie them together into a high-level voiD description of a dataset, even if that dataset contains subsets that require different licensing terms. The key voiD term to use in this case is void:subset, as shown in Figure 5.

Further Reading

Linked Data and the Public Domain, by Ian Davis http://blogs.talis.com/nodalities/2009/07/linked-data-public-domain.php
voiD Guide - Using the Vocabulary of Interlinked Datasets http://rdfs.org/ns/void-guide

Acknowledgements

Thanks to Leigh Dodds, Ian Davis, and the voiD team, for contributing in various ways to this article, which is based on tutorial slides from ISWC2009: http://tomheath.com/slides/2009-10-washington-licenses-and-waivers-in-practice.pdf

Tom Heath is a resercher for the Talis Platform

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 detailed below.

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

talis® Platform™
data, connected

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
talis.com/platform
nodalities-magazine@talis.com