Linked Data:
thinking big, starting small

Peter Neish
Systems Officer
Victorian Parliamentary Library
peter.neish@parliament.vic.gov.au

Abstract:
The concept of using Linked Data in libraries is gaining momentum; however, there are limited concrete examples that demonstrate the benefits of this approach. This paper explores the use of Linked Data at the Victorian Parliamentary Library, and discusses whether the internal benefits on their own are enough to justify the investment in this new technology.

Introduction

What is Linked Data?

The concepts of Linked Data and the Semantic Web have emerged as the next development of the World Wide Web, and comprise a set of best practices for publishing and connecting structured data on the Web (Berners-Lee 2009). Most content of the web today is linked only in the most basic way, through the use of hyperlinks between pages. Even though this simple linking has been exploited to great effect to build effective search engines and relevance ranking of content, there are huge shortcomings when it comes to identifying what a string of characters actually means. Concepts with the same name cannot be disambiguated easily and the meaning of much data on the web is not readily understood by computers (e.g. numbers, currency, dates and addresses). By implementing Linked Data technology, it is possible to identify things unambiguously and make links from one thing to another. Linked Data essentially involves creating identifiers for things on the Web and then linking these things together, using statements in a standard format called the Resource Description Framework (RDF). This consists of simple 'triple statements' to describe the relationships between things (Figure 1).

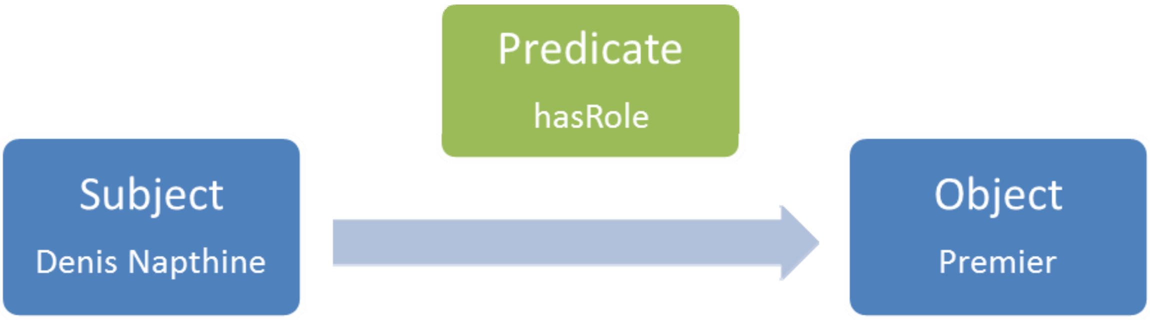


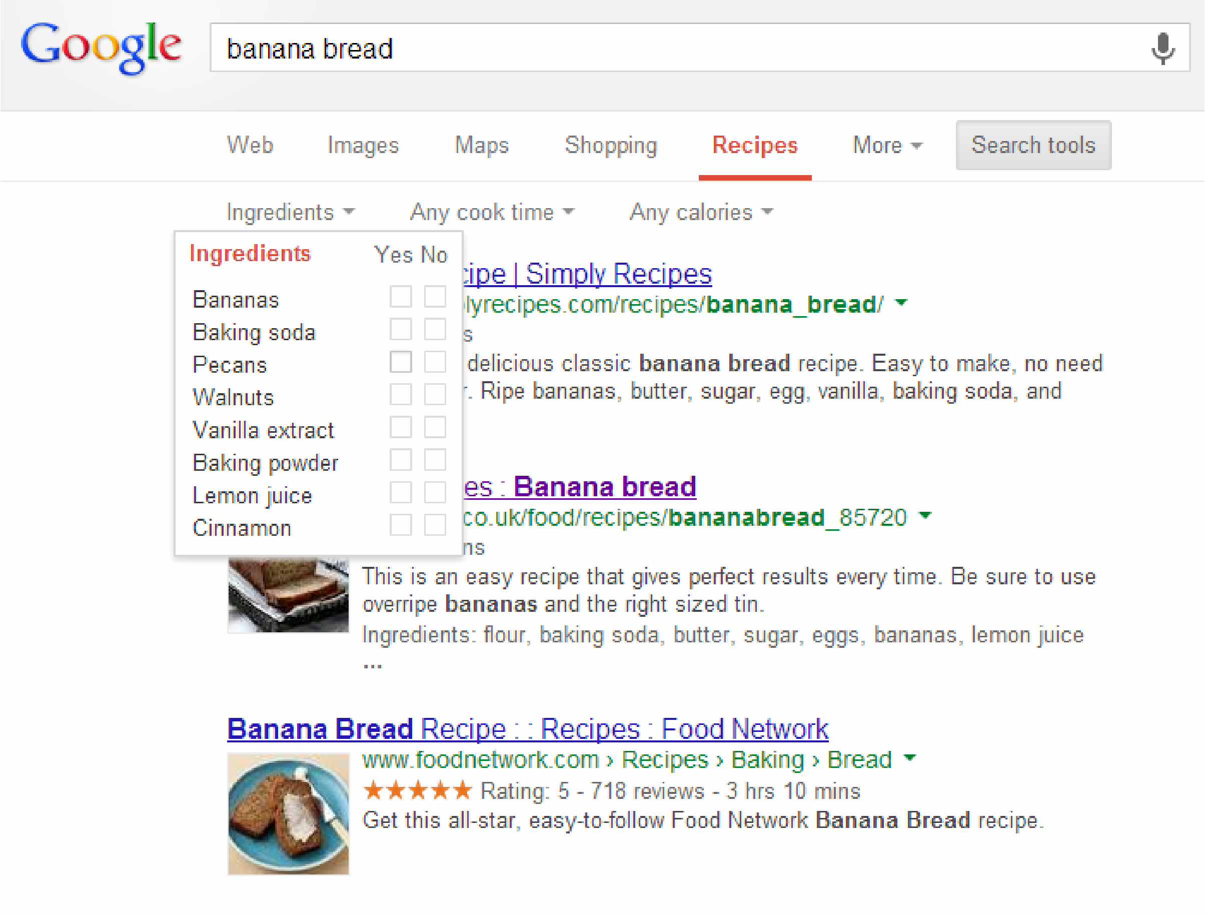
Figure 1 Example Subject Object Predicate triple statement

Simplifying information into these triple statements means that computers can parse and build a graph of how the information is linked. This can lead to some powerful techniques, where computers navigate the graph of information and make inferences about the data by analysing and comparing the data to predefined definitions or ontologies. New linkages and information can be discovered in ways that were not possible before the data was linked.

Examples of Linked Data Initiatives

Despite Linked Data and the Semantic Web having garnered much publicity and support, there are still relatively few concrete examples that demonstrate how implementing Linked Data can benefit an institution. In some respects, there is a chicken-and-egg problem with Linked Data. One of the main benefits of Linked Data is the *Linked* part; however, if there are few other datasets to link to, then this benefit is not realised. Even so, the projects that have implemented Linked Data give a glimpse into what might be possible when a critical mass of data is reached.

