

Data-Gov Wiki: Towards Linking Government Data

Li Ding, Dominic Difranzo, Alvaro Graves, James R. Michaelis, Xian Li,
Deborah L. McGuinness and Jim Hendler

Tetherless World Constellation, Rensselaer Polytechnic Institute, 110 8th st. Troy, NY 12180

Abstract

Data.gov is a website that provides US Government data to the general public to ensure better accountability and transparency. Our recent work on the Data.gov Wiki, which attempts to integrate the datasets published at Data.gov into the Linking Open Data (LOD) cloud (yielding "linked government data"), has produced 5 billion triples covering a range of topics including: government spending, environmental records, and statistics on the cost and usage of public services. In this paper, we investigate the role of Semantic Web technologies in converting, enhancing and using linked government data. In particular, we show how government data can be (i) inter-linked by sharing the same terms and URIs, (ii) linked to existing data sources ranging from the LOD cloud (e.g. DBpedia) to the conventional web (e.g. the New York Times), and (iii) cross-linked by their knowledge provenance (which captures, among other things, derivation and revision histories).

Introduction

Increasingly, the US Government has made efforts to make data from its constituent agencies available for public consumption. As part of this, the website *Data.gov* was launched, which hosts datasets from over 50 US government agencies. These datasets typically contain records with temporal, spatial, and numerical properties. For example, Dataset-353 (State Library Agency Survey 2006) contains information on how many visits were made to the New York State Library in the 2006 Fiscal Year.

As of October 15, 2009, Data.gov has published 942 datasets (see Figure 1) including 629 *Raw Data Catalogs*, which are downloadable government datasets, 312 *Tool catalogs* which allow users to browse government datasets, and a catalog (Dataset-92) which contains provenance metadata for datasets published at Data.gov.

On Data.gov, raw data catalogs are available either as downloadable data files or through interactive query interfaces, and are encoded in formats such as CSV, Text, Microsoft Excel, XML, ERIS Shape, and KML. This variety of formats complicates inter-dataset aggregation, as well as integration with external data sources. Therefore, we adopt a *linked data* (Berners-Lee 2007) based strategy to enhance

Copyright © 2010, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

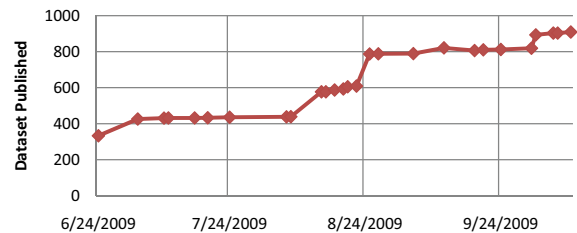


Figure 1: Data.gov has doubled its datasets in three months

the raw data from Data.gov, which leverages external data sources including the *Linking Open Data* (LOD) cloud¹ to produce *linked government data* (Berners-Lee 2009). In this paper, we investigate practical strategies for producing, processing and utilizing linked government data with special focus on the following issues:

- **Making Government Data Linkable** Simply publishing raw data on the web won't guarantee it can be easily consumed and referenced by web developers and users. For instance, some datasets on Data.gov lack table-based structuring (a header row followed by content rows). Likewise, additional datasets are known to have format-specific syntactic errors and/or incomplete data entries. Therefore, the Data.gov datasets should be preprocessed (e.g. normalized and dereferenceable) before they are added to the Linking Open Data cloud.
- **Linking Government Data** The value of Data.gov datasets can be greatly improved through interlinking based on relevance. For example, the EPA has published multiple datasets from the *Clean Air Status and Trends Network* (CASTNET) on Data.gov. Now, assume a user is browsing Dataset-8 (CASTNET ozone), which records ozone layer readings at different CASTNET sites in the US between 1987 and 2009. As the user browses the ozone readings of a particular observation site, he may be interested in knowing the geographic location of the site. Since geographic information on CASTNET sites is not included in Dataset-8, the user will need to reference additional information from the EPA's website in Dataset-10001 (CASTNET sites). In addition, the EPA has also

¹<http://linkeddata.org/>

published Toxic Release Inventory datasets on Data.gov (partitioned by year and states), and these datasets can be linked to support, for instance, multi-year data analysis.

