Exploiting U.K. Government Linked Data

Alasdair Logan

CS5577 Technological, Scientific and Market Research

Literature Review Report

University of Aberdeen.

Department of Computing Science

April 23, 2010

Declaration

No portion of the work contained in this document has been submitted in support of an application for a degree or qualification of this or any other university or other institution of learning. All verbatim extracts have been distinguished by quotation marks, and all sources of information have been specifically acknowledged.

Signed:

Date: April 23, 2010

Abstract

The U.K. government, in a bid to increase transparency and accountability, has created the data.gov.uk Web site. The government is using this site to make over 3,000 different sets of data created by different government departments available to the general public.

The data.gov.uk Web site was designed with Linked Data in mind as all the datasets eventually will be converted into Linked Data. Linked Data is an essential part of the Semantic Web and is a way of sharing and connecting data over the Web.

This report has looked at a number of different tools associated with Linked Data including tools for transforming data into Linked Data, browsing Linked Data and creating mashups with Linked Data.

Contents

Chapter 1

Introduction

The Web has revolutionised the way that information and data is published and shared around the world. Whilst the Web had been designed to allow this data to be easily published and exchanged until recently there has been no attempt at linking the meaning of data together in order to maximise its utility. The adoption of the principles of Linked Data are an attempt to tackle this problem and change the current Web of Documents into a Web of Data.

1.1 Linked Data

In 2006, Sir Tim Berners-Lee first put forward the idea of Linked Data (Berners-Lee, 2006) and it is considered a central part of the evolving Semantic Web. Linked Data is a way of sharing and connecting data over the Web by using Unique Resource Identifiers(URIs) to identify 'things' and storing the data using the principles of the Resource Description Framework(RDF). By using a unique identifier and storing the data in a standard form the interoperability of data on the Web is greatly improved and allows for previously unknown relationships between different datasets to be uncovered.

1.2 Data.gov.uk

The data.gov.uk Web site is an attempt by the government to increase their transparency and accountability by releasing non-personal data that the government has collected.

The Web site follows similar ideas carried out in the United States(data.gov), Australia(data.australia.gov.au) and New Zealand(data.govt.nz) where they act as a central repository for the datasets published by the various departments of the government. The major difference between the U.K. version and other government data sites is the prominence that Linked Data takes in their plans.

Currently the data.gov.uk Web site contains over 3,000 different datasets that the government have produced from transport data to school information. At the moment only 6 of these datasets have been turned into Linked Data and placed behind a SPARQL endpoint but the government aims to convert all their datasets into Linked Data eventually.

This report will focus on three areas centred around the data.gov.uk website and Linked Data. The first section will focus on the issues of data transformation. This will deal with the process data must undergo to be transformed from its raw form into Linked Data.

The next section will involve the process of Linked Data browsing. One of the major issues when dealing with Linked Data is understanding what the data contains and how it can be used. There are currently many tools that allow users to browse and search through Linked Data and then generate SPARQL queries. This section will discuss the approaches and programs currently available in this area.

The final section will tackle mashups which is one of the most popular ways of using Linked Data. This chapter will be a review of the different tools to carry out this task as well as look at some of the different mashups that have already been created with Linked Data.

Chapter 2

Data Transformation

Data transformation is one of the main challenges for the Semantic Web's vision of moving towards a Web of Data. There are currently huge amounts of data in many different formats, stored both online and offline that would add value to the Semantic Web by being converted into Linked Data. It is therefore very important to create tools that can aid with the transformation to Linked Data with these tools needing to cater for both Linked Data experts as well as novice developers. These tools need to allow developers to convert data into RDF quickly and efficiently in order to allow more RDF data to be published on the Web and hopefully creating a network effect for RDF when developers create something new from the Linked Data.

This chapter will discuss the tools currently available that transform data from a number of different formats including spreadsheets, XML and Relational Databases(RDB). All of these tools need a mapping system for converting data to RDF so this chapter will start by looking at the different mapping styles and implementations available.

2.1 Mapping Style

Their are two main types of mapping techniques when converting from structured data to RDF: generic mappings and domain specific mappings.