One significant driver of semantic technology is search. Three of the major search engine companies (Bing, Google and Yahoo!) joined forces to create Schema.org to provide tools for web content creators to mark up their pages with semantic data. Doing this assists search engines to properly search and categorise not just the web page, but individual data elements on the page. A simple example of this is searching for recipes, where search results can be filtered by meaningful terms like cooking time, calories and ingredients (see Figure 2 below).



banana bread

ipe | Simply Recipes
lyrecipes.com/recipes/**banana_bread**/
s
delicious classic **banana bread** recipe. Easy to make, no need
. Ripe bananas, butter, sugar, egg, vanilla, baking soda, and

es : **Banana bread**
co.uk/food/recipes/**bananabread**_85720
ns
This is an easy recipe that gives perfect results every time. Be sure to use overripe **bananas** and the right sized tin.
Ingredients: flour, baking soda, butter, sugar, eggs, bananas, lemon juice ...

Banana Bread Recipe : : Recipes : Food Network
www.foodnetwork.com › Recipes › Baking › Bread
★★★★★ Rating: 5 - 718 reviews - 3 hrs 10 mins
Get this all-star, easy-to-follow Food Network **Banana Bread** recipe.

Figure 2 Example of semantic searching

This can only be done because individual recipe web pages have been marked up as structured data, with individual elements (ingredients, cooking time, temperature etc.) identified within the page. This allows a search engine to index meaningful semantic information from each page, without having to try to guess what everything means.

Some of the best examples of successful implementations are in the life sciences and medical fields, where genomics, proteomics and other related fields benefit hugely by being able to link data between genes, proteins, medical trials and other data to discover new information from existing bodies of work. In the last few years, a portal to Australia's biodiversity, the Atlas of Living Australia has been built. At the heart of this service has been an effort to combine data about the names and descriptions, images and literature of all life forms in Australia. By creating permanent identifiers for species names and concepts, data about the same species held in different institutions can be combined, without having to resort to error-prone string matching. Also, name changes of organisms can be tracked unambiguously to allow users to find relevant matches. For example, a search on an older name *Eriostemon myoporoides* can still find results, because behind the scenes there will be triple statements pointing from the older name to the accepted name.

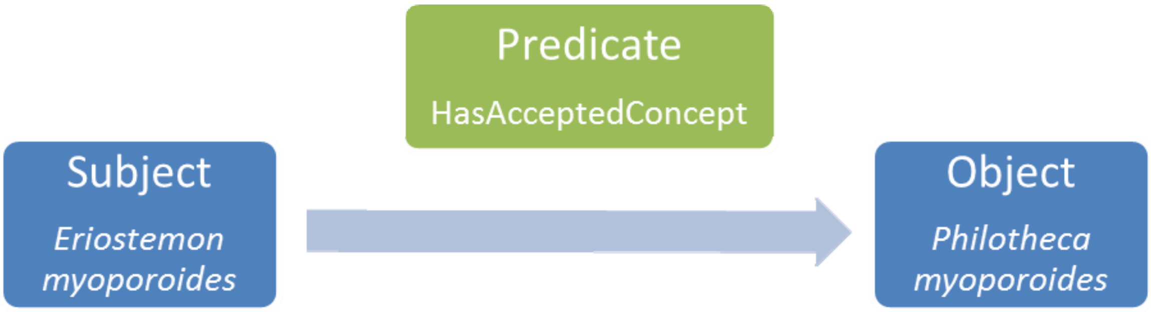


Figure 3 Triple statement relating an old name to a new name in the Atlas of Living Australia

Linked Data in Libraries

Within the library and related cultural institutions there have been some significant initiatives to explore Linked Data (Godby & OCLC Research 2013; Library of Congress 2012). One of the drivers of this change has been the development of the Resource Description and Access (RDA) cataloguing rules and the underlying conceptual model, the Functional Requirements for Bibliographic Records (FRBR). Testing of RDA in the last few years has shown that the flat file MARC21 standard does not accommodate the use of relationships between bibliographic entities that both FRBR and RDA internalise (Coyle 2012). The Bibliographic Framework Initiative (BIBFRAME) is working to provide tools and standards to help the library community transition to Linked Data. An immediate goal is to facilitate a transition from the MARC21 exchange format to a more Web based, Linked Data standard. One of the most challenging aspects of this is dealing with the huge amount of legacy data held in library catalogues, and the transition to future data formats.

In parallel with the BIBFRAME initiative, OCLC and the W3C Schema Bib Extend Community Group have been working toward extending Schema.org to better describe library resources for the purpose of search and discovery. This is a work in progress and is complementary to BIBFRAME. OCLC "envision[s] a model for describing library resources in which key concepts required for discovery are expressed in Schema.org, while the details required for curation and management are expressed in BIBFRAME and associated standards defined in the library community" (Godby & OCLC Research 2013).

While most Linked Data initiatives have been developed within discrete user communities (Life Science, Geology, Archaeology etc.), there are some groups aiming to integrate data across wider domains. Europeana is actively aggregating and sharing data from the entire European cultural sector. In October 2012, they converted a large subset of data into Linked Data and made it available on the web. In Australia, the Humanities Networked Infrastructure (HuNI) project is combining information from 28 of Australia's most significant cultural datasets into a semantic data store. Another prominent group that is actively bringing together members of the GLAM (Galleries, Libraries, Art Galleries and Museums) sector is the LODLAM group. By bringing together members of different communities, several successful pilot projects have been built using Linked Data (Civil War Data 2011; Culture Victoria 2013) and importantly, cross-domain linking of data and idea are being developed.

Linked Data in Parliament and Government

In the last few years, there has been a strong push for governments to publicly release datasets under its control. Under the banner of 'Open Government', many country, state and local government bodies have embraced this initiative. The site datacatalogs.org currently lists 362 data catalogues mainly from government organisations. Although there are more datasets than ever before published, the quality, timeliness and format all vary greatly. Linked Data provides an opportunity to harmonise this data and to create linkages between the disparate datasets. The US data.gov site is experimenting with converting data on its site to RDF, and now has over 6.4 billion triples of open government data available (Anon 2013b). In parallel, there has been a growing number of grass-roots organisations that are working towards making governments more accountable and open. My Society in the United Kingdom, the Sunlight Foundation in the US and the Open Australia Foundation in Australia have been building systems and tools to make releasing and working with government data easier. The Australian Parliament has opened its archive of Hansard (the proceedings of parliamentary debates) and provides this data in a standard format through their ParlInfo service. Importantly, it allows its data to be reused by others, including the Open Australia Foundation, which has incorporated it into its OpenAustralia site.

At the Victorian Parliamentary Library, we have a number of databases that have potential to be integrated using Linked Data. These can be grouped into the following broad themes (Figure 4).