- **Supporting the Use of Linked Government Data** In order for linked government data to be useful outside specialized research circles, it must be adopted and consumed by both web developers and users. Web developers will want access to linked government data in easy-to-use formats (such as JSON and XML) to design applications and interfaces around. In turn, normal web users will want a decent selection of web applications for browsing linked government data. Moreover, they could also be interested in not just consuming but also contributing to the data and later seeing their contributions integrated into the government data.
- **Preserving Knowledge Provenance** Provenance is pervasive throughout the linked government data. For instance, Dataset-92 records provenance annotations on all Data.gov datasets. The processes that convert, enhance, and use government datasets also generate provenance on how datasets were derived or used. Moreover, certain constantly updated datasets further produce revision provenance links that capture dataset histories enabling users to analyze changes.

In this paper, we discuss our efforts in developing the Data-gov Wiki - a website designed to host linked government data and demos - to address the issues above. The remaining sections in this paper are organized as follows. First, we provide an overview of the Data-gov Wiki. Second, we discuss strategies for making data.gov datasets linkable. Third, we describe how we linked datasets that were converted from Data.gov raw data catalogs. Fourth, we discuss the role of provenance in linked government data. Finally, we discuss related work and make concluding remarks.

The Data-Gov Wiki

The Data-gov Wiki (Ding et al. 2009) is not merely a wiki but a website² that delivers our research results on linked government data including: RDF datasets converted from Data.gov datasets, demos and tools that leverage the RDF data, and documents describing the Semantic Web techniques behind these efforts. As shown in Figure 2, the Data-gov Wiki runs in the following processes:

- **Conversion:** This process is fairly straightforward and highly extensible. First, the raw government data is cleaned-up and preserved through RDF-based representation. Second, these *converted* datasets use dereferenceable URIs, so that both the datasets and their ontologies can be extended by third party users.
- **Enhancement:** This focuses on extracting the semantics of literal values in government data into meaningful URIs and linking datasets by declaratively associating URIs mentioned in different datasets. Both automated and manual steps are involved in adding, deriving, linking and integrating government data. As a basis for the Data-gov

²This website is delivered by the Tetherless World Constellation at Rensselaer Polytechnic Institute. See <http://data-gov.tw.rpi.edu/>

Wiki, we use the Semantic MediaWiki(Krötzsch, Vrandečić, and Völkel 2006) platform, which allows users to collaboratively edit semantic content. This enables the Data-gov Wiki to allow consumers of government data to access raw datasets, as well as subsequent enhancements contributed by third party users.

- **Usage:** Here, the practical value of linked government data is highlighted through the design of interesting demos and tools combining both semantic and standard web technologies. For instance, in a number of demos hosted on the Data-gov Wiki, a SPARQL web service was used to bridge linked government data with conventional web APIs (such as the Google Visualization API). Since this SPARQL web service formats query results in both XML and JSON, web developers are given significant freedom in choosing how to process its results.
- **Knowledge Provenance:** On the Data-gov wiki, we pay special attention to representing and maintaining provenance metadata that was embedded in the original government data, as well as provenance that can be inferred from the processes executed in the three tasks above.