2.1.1 Generic Mappings

Generic mappings convert the data automatically without specific knowledge of the domain allowing for generic RDF to be created. They often produce RDF graphs of the data that are star shaped where the RDB or spreadsheet table row is a node, the spreadsheet or RDB column is the relation and the cell content is the instance (Han et al., 2010).

Byrne (2010) showed that these automatically generated mappings create simple RDF graphs that do not allow for the capture of complex domain semantics. However, they are a useful starting point to create more customised domain specific mappings with fewer unnecessary triples .

2.1.2 Domain Specific Mappings

Domain specific mappings are harder to implement due to the prior knowledge needed in preparing the mappings but they do generate more relevant RDF with fewer unnecessary triples (Byrne, 2010). Tools like XLWrap and RDF123 both follow domain specific mapping patterns which increase their usability.

2.2 Mapping Implementation

Once a map has been created there are two main ways of exposing the RDF (Sahoo et al., 2009). Extract, Transform, Load or Query Driven Implementation.

2.2.1 Extract, Transform, Load

Extract, Transform, Load (ETL) is a process that first extracts the data from the spreadsheet or RDB, transforms it into RDF and then allows that RDF to be loaded or published on the Web. An advantage of transforming the data using ETL is that it is quicker to access at query time. The flip side of this is that the data is harder to maintain consistency if the original data source changes constantly.

2.2.2 Query Driven

The alternative to ETL is a query driven implementation where a wrapper is used to convert the SPARQL query into a set of commands so that data can be retrieved from the original store. For example, with a RDB, the SPARQL query would be converted into SQL commands which would be run against the RDB. The data would then be retrieved and converted to RDF based on the mapping file. This approach has the opposite positives and negatives compared to ETL as the data is always up to date but querying can take longer.

Han et al. (2010) suggested that for spreadsheets there is no value to using a query driven implementation due to the generally small amounts of data contained in a spreadsheet and generally an ETL implementation is preferred for databases as well.

2.3 Spreadsheet to RDF Tools

One of the most common structures for data storage is a spreadsheet or Comma-Separated Values(CSV) file. With the majority of the data from the data.gov.uk website being stored in spreadsheet or CSV file format and with the popularity of online collaborative spreadsheets like Google Docs increasing, these tools are extremely important for Linked Data conversion. The four tools looked at for this report are Convert To Rdf, RDF123, XLWrap and TopBraid Composer.

2.3.1 Convert To RDF

Convert To RDF (Golbeck et al., 2002) is a GUI based tool for converting CSV's and basic spreadsheets into RDF based on the earlier Excel To RDF tool. The GUI shows the contents of the imported spreadsheet, allowing the user to manually select the cells that are required to be converted into RDF. The selected cells must contain a header row with the names for each column which will be used for the naming conventions of the triples.

Whilst the GUI makes Convert To RDF an easy to use tool it only offers a

generic RDF mapping system allowing for only simple RDF graphs to be created. It also can't handle spreadsheets with multiple sheets. Selecting the cells manually can be arduous if the spreadsheets are sufficiently large and contain many heterogeneous sections as mappings would need to be created for every section. Convert To RDF uses an ETL mapping implementation.

2.3.2 RDF123

RDF123 (Han et al., 2010) is a spreadsheet to RDF tool that improves on the work of Convert To RDF by allowing for more complex mappings to be created. Where Convert To RDF allows mappings that can only create basic RDF graphs, RDF123 allows for the user to create domain specific mappings. RDF123 also allows rows to be converted with slightly different RDF schemas and a more complex naming schema can be adopted.

RDF123 follows on from the principles of the GRDDL syntax by adding metadata to the spreadsheet to be converted. In this metadata, a link to a translation file can be placed so that automatic agents that come across the spreadsheet file can also follow the translation link and convert the file into RDF automatically.The main drawback of RDF123 is that, like Convert To RDF, it is unable to handle spreadsheets with more than one sheet.

2.3.3 XLWrap

XLWrap (Langegger and Wöß, 2009) is another spreadsheet to RDF tool that offers the same advantages as RDF123 in that it allows the user to define domain specific mappings but unlike RDF123 it can understand and convert tables with multiple sheets which is one of the most popular formats for data, especially in the data.gov.uk Web site.