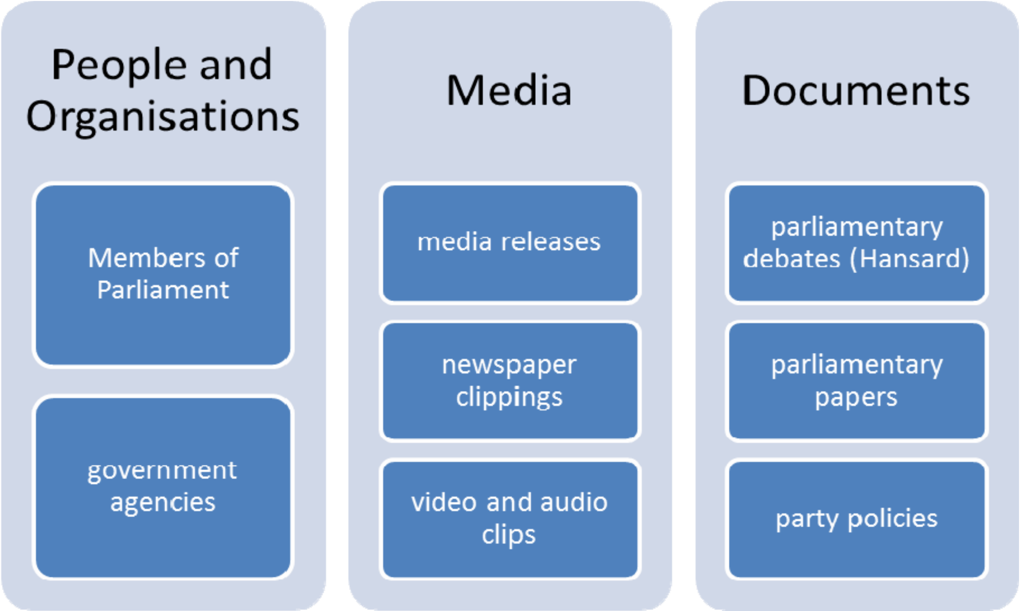


Figure 4 Selected databases at the Parliamentary Library

There is also potential to link to information held by third parties, including elections, electorates, parties and candidates, as well as bills and legislation.

Why Linked Data?

The prime reason for working with Linked Data is to add value to our collections for our own purposes. Much of the justification for using Linked Data comes from the ability to share data or link out to external data sources. However, the scope of our data is necessarily state-based and the opportunities for linking out to other datasets are limited. If we could provide added benefits to other users in sharing our data, then of course that would be a nice side-effect, but is not the prime reason for our interest in Linked Data.

Our main requirement was to create a system where we could query across our databases to discover information that would not be possible in a traditional relational database or federated search. For example, one database records biographical details for each Member of Parliament (MP), while another keeps track of newspaper articles (Figure 4). Although we can easily retrieve a list of newspaper articles mentioning a particular MP, it is quite difficult to return all articles mentioning members of a particular party. The only way to do this at present is to query the members database to discover all the MPs for a particular party and then query the Newspaper database for this set of members. By storing this information as Linked Data, we have the potential to assign properties and classes to MPs that can be used for querying. It would then be straightforward to search for all newspaper articles mentioning a particular party. We could also do more sophisticated queries, for example, 'find me all newspaper articles about MP entitlements that mention MPs who have served in the parliament for more than 10 years'.

Developing the System

The following discussion of our development methodology draws on resources such as freeyourmetadata.org, the Linked Data book and the EUCLID project. These resources have been set up to educate users about Linked Data and give examples and recipes on the steps needed to create and use Linked Data.

The basic workflow to implement Linked Data is as follows:

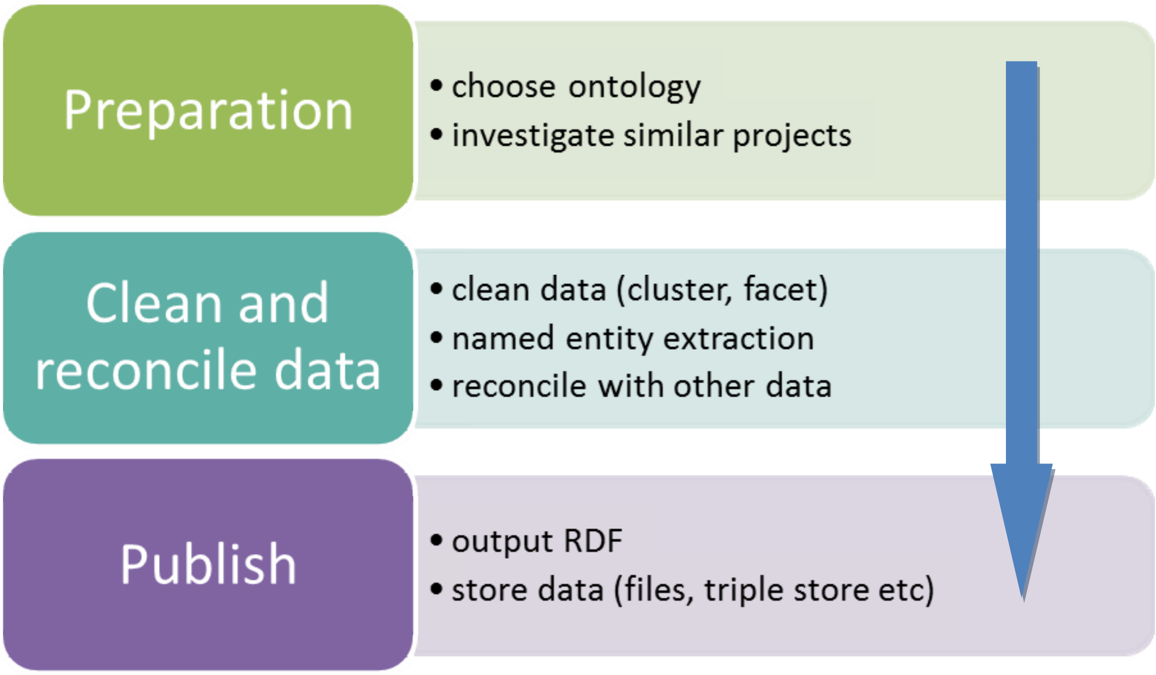

Preparation
• choose ontology
• investigate similar projects
Clean and reconcile data
• clean data (cluster, facet)
• named entity extraction
• reconcile with other data
Publish
• output RDF
• store data (files, triple store etc)

Figure 5 Workflow for creating Linked Data

Choosing a Vocabulary or Ontology

Although this is listed first, investigation of vocabulary terms could be done after data is cleaned and reconciled. The advantage of doing it first is that the process of examining ontologies can give you an insight into common patterns of publishing Linked Data and might lead to the discovery of similar projects.

Choosing a vocabulary involves working out what formal definitions exist for your data items. These formal definitions are defined as a vocabulary or ontology: an ontology is a formal specification of the kinds of things and relationships that can exist within a domain (Anon 2013a). Ontologies are usually specified in a web ontology language such as OWL. The ontology used could be an embedded ontology that is only used within a system and is not designed to be applicable in external situations. Alternatively, terms from existing community-developed ontologies can be used. Each approach has its problems.

With a community-developed ontology, there is no guarantee that the meaning of terms will remain the same. Therefore choosing a stable and widely-used ontology is critical, and is usually used for data elements that are widespread, well understood and unlikely to change (Dublin Core is a good example of a stable community-developed ontology). You are also restricted to the terms and definitions already defined, which may or may not suit your own situation. The big advantage, however, is that your data is immediately aligned with other datasets that use the same ontology, and no mapping or transformation is required to deliver your data. The alternative is to use an embedded ontology that is customised to your internal system. This has the advantage that it can be flexible and support new requirements and iterations. However, the problem with this approach is term proliferation and the additional work needed if you want to map or integrate your data with other systems. Rogers (2013) discusses embedded versus community ontologies and the reasons the BBC chose an embedded ontology for their Linked Data platform. In summary, if the primary goal of using Linked Data is to build a closed system for working with your own data, then an embedded ontology is entirely appropriate. However, if your goal is to link out to existing data in the wider Linked Data ecosystem, then community-developed ontologies are more appropriate.