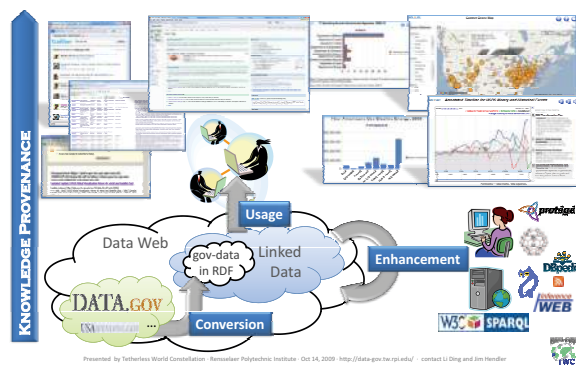


Figure 2: The Architecture of the Data-gov Wiki

The Data-gov Wiki is currently hosting 116 RDF datasets, covering 401 (65%) of the Data.gov raw data catalogs plus the Data.gov catalog(Dataset 92)³. While most of these datasets are static, a few are updated frequently, such as Dataset-34 (Worldwide M1+ Earthquakes, Past 7 Days) which gives daily updates on earthquakes occurring worldwide. The RDF datasets based off Data.gov raw data constitute a total of 5 billion triples, describing 459,412,871 table entries. These datasets form 11 topic clusters, as shown in Figure 3. Aside from datasets distributed by Data.gov, the Data-gov Wiki also stores RDF datasets based on raw data from other sources, such as *usaspending.gov*, *epa.gov*, *usps.com*, and *ca.gov*.

³There are 4 national EPA Toxic Release Inventory datasets covering 286 sub-datasets (corresponding to individual states over different years); so we skipped converting the sub-datasets to avoid duplication.

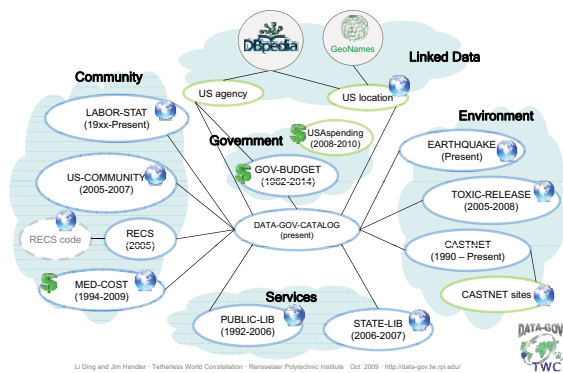


Figure 3: The cloud of RDF datasets on Data.gov Wiki

Making Government Data Linkable

On the Data.gov Wiki, we limited our conversion efforts to well-formed CSV files with headers. This was done since CSV files proved easy to convert into RDF and many datasets in data.gov were published in the CSV format. Our conversion strategy adopts the following principles.

First, the conversion should be *minimal and extensible* - just enough to preserve the structure and content of the raw government data. Doing this helps simplify access and usage of generated RDF datasets, which will allow third parties to extend them later on. During conversion, each non-header table row (entry) is assigned a URI, and table column names (on the header row) are mapped to RDF properties with two simple rules: (i) the namespace of each property must be uniquely tied to the dataset it comes from; and (ii) the local name must be the normalized column name. We did not use properties from existing ontologies, as this requires manual moderation. Since properties from different datasets sharing the same name could be semantically different, two datasets should not use the same property namespace unless explicitly stated by a dataset curator. In turn, all non-header cells are mapped to literal strings - we don't create new URIs unless the cell value is either a URL or a hyperlink whose anchor text is the same as its link.

Second, the converted government data should be *web accessible*. To enable this, the RDF/XML encoding syntax was chosen to ensure that converted government data could be consumed by both RDF and XML parsers, and that it could be further queried using both SPARQL and XQuery. Each table row and dataset has a dereferenceable URI via the HTTP protocol. RDF data converted from very large tables⁴ should be partitioned into multiple small files (or stored in a scalable triple store) to support linked data browsing and to prevent out-of-memory complications.

Linking Government Data

Government data, imported from Data.gov or elsewhere, can be linked at different processing stages in the Data.gov Wiki.

⁴The largest imported raw dataset from Data.gov contains a 7GB CSV file. See Dataset-91 (US Census population) which provides US census data by regions and years.

Such links can be contributed by dataset curators, as well as third parties (such as web developers and users).

Linking at Conversion Time

During the conversion process, we found that many data.gov datasets are closely linked, through both equivalent (or very similar) topics and table headers. For example, Data.gov hosts a set of datasets from the EPA on Toxic Release Inventories for each individual state from 2005 to 2008⁵. Here, there is no need to define unique namespaces for the properties used in every single dataset. We therefore set one common namespace for these datasets, which serves to interlink their content.

Linking through Semantic MediaWiki

Following conversion, we kept RDF datasets extensible through Semantic MediaWiki. Properties used in these converted RDF data are dereferenceable to terms in well-known ontologies (e.g. FOAF and Dublin Core) or RDF/XML pages generated by Semantic MediaWiki.