XLWrap converts spreadsheets into RDF by using a mapping file which allows for complex triples to be created programmatically. The mapping file itself is written in the TriG RDF syntax and offers more options for conversion than the other spreadsheet to RDF converters and improves on RDF123 by allowing for

completely different schemas to be assigned to different rows.

XLWrap is bundled with the XLWrap Server which provides a SPARQL endpoint and a Linked Data browser for the triples created. Whilst XLWrap utilises an ETL mapping implementation, the XLWrap Server also monitors any updates made to the files and reconverts to RDF when needed, which saves time at conversion time. XLWrap also allows for files to be converted from multiple files and also from different sources online.

2.3.4 TopBraid Composer

TopBraid Composer is a premium Semantic Web modelling and application development environment from TopQuadrant. The product offers many different services for Semantic Web developers including ontology creation and population, SPARQL querying and importing data from spreadsheets, RDBMS's or XML data. TopBraid offers three versions of its software, a Free Edition, Standard Edition and Maestro Edition. The Free Edition was accessed for this report and only allows for the creation of RDF/OWL files and SPARQL querying. The Standard Edition retails at $1,650 and includes the ability to import data from multiple formats. The Maestro edition retails at $3,000.

TopBraid Composer is an easy to use tool and will be familiar to anyone already familiar with Eclipse. Langegger and Wöß (2009) critically evaluated the spreadsheet converting abilities of the TopBraid Composer stating that the spreadsheet import feature is basic with support for basic mappings similar to ConvertToRDF.

2.4 RDB to RDF Tools

The majority of data currently stored on the Web is stored in Relational Databases(RDB's) (Sahoo et al., 2009) meaning that changing this data from it's traditional format into RDF has been an important task for the Semantic Web community.

2.4.1 Virtuoso RDF View

Virtuoso RDF View is part of OpenLink's Virtuoso Universal Server. RDF Views are used to map relational data into RDF and allow the RDF structure to be customised. It includes a declarative Meta Schema Language to define the dynamic mapping of SQL data to RDF ontologies. RDF View can create basic mappings of databases and takes the table name as an RDFS class. It also understands primary key and foreign key relations and follows a query driven implementation by creating virtual RDF graphs without the need for a physical creation of datasets allowing for the data to be constantly up to date.

2.4.2 D2RQ

D2RQ (Bizer and Seaborne, 2004) is a mapping language for treating relational databases as virtual RDF graphs. The D2RQ platform includes a D2R server and D2RQ Engine which use the mappings to enable applications to access these graphs through APIs and over the Web via SPARQL as Linked Data. D2RQ can be used to create both automatic and domain specific mappings allowing for domain knowledge to be applied to the mappings.

2.4.3 Triplify

Triplify (Auer et al., 2009) is an approach to publish RDF and Linked data from relational databases on the Web by mapping HTTP-URI requests onto relational database queries. Triplify then converts the results from these queries into triples and publishes them. Triplify doesn't define a mapping language but exploits SQL notions with suitable conventions to transform query results into Linked Data.

2.5 XML to RDF Tools

Converting data from XML into RDF is another popular task due to the increasing amount of data stored as XML and the versatility of the format. The two ways of tackling this topic discussed here are GRDDL and XSL Transformations(XSLT).

2.5.1 GRDDL

Gleaning Resource Descriptions from Dialects of Languages (GRDDL) is a specification for introducing markup to a XML document in order to declare that it contains data compatible with RDF and for linking to transformation algorithms for extracting this data from the document (Connolly, 2007). These transformation algorithms are often XSL Transformations but other transformations are available.

2.5.2 XSLT

XSL Transformations(XSLT) is a language for transforming a XML document into other XML documents and is often used to transform XML documents into XHTML to create dynamic web pages. XSLT can also be used to transform XML into RDF/XML by creating a XSLTransformation file.

Both of these techniques can easily be used to convert XML files into RDF which means that anything that can be converted into XML can then be converted to RDF. Only a small proportion of the data from data.gov.uk is currently stored in XML format so having to convert to XML would be an extra and possibly unnecessary step. Any changes made to the original data source would also need to be updated in the XML file making it slightly more complicated task.