If you choose to re-use existing vocabularies or ontologies, it can be difficult to discover the best standard for your system, because there are so many standards to choose from (for a visualisation of the large number of metadata standards see Riley 2009).

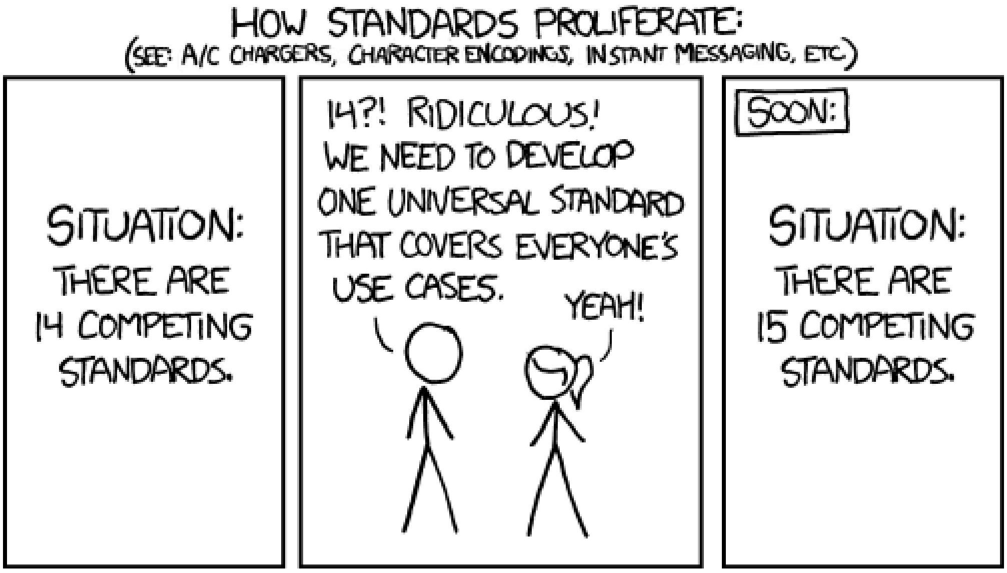

HOW STANDARDS PROLIFERATE:
(SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC)
SITUATION: THERE ARE 14 COMPETING STANDARDS.
14?! RIDICULOUS! WE NEED TO DEVELOP ONE UNIVERSAL STANDARD THAT COVERS EVERYONE'S USE CASES.
YEAH!
SOON:
SITUATION: THERE ARE 15 COMPETING STANDARDS.

Figure 6 Why standards proliferate (source: xkcd.com/927)

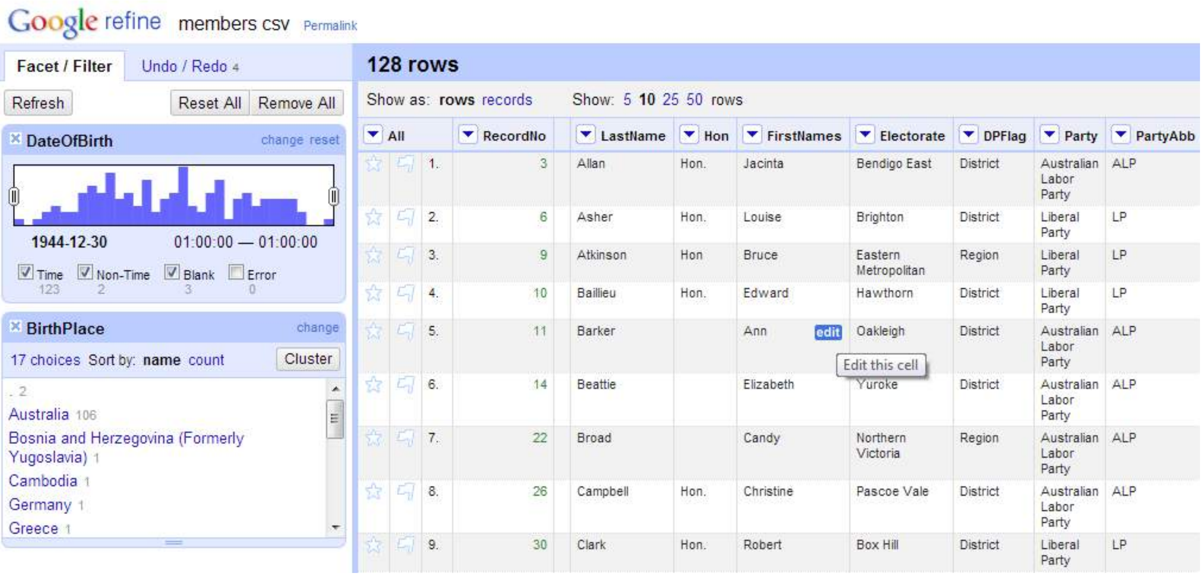
Fortunately, numerous registries and tools have been created to enable discovery of relevant vocabularies. The Popolo project, a community-developed ontology for international open government data specifications, made use of the following resources to discover candidate terms to reuse in its ontology.

- Linked Open Vocabularies by the Open Knowledge Foundation allows you to search and explore over 350 RDF vocabularies by theme and term.
- The Interoperability Solutions for European Public Administrations (ISA) programme of the European Commission lists semantic assets (domain models, ontologies, thesauruses and code lists) from more than 20 online repositories.
- Various ontology search engines exist, including Watson, Swoogle and FalconS.

The Popolo ontology itself is directly applicable to parliamentary information and was a likely candidate for use within the Parliamentary Library. The main drawback was that it had not been fully developed and is currently undergoing rapid change. Despite this, or rather because of this, it was seen as a relevant ontology to use and provided an opportunity to test the ontology with some real data. We were able to provide feedback to the Popolo community on how well the ontology represented parliamentary information in the Victorian context and suggest improvements to the ontology. The Popolo ontology favours reuse over novelty, which means that the majority of its terms reuse existing well-known ontologies.

Cleaning Data

An optional, but highly recommended, step in creating Linked Data is cleaning up your data. This may include fixing problems such as misspellings, duplications, formatting inconsistencies or missing data. The tool of choice for this is Open Refine (previously Google Refine), which allows large amounts of data to be processed in a browser-based interface. By using various clustering, faceting and batch editing tools, it is possible to find duplicates and clean up messy data (Figure 7). If more complex data processing is needed, Open Refine includes the General Refine Expression Language (GREL), which is quite powerful at manipulating data. Importantly, Open Refine allows you to record every step that you take, so that you are able to replicate the same sequence of steps to clean up any subsequent batches of data.