Chapter 3

Linked Data Browsing

Making data machine readable is one of the central concepts of the Semantic Web allowing for knowledge to be accessed and utilised by automatic agents. Whilst correctly structured Linked Data satisfies this condition, the RDF itself is not designed to be massively human readable. This is where RDF and Linked Data browsing tools play an important part in the development of the Semantic Web by improving the accessibility of Linked Data allowing the data and links between the data to be explored.

This section of the report will look at faceted browsers, data browsers and RDF graph visualisations.

3.1 Faceted Browsers

Yee et al. (2003) proposed the faceted browser as an alternative way of exploring data compared to the traditional methods of text match searching and browsing through the results sequentially. A faceted browser offers an intuitive interface that allows the search results to be grouped into facets which, when selected, narrow the search results to show only the items that share the attributes of that particular facet.

For example when browsing through a dataset of paintings, a faceted browser could allow the user to view paintings by Impressionist painters. The user could then refine his search by only viewing Impressionist paintings painted by someone from France or paintings that contained a water scene. Grouping by facets in

this way often leads to previously unknown relationships and groups of objects to be discovered.

3.1.1 Sparallax

Sparallax is an open source faceted browsing interface for SPARQL endpoints based on Freebase Parallax (Huynh and Karger, 2009). Sparallax provides a faceted browsing interface and works as a proxy over SPARQL endpoints, converting the users interactions into SPARQL queries.

The faceted browser allows queries to be easily generated over different endpoints, utilising the Linked part of Linked Data. The easy to use interface allows someone with no knowledge of Linked Data or the Semantic Web to use these technologies. However, Sparallax is restricted as only Virtuoso SPARQL endpoints are supported and all properties should explicitly specify their range and domain which limits the usability of this tool. Sparallax can also become rather slow the more complex a query becomes.

3.1.2 /Facet

/Facet is another open source faceted browser for RDF data. /Facet uses RDF files rather than SPARQL endpoints and is designed to handle any RDF schema making up for the limitations of Sparallax. /Facet also allows users to use Linked Data without needing to know any of the underlying principles involved.

Like Sparallax, /Facet slows down as the query becomes larger and more complicated. It also doesn't allow for a traversal of the Linked Data so links to other datasets are not explored.

3.2 Data Browsers

An RDF data browser allows RDF to be accessed with the traditional point and click browsing paradigm. They display RDF in a structured way, where URI's are links that when clicked display relevant data. These browsers access RDF as

either a single RDF file or they browse the Web as an unbounded set of RDF data stores. The first option limits the usability of the RDF as no new links to other datasets can be found but this can be suitable for tasks where the RDF is used in a closed world environment. The open world environment of allowing the browsers to search the whole of the Semantic Web allows can lead to very slow query times due to the vast data being searched.

3.2.1 Disco - Hyperdata Browser

Disco is a hyperdata browser interface for the Semantic Web Client Library (Bizer et al., 2009). The Semantic Web Client Library views the Semantic Web as a single query-able RDF graph. The semantic Web Client Library answers queries by dereferencing HTTP URIs, following rdfs:seeAlso links and using the Sindice search engine by looking for other possible sources for the same URI. The Disco browser acts a simple interface over the top of the Semantic Web Client Library displaying the gathered data including where the data came from.

Whilst Disco allows for searching and browsing through the Semantic Web to find links to datasets and the information they contain, the information returned can't be edited or utilised in any manner.

3.2.2 Tabulator

Tabulator is an RDF browser that treats the Semantic Web as a connected RDF graph and was designed a generic RDF browser for exploring all RDF data (Berners-Lee et al., 2006). Tabulator was originally designed as a simple RDF browser but the most recent version of the Tabulator has included the ability to edit RDF as well as generating specific SPARQL queries (Berners-Lee et al., 2008).

3.2.3 Pubby

Pubby (Cyganiak, 2007) is a Linked Data frontend for SPARQL endpoints. Pubby is a server side and uses a configuration file to translate non dereferencable URI's

by querying the SPARQL endpoint. Pubby can only work with a SPARQL endpoint that can answer DESCRIBE queries and will not be able to support more than one dataset at a time. Pubby is used as the front end for DBPedia the Linked Data version of Wikipedia.

3.3 RDF Graph Visualisations

RDF is represented in a graph form so it is only natural that tools have been created to display RDF as a graph. Schraefel and Karger (2006) state that displaying RDF in its natural graph form has both positive and negative aspects. They say that graphs are good for showing density and clusters of the RDF and are suitable for giving an overview of the RDF data. However, RDF graph visualisations also support limited tasks that aren't really suited for the needs of the Semantic Web. They generally state that graphs are used due to the fact they fit the structure of RDF.

3.3.1 RDF Gravity

RDF Gravity (Goyal and Westenthaler, 2003) is a configurable RDF visualisation tool that provides a wide range of visualising options. These include zooming in on a particular area and filtering out global or local facets to change the views on the graph. RDF Gravity also provides a full text search option for searching over concepts, properties and instances in the RDF file. RDF Gravity also offers the ability to visualise over multiple RDF files but doesn't offer editing services.

3.3.2 ISAViz

IsaViz is a visual authoring tool for RDF offering both searching and browsing of the visualised RDF but also editing of the underlying RDF document. A combined visualising and editing tool can offer the advantage of being able to easily spot incorrectly created triples and edit links between data. ISAViz also allows for easily spotting unknown relations between instances of the RDF but ISAViz is again limited for anything other than viewing and editing the RDF.

Chapter 4

Mashups

Mashups is a loose term for a very important area of the current Web 2.0 and proposed Web 3.0 movement. A mashup can generally be described as a web application that combines data from two or more sources to produce a new service (Hartmann et al., 2008). This data can come from a number of different sources including APIs or raw data.

Linked Data is extremely well suited towards creating mashups due to the structure of RDF data. Tuchinda et al. (2008) state that when creating a data mashup the main stages involved are Data Retrieval, Source Modelling, Data Cleaning, Data Integration and Data Visualisation. Without RDF the two hardest tasks would be Source Modelling, which involves modelling the relations between data sources, and Data Integration, integrating the different data sources together. Both of these tasks are made a lot simpler by RDF as RDF relations between data sources are implicit due to the structure and RDF is designed to allow for interoperability making data integration easy.

This section of the report will look at some existing data mashups currently created from Linked Data and non Linked Data. Some of the tools used to create these mashups will also be evaluated.

4.1 Existing Data Mashups

4.1.1 Research Funding Explorer

The Research Funding Explorer (http://bis.clients.talis.com/) is a joint project between the Department of Business Innovation and Skills (BIS), the Research Councils UK, the Funding Councils, Intellectual Property Office (IPO) and Technology Strategy Board (TSB). It takes Linked Data about networks and patents from these sources and visualises it over a map in order to show clusters of research areas. The goal of the project is to provide possible investors seeking to start a project in a certain research field with a simple map showing pre-existing areas of expertise. This mashup was created using Flash and Google Maps

4.1.2 ITO: Visualising Traffic Data

ITO (http://www.itoworld.com/static/gallerytraffic.html) are a web-based service company specialising in visualising transport data. They were invited to make use of the traffic Linked Data offered by the data.gov.uk Web site. The data contains traffic counts for certain types of vehicles on U.K. roads and they used their own In-house visualisation software to come up with a 3D visualisation tool showing the busiest of the U.K.'s roads. The mashup allows for busier roads to be clearly depicted by colour and a time slider allows for increases and decreases in the traffic through the years to be easily viewed.

4.1.3 Facts about Transportation Energy

This demo (http://data-gov.tw.rpi.edu/demo/linked/demo-10029-fuel.html) from the U.S. data.gov wiki takes Linked Data from two sources and combines it in a Google Motion Chart allowing data from different sources to be investigated and charted against each other. It takes Linked Data about the U.S. budget and from a dataset containing information about governmental fuel consumption by department.