. 2
Australia 106
Bosnia and Herzegovina (Formerly Yugoslavia) 1
Cambodia 1
Germany 1
Greece 1

All	RecordNo	LastName	Hon	FirstNames	Electorate	DPFlag	Party	PartyAbb
1.	3	Allan	Hon.	Jacinta	Bendigo East	District	Australian Labor Party	ALP
2.	6	Asher	Hon.	Louise	Brighton	District	Liberal Party	LP
3.	9	Atkinson	Hon	Bruce	Eastern Metropolitan	Region	Liberal Party	LP
4.	10	Baillieu	Hon.	Edward	Hawthorn	District	Liberal Party	LP
5.	11	Barker		Ann	Oakleigh	District	Australian Labor Party	ALP
6.	14	Beattie		Elizabeth	Yuroke	District	Australian Labor Party	ALP
7.	22	Broad		Candy	Northern Victoria	Region	Australian Labor Party	ALP
8.	26	Campbell	Hon.	Christine	Pascoe Vale	District	Australian Labor Party	ALP
9.	30	Clark	Hon.	Robert	Box Hill	District	Liberal Party	LP

Figure 7 Open Refine screenshot (note facets on left)

Named Entity Extraction

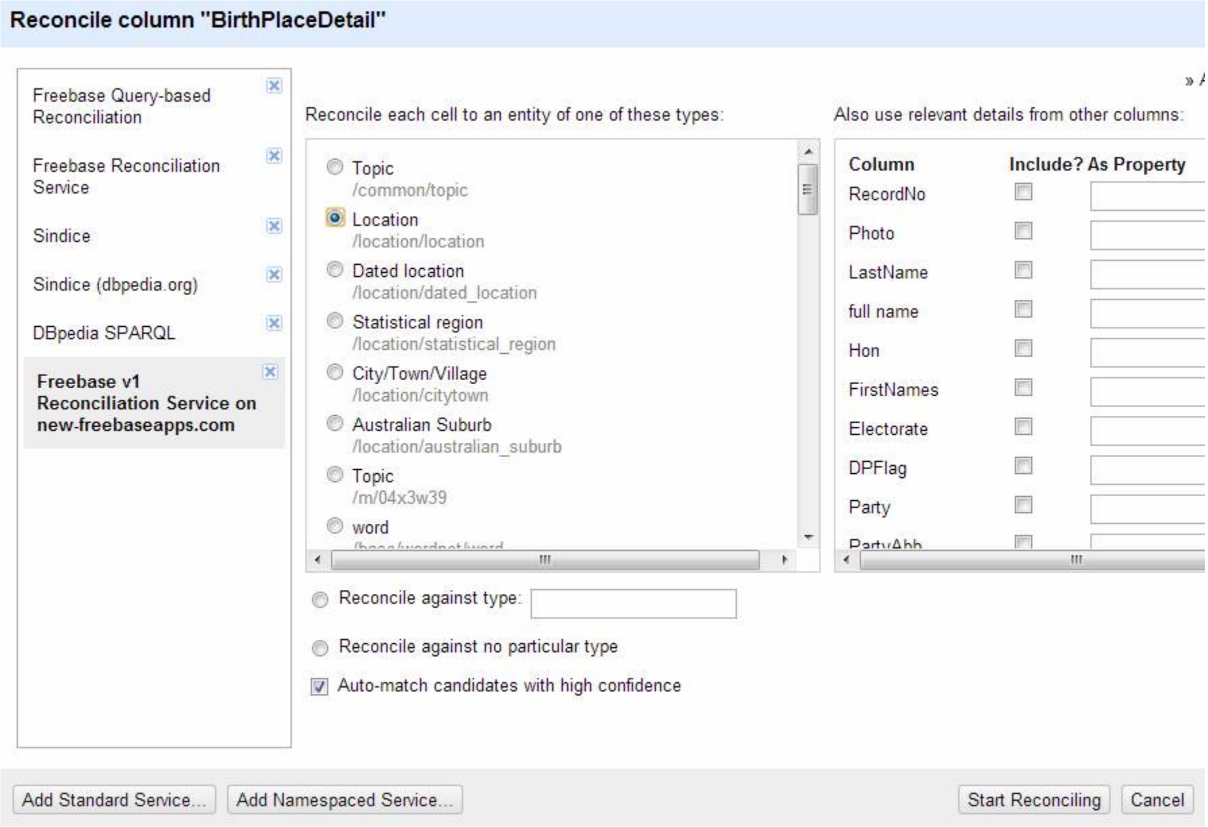
Another optional step, but one that can provide additional value, is extracting named entities using third party APIs. Named entities include such things as people, organisations, events and places. The Free Your Metadata group has released an Open Refine extension that allows you to use three different services out-of-the-box for this purpose: Alchemy API, DBpedia Spotlight, and Zemantia. Results were impressive and can be quantified using the F1 score, which is a measure of a test's accuracy. Alchemy returned an F1 score of 40% and Zemantia 60%, with almost all results being highly relevant (van Hooland et al. n.d.). This tool allows highly relevant entities to be extracted and potentially linked to other data sources for no cost.

Reconcile Data

Once data has been cleaned up, it can be reconciled against external data sources. This is where we take a data element and link it to an existing entity. For example, we might have identified the person Justin Madden and we can link it to the URI in DBpedia: http://dbpedia.org/resource/Justin_Madden. By asserting that the database record we have is linked to the DBpedia entry, we can now infer any of the information that is present in the DBpedia data.

Open Refine can reconcile data against a large number of services, including Freebase, taxonomic databases and organisation information. It can also reconcile against any triple store that exposes its data as a service that can be queried using SPARQL (SPARQL Protocol and RDF Query Language). The complexities are hidden from the user as the user interface takes care of querying the external service and importing the reconciled terms.

For example, we can reconcile the birthplace of Members of Parliament against Freebase (a community-curated database of well-known people, places and things). First, we select the Location type in Freebase to match against and start reconciling (Figure 8 Open Refine reconciliation options).



Reconcile column "BirthPlaceDetail"

Freebase Query-based Reconciliation
Freebase Reconciliation Service
Sindice
Sindice (dbpedia.org)
DBpedia SPARQL
Freebase v1 Reconciliation Service on new-freebaseapps.com

Reconcile each cell to an entity of one of these types:

○ Topic /common/topic
● Location /location/location
○ Dated location /location/dated_location
○ Statistical region /location/statistical_region
○ City/Town/Village /location/citytown
○ Australian Suburb /location/australian_suburb
○ Topic /m/04x3w39
○ word