4.1.4 Mapumental

Mapumental is public service mashup produced by mysociety.org and 4IP that doesn't use Linked Data but does use publicly available transport data. It is designed to help the public find the distance that can be traveled in a certain amount of time by using public transport. A time slider allows for a certain amount of minutes to be selected and a OpenStreetMap map lights up the areas that are travel-able in the given time from a defined starting point. This can be used by homeowners to see how long it would take to commute to work each day when purchasing a house.

4.2 Tools

4.2.1 Google Visualization API

Google Visualization API is a popular way of presenting data on the Web. It allows the user to expose their data stored in any Web based data store allowing visualisations to be created from different sources . The API can be used to embed interactive visualisations and can be connected to a live source of data keeping the visualisations will be up to date. The visualisation can be embedded using either JavaScript or simple HTML.

4.2.2 Many Eyes

Many Eyes is a Web based visualisation service that takes data input as spreadsheets or tab-delimited text files and allows the users of the service to create a number of different visualisations from the data. The options for visualisations that Many Eyes offers includes treemaps, bar charts, word maps and pie charts to name but a few and it was designed to use the 'power of human visual intelligence to find patterns' in the data (abm, 2010).

Higgins et al. (2008) used Many Eyes to visualise polio vaccination trends in the United States. They took data from the National Immunisation Survey about the polio vaccinations and used a variety of different Many Eyes visualisations in order to find information about the data. They found the clusters of areas where

children are unvaccinated and also used visualisations to find that more unvaccinated children come from families below the poverty line.

Many Eyes takes its input data as CSVs or spreadsheets, meaning the data would have to be converted from Linked Data into this form.

4.2.3 Exhibit

Exhibit is a framework that allows inexperienced developers the ability to create data visualisations. It is designed to be easy to use with only an HTML file and a data file needed to create advanced data visualisations and mashups It can take information from a number of different input types including JSON and spreadsheets. Exhibit can display the data in many different visualisation including timelines, maps and faceted browsing.

4.2.4 Potluck

Potluck (Huynh et al., 2007) is from the same SIMILE project that created Exhibit and takes data structured for Exhibit as its input. It is a mashup tool designed to allow completely novice developers the ability to generate mashups with no programming ability. Potluck takes two Exhibit data sources as an input and allows the user to combine them using a simple drag and drop interface. Potluck also allows the user to clean up the data and explore the data through faceted browsing.

Chapter 5

Conclusion

This report has given an overview of a number of tools associated with different aspects of Linked Data. Tools for transforming, browsing and mashing up RDF were investigated to see if they have value for using in conjunction data.gov.uk.

When carrying out data transformation tasks there are many suitable tools that can handle these tasks. Due to the nature of the data from data.gov.uk Web site, spreadsheet to RDF tools are the most relevant for this project. Out of the four tools surveyed XLWrap was the most useful by allowing RDF to be easily created from spreadsheet data and it also included a server as well to enable querying and hosting of the data. XLWrap was tested on a spreadsheet from the data.gov.uk Web site and was successfully converted into a basic RDF file.

When it came to Linked Data browsing tools faceted browsers and data browsers both have value for the Linked Data and the data.gov.uk Web site. When data.gov.uk produces more SPARQL endpoints a faceted browser like /Facet would be useful.

RDF graphing tools are useful up to an extent when it comes to Linked Data. They are useful for viewing clusters and density of RDF but they have limited ability above that. Data browsing tools like Tabulator and Pubby have a lot of value for Linked Data by allowing the data to be browsed and new relations in RDF to be discovered.

There are also a good amount of mashup tools available to create new services from RDF. Potluck is a great example of a tool that can allows users to create mashups without having to be aware of any of the underlying technological principles. Tools like this will play an important part in creating the network effect needed for Linked Data to take off properly.

Bibliography

(2010). About many eyes. manyeyes.alphaworks.ibm.com/manyeyes/page/About.html.

Auer, S., Dietzold, S., Lehmann, J., Hellmann, S., and Aumueller, D. (2009). Triplify: light-weight linked data publication from relational databases. In *WWW '09: Proceedings of the 18th international conference on World wide web*, pages 621–630, New York, NY, USA. ACM.