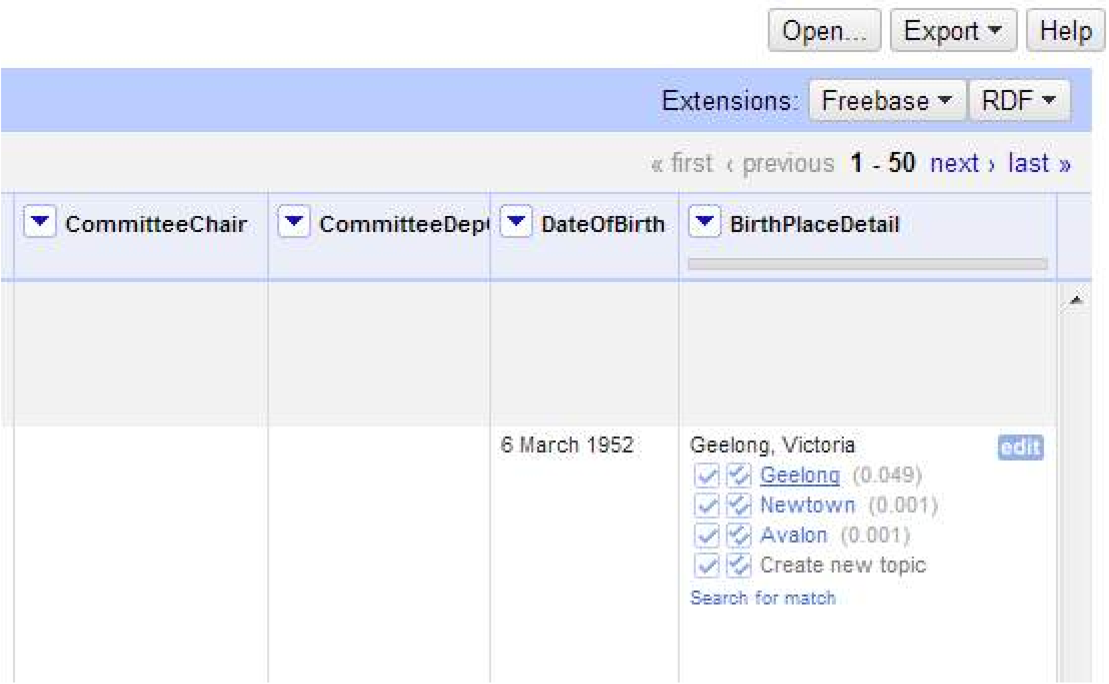
○ Reconcile against type:
○ Reconcile against no particular type
☑ Auto-match candidates with high confidence

Also use relevant details from other columns:

Column	Include?	As Property
RecordNo	□	
Photo	□	
LastName	□	
full name	□	
Hon	□	
FirstNames	□	
Electorate	□	
DPFlag	□	
Party	□	
PartyAbb	□	

Figure 8 Open Refine reconciliation options

If an item cannot be matched exactly or if there is some ambiguity, the user needs to select the correct match. For example, in Figure 8 the user needs to decide which match is correct for "Geelong, Victoria". A hyperlink is provided that brings up the relevant web page for that match to help with the matching process.



CommitteeChair | CommitteeDep | DateOfBirth | BirthPlaceDetail

6 March 1952

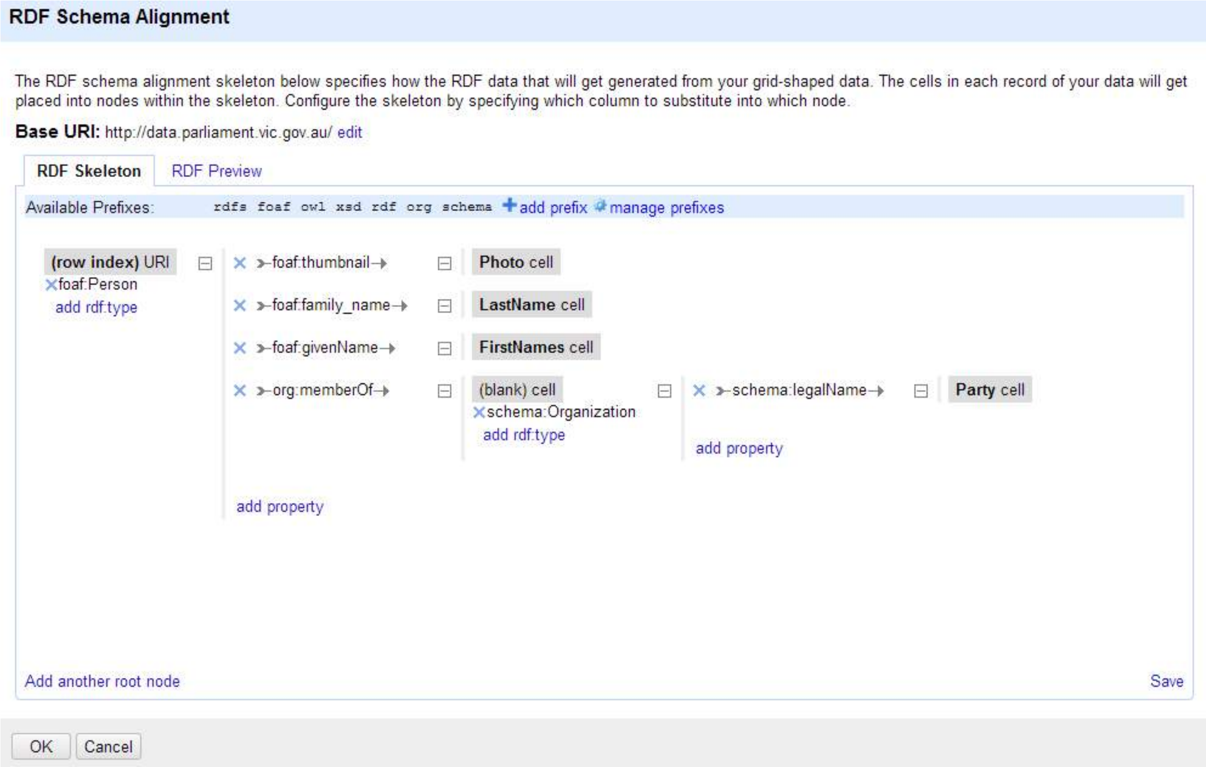
Geelong, Victoria

☑ ☑ Geelong (0.049)
☑ ☑ Newtown (0.001)
☑ ☑ Avalon (0.001)
☑ ☑ Create new topic

Figure 9 Choosing correct match in Open Refine

Output RDF

Again, Open Refine can be used here. By using the RDF refine extension, it is possible to define an RDF template that then aligns values from the data to defined Linked Data vocabulary terms (Figure 10).



RDF Schema Alignment

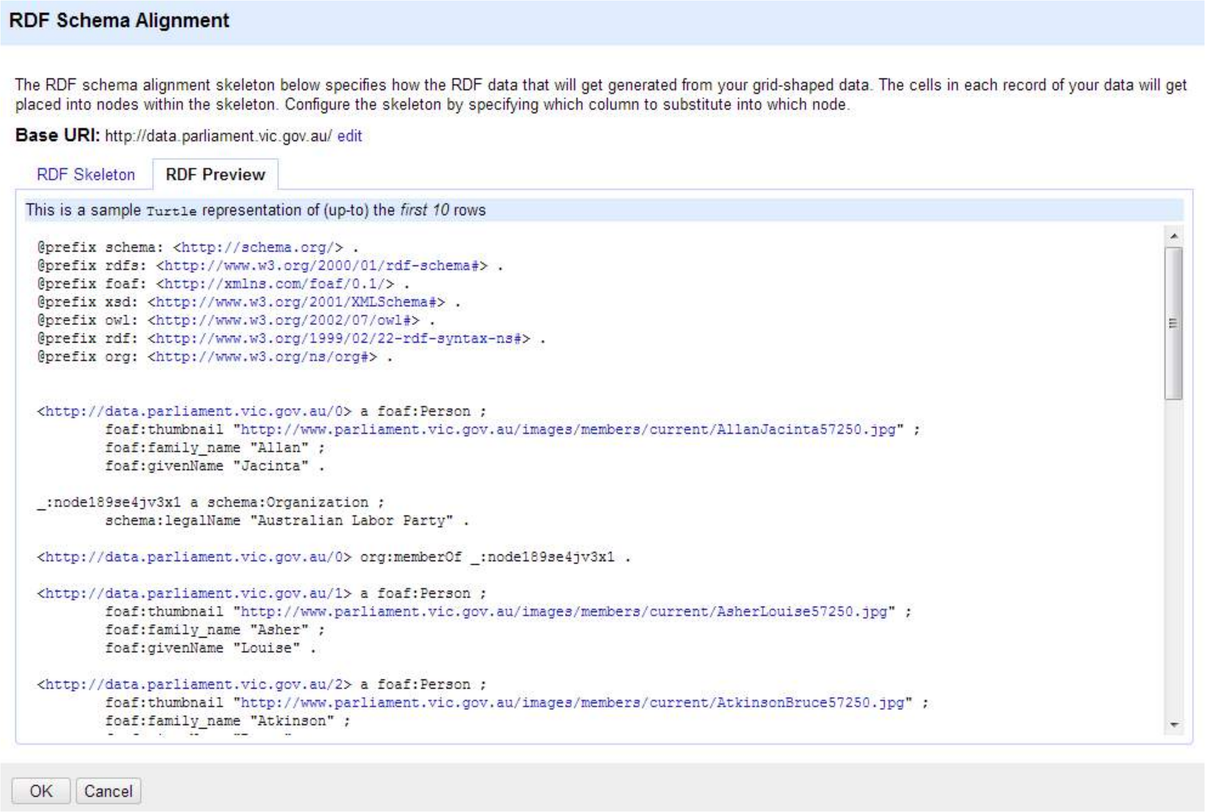
The RDF schema alignment skeleton below specifies how the RDF data that will get generated from your grid-shaped data. The cells in each record of your data will get placed into nodes within the skeleton. Configure the skeleton by specifying which column to substitute into which node.

Base URI: http://data.parliament.vic.gov.au/ edit

RDF Skeleton
RDF Preview
Available Prefixes: rdfs foaf owl xsd rdf org schema +add prefix manage prefixes
(row index) URI
×foaf:Person
add rdf:type
× foaf:thumbnail→ Photo cell
× foaf:family_name→ LastName cell
× foaf:givenName→ FirstNames cell
× org:memberOf→ (blank) cell
×schema:Organization
add rdf:type
× schema:legalName→ Party cell
add property
add property
Add another root node
Save
OK
Cancel

Figure 10 Editing RDF template

The template system is quite powerful and can make use of GREL expressions. Once a template has been configured, data can be exported to RDF files ready to be published either as static files, or for loading into an RDF database. At any time while developing the mapping, the template can be previewed with real data from your table (Figure 11).



RDF Schema Alignment

The RDF schema alignment skeleton below specifies how the RDF data that will get generated from your grid-shaped data. The cells in each record of your data will get placed into nodes within the skeleton. Configure the skeleton by specifying which column to substitute into which node.

Base URI: http://data.parliament.vic.gov.au/ edit

This is a sample `Turtle` representation of (up-to) the *first 10* rows

```
@prefix schema: <http://schema.org/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix org: <http://www.w3.org/ns/org#> .


<http://data.parliament.vic.gov.au/0> a foaf:Person ;
        foaf:thumbnail "http://www.parliament.vic.gov.au/images/members/current/AllanJacinta57250.jpg" ;
        foaf:family_name "Allan" ;
        foaf:givenName "Jacinta" .

_:node189se4jv3x1 a schema:Organization ;
        schema:legalName "Australian Labor Party" .

<http://data.parliament.vic.gov.au/0> org:memberOf _:node189se4jv3x1 .

<http://data.parliament.vic.gov.au/1> a foaf:Person ;
        foaf:thumbnail "http://www.parliament.vic.gov.au/images/members/current/AsherLouise57250.jpg" ;
        foaf:family_name "Asher" ;
        foaf:givenName "Louise" .

<http://data.parliament.vic.gov.au/2> a foaf:Person ;
        foaf:thumbnail "http://www.parliament.vic.gov.au/images/members/current/AtkinsonBruce57250.jpg" ;
        foaf:family_name "Atkinson" ;
```

Figure 11 RDF preview in Open Refine

Publishing Data

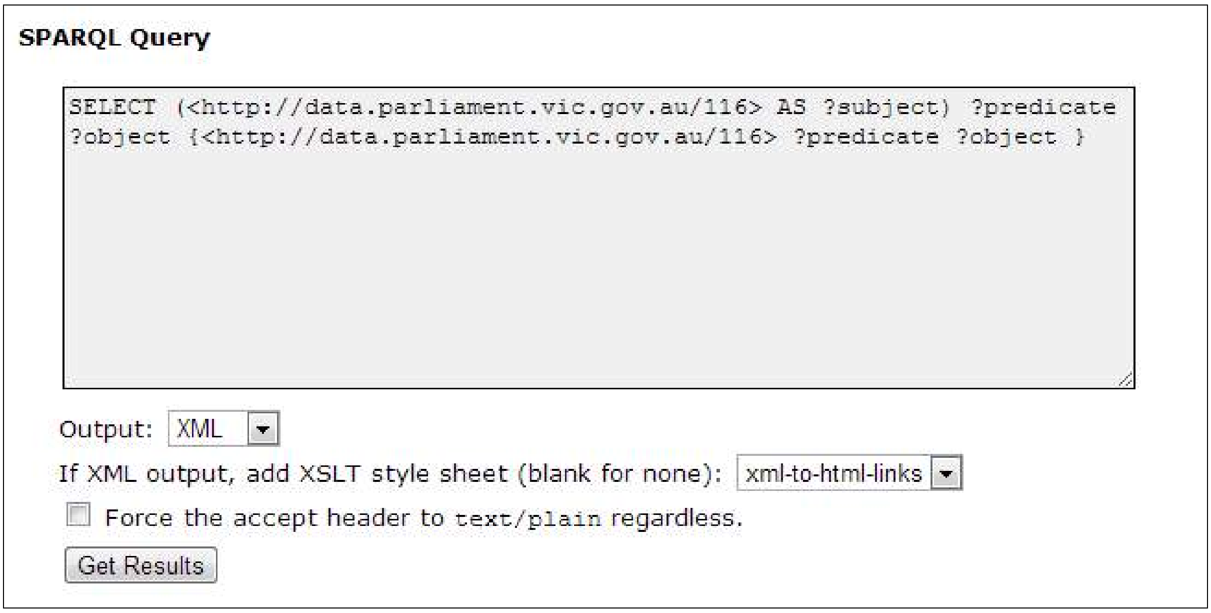
There are numerous patterns for publishing Linked Data, and which you choose depends on how the data is stored and also on how you plan to consume the data. The main patterns are serving Linked Data from:

- static RDF Files
- RDF Embedded in HTML Files or linked from HTML files
- Mapping Relational Database
- RDF Triple Store.