Berners-Lee, T. (2006). Linked data - design issues. www.w3.org/DesignIssues/LinkedData.html.

Berners-Lee, T., Chen, Y., Chilton, L., Connolly, D., Dhanaraj, R., Hollenbach, J., Lerer, A., and Sheets, D. (2006). Tabulator: Exploring and analyzing linked data on the semantic web. http://swui.semanticweb.org/swui06/papers/Berners-Lee/Berners-Lee.pdf.l.

Berners-Lee, T., Hollenbach, J., Lu, K., Presbrey, J., and Schraefel, M. (2008). Tabulator redux: Browsing and writing linked data.

Bizer, C., Gauß, T., Cyganiak, R., and Hartig, O. (2009). Semantic web client library. http://www4.wiwiss.fu-berlin.de/bizer/ng4j/semwebclient/.

Bizer, C. and Seaborne, A. (2004). D2rq treating non-rdf databases as virtual rdf graphs. http://www.w3.org/TR/grddl/.

Byrne, K. (2010). Putting hybrid cultural data on the semantic web. *Journal of Digital Information*, 10(6).

Connolly, D. (2007). Gleaning resource descriptions from dialects of languages (grddl). http://www.w3.org/TR/grddl/.

Cyganiak, R. (2007). About many eyes. www4.wiwiss.fu-berlin.de/pubby/.

Golbeck, J., Grove, M., Parsia, B., Kalyanpur, A., and Hendler, J. (2002). New tools for the semantic web. In *Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web*, volume 2473 of *Lecture Notes in Computer Science*, pages 23–38. Springer.

Goyal, S. and Westenthaler, R. (2003). Rdf gravity (rdf graph visualization tool). `semweb.salzburgresearch.at/apps/rdf-gravity/user\_doc.html`.

Han, L., Finin, T., Parr, C., Sachs, J., and Joshi, A. (2010). Rdf123: From spreadsheets to rdf. In *The Semantic Web - ISWC 2008*, volume 5318 of *Lecture Notes in Computer Science*, pages 451–466. Springer.

Hartmann, B., Doorley, S., and Klemmer, S. R. (2008). Hacking, mashing, gluing: Understanding opportunistic design. In *IEEE Pervasive Computing*, pages 46–54. IEEE Computer Society.

Higgins, R., Li, G., Ahmad, T., and Ytzen, R. (2008). Visualizing polio vaccination trends in the united states:status of the nation. `cluster.ischool.drexel.edu/~cchen/courses/INFO633/07-08/g3.pdf`.

Huynh, D. F. and Karger, D. R. (2009). Parallax and companion: set-based browsing for the data web. `http://davidhuynh.net/media/papers/2009/www2009-parallax.pdf`.

Huynh, D. F., Miller, R. C., and Karger, D. R. (2007). Potluck: Data mash-up tool for casual users. In *The Semantic Web*, volume 4825 of *Lecture Notes in Computer Science*, pages 239–252. Springer.

Langegger, A. and Wöß, W. (2009). Xlwrap querying and integrating arbitrary spreadsheets with sparql. In *The Semantic Web - ISWC 2009*, volume 5823 of *Lecture Notes in Computer Science*, pages 359–374. Springer.

Sahoo, S. S., Halb, W., Hellmann, S., Idehen, K., Jr, T. T., Auer, S., Sequeda, J., and Ezzat, A. (2009). A survey of current approaches for mapping of relational databases to rdf. `www.w3.org/2005/Incubator/rdb2rdf/RDB2RDF_SurveyReport.pdf`.

Schraefel, M. and Karger, D. (2006). The pathetic fallacy of rdf. `eprints.ecs.soton.ac.uk/12911/1/the_pathetic_fallacy_of_rdf-33.html`.

Tuchinda, R., Szekely, P., and Knoblock, C. A. (2008). Building mashups by

example. In *IUI '08: Proceedings of the 13th international conference on Intelligent user interfaces*, pages 139–148, New York, NY, USA. ACM.

Yee, K.-P., Swearingen, K., Li, K., and Hearst, M. (2003). Faceted metadata for image search and browsing. In *Conference on Human Factors in Computing Systems*, pages 401–408. ACM Press.