The Linked Data Book has detailed information and recipes for how to implement these patterns. If your data is stored in a relational database, then the easiest way would be to use a platform such as D2RQ, which allows you to map your relational data to Linked Data terms and output Linked Data.

At the Parliamentary Library, we were mainly interested in using the data for querying within the library, and so we chose to store the data in a Fuseki / TDB RDF triple store.

Setting up Fuseki is very quick, and once running, the RDF data can be easily uploaded through a web interface. Fuseki provides a simple SPARQL query interface to interrogate the data (Figure 12).



SPARQL Query

```
SELECT (<http://data.parliament.vic.gov.au/116> AS ?subject) ?predicate
?object {<http://data.parliament.vic.gov.au/116> ?predicate ?object }
```

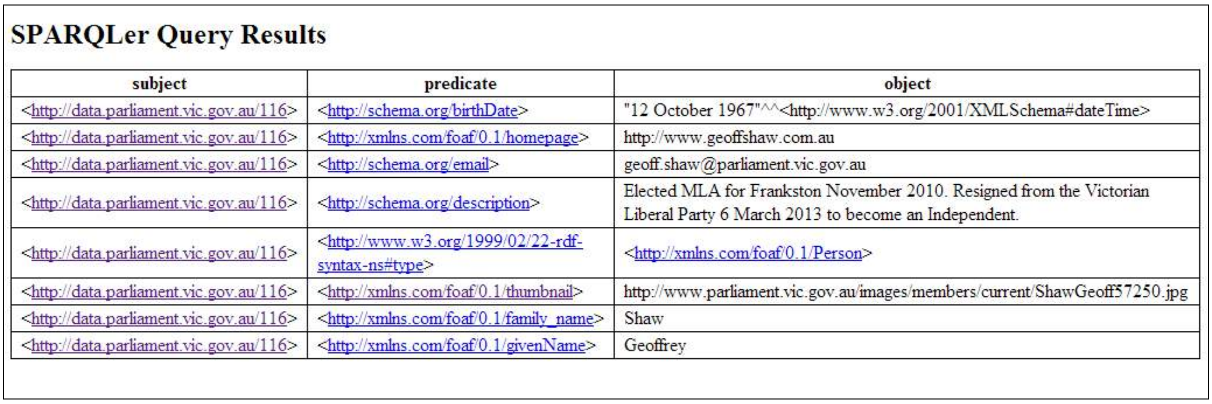
Output: XML

If XML output, add XSLT style sheet (blank for none): xml-to-html-links

□ Force the accept header to text/plain regardless.

Figure 12 Fuseki query interface

And results to queries are displayed in the browser (Figure 13).



SPARQLer Query Results

subject	predicate	object
<http://data.parliament.vic.gov.au/116>	<http://schema.org/birthDate>	"12 October 1967"^^<http://www.w3.org/2001/XMLSchema#dateTime>
<http://data.parliament.vic.gov.au/116>	<http://xmlns.com/foaf/0.1/homepage>	http://www.geoffshaw.com.au
<http://data.parliament.vic.gov.au/116>	<http://schema.org/email>	geoff.shaw@parliament.vic.gov.au
<http://data.parliament.vic.gov.au/116>	<http://schema.org/description>	Elected MLA for Frankston November 2010. Resigned from the Victorian Liberal Party 6 March 2013 to become an Independent.
<http://data.parliament.vic.gov.au/116>	<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>	<http://xmlns.com/foaf/0.1/Person>
<http://data.parliament.vic.gov.au/116>	<http://xmlns.com/foaf/0.1/thumbnail>	http://www.parliament.vic.gov.au/images/members/current/ShawGeoff57250.jpg
<http://data.parliament.vic.gov.au/116>	<http://xmlns.com/foaf/0.1/family_name>	Shaw
<http://data.parliament.vic.gov.au/116>	<http://xmlns.com/foaf/0.1/givenName>	Geoffrey

Figure 13 Fuseki results

Additionally, we also wanted to be able to generate linkages on the fly for web applications, so we plan to embed Linked Data terms within HTML using RDFa. This will allow the discovery of meaningful metadata by RDFa aware agents such as conventional search robots as well as semantic web search engines.

Conclusion

While not complete, our investigations are showing that Linked Data is providing benefits to the Parliamentary Library.

The actual process of investigating Linked Data has forced us to think critically about our data and systems. Investigating related systems, ontologies and vocabularies has given us ideas on how to structure our data and revealed possibilities for linking to other datasets. It has also allowed us to discover other initiatives that are working with Linked Data and provide an opportunity to contribute to these initiatives. In particular, the Popolo Project is emerging as a standard way of representing parliamentary information, and our aim is to make our data conform to this specification. Aligning with the Popolo Project also means we can potentially make use of open source systems already developed for working with government information.

We have also discovered the power of tools such as Open Refine to clean and reconcile our data. This has added to the value of our data regardless of whether it is being shared externally.

Finally, by having data aligned with a standard ontology and stored in a triple store, we are able to query our data in ways that were impossible before.

References

Anon, 2013a. Ontology (information science). *Wikipedia, the free encyclopedia*. Available at: http://en.wikipedia.org/wiki/Ontology_(information_science) [Accessed 26 September, 2013].

Anon, 2013b. Semantic Developer | Data.gov. Available at: http://www.data.gov/developers/page/semantic-web [Accessed 25 September, 2013].

Berners-Lee, T., 2009. Linked Data-The Story So Far. *International Journal on Semantic Web and Information Systems*, 5(3), pp.1–22.

Civil War Data, 2011. Civil War Data 150 | Linking Civil War Data Across State and Federal Archives and Libraries. Available at: http://www.civilwardata150.net/ [Accessed 25 September, 2013].

Coyle, K., 2012. Chapter 1: Introduction. *Library Technology Reports*, 48(4), pp.6–9.

Culture Victoria, 2013. Linking History, an exercise in linked open data | Culture Victoria News. Available at: http://blogs.cv.vic.gov.au/news/linking-history-an-exercise-in-linked-open-data/ [Accessed 25 September, 2013].

Godby, C.J. & OCLC Research, 2013. *The relationship between BIBFRAME and the Schema.org "Bib Extensions" Model a working paper*, Dublin, Ohio: OCLC Research. Available at: http://www.oclc.org/content/dam/research/publications/library/2013/2013-05.pdf [Accessed 20 September, 2013].

Van Hooland, S. et al., Named-Entity Recognition: A Gateway Drug for Cultural Heritage Collections to the Linked Data Cloud? Available at: http://freeyourmetadata.org/publications/named-entity-recognition.pdf [Accessed 26 September, 2013].

Library of Congress, 2012. Bibliographic Framework as a Web of Data: Linked Data Model and Supporting Services - marcld-report-11-21-2012.pdf. Available at: http://www.loc.gov/bibframe/pdf/marcld-report-11-21-2012.pdf [Accessed 20 September, 2013].

Riley, J., 2009. Seeing Standards. Available at: http://www.dlib.indiana.edu/~jenlrile/metadatamap/ [Accessed 26 September, 2013].

Rogers, D., 2013. Ontologies in software: A conflict of interest? *Dave's Blog*. Available at: http://daverog.wordpress.com/2013/05/30/ontologies-in-software-a-conflict-of-interest/ [Accessed 20 September, 2013].