This approach allows for improved usage of the converted Data.gov datasets by Semantic Web tools. Tabulator⁶ is a linked data browser. When displaying RDF resources, tabulator will look for an *rdfs:label* (or corresponding subproperty) value to display in place of a URI. When browsing Dataset-92 with Tabulator, English-based labels (*rdfs:label*) could be significantly more descriptive than URIs. To this end, we use Semantic MediaWiki to update the description of *dgp92:title*⁷ by asserting it as a *rdfs:subPropertyOf rdfs:label*. Following this, the enhanced definition will be immediately available for Tabulator to load and use.

We have also linked Data.gov datasets to the Linking Open Data cloud. For instance, we first obtain the RDF version of Dataset-92 by generating URIs for each mentioned US state/territory. We then map these URIs to equivalent URIs in DBpedia and Geonames via the *owl:sameAs* predicate. In practice, we adopted a semi-automated approach to build these mappings. First, we generate preliminary mappings using simple heuristics, such as: "Given a state name *X Y* from the dataset, add the DBpedia namespace (e.g. DBpedia:*X Y*) and replace whitespaces with underscores (DBpedia:*X_Y*)". Here, "New York" can be mapped to http://dbpedia.org/resource/New_York. However the US state "Georgia" has a different DBpedia uri [http://dbpedia.org/resource/Georgia_\(U.S._state\)](http://dbpedia.org/resource/Georgia_(U.S._state)), to distinguish it from the country "Georgia". Therefore we use a widget (MediaWiki extension) to help end users to detect and correct auto-generated DBpedia links (see Figure 4).

⁵Dataset-135-308, 628-699 and 700-743, see http://data.gov.tw.rpi.edu/wiki/Environmental_Protection_Agency

⁶<http://www.w3.org/2005/ajar/tab>

⁷*dgp92* is the prefix of the namespace for Dataset-92's properties. The uri of this property is <http://data.gov.tw.rpi.edu/wiki/Property:92/title>



Figure 4: Using semantic wiki to create links to DBpedia

Linking by Other Means

Aside from the ways above, we also determined strategies for linking data by reusing literal strings. For example, Dataset-8 (CASTNET Ozone) can be linked to Dataset 10001 (CASTNET sites), because they use the same literal identifiers for CASETNET sites. Below is a SPARQL query linking the two datasets.

```

PREFIX dgpl: <http://data-gov.tw.rpi.edu/vocab/p/10001/>
PREFIX dgtwc: <http://data-gov.tw.rpi.edu/2009/data-gov-twc.rdf#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
SELECT ?id ?lat ?long ?value
FROM <http://data-gov.tw.rpi.edu/raw2/10001/data-10001.rdf>
FROM <http://data-gov.tw.rpi.edu/raw/8/rindex.rdf>
WHERE {?o2 dgpl:site_id ?id .
       ?o2 dgpl:latitude ?lat .
       ?o2 dgpl:longitude ?long .
       ?url rdf:type ?id .
       ?url dgtwc:avgValue ?value .
}

```

Different datasets on similar topics can also be linked by temporal or spatial properties. In one case, we have shown that Data.gov datasets can be related to information from external sources. One of our demos leverages two data sources: (i) Dataset-401 (Public Budget Database - Budget Authority and offsetting receipts 1976-2014) from Data.gov, and (ii) news article listing from the New York Times (NYT), i.e. an archival dataset (accessible via API⁸) which allows users to query for past news from NYT and obtain search results in JSON format.

For querying the New York Times archive via the provided API, each query string consists of an agency name, a keyword (e.g. “budget”), and a time range (e.g. between 1981-01-01 and 2006-12-31). Shown in figure 5, the budget for the Civil Works of the Corps of Engineers got a significant budget increase in 2006. Meanwhile, around the year 2005, we can see in point “J” news about design flaws in the New Orleans levees. This event (shown in the New York Times data) may help explain the findings from Dataset-401.

The above examples are mostly done by web application developers. However, with the help of Web 2.0 technology, we can further allow end users to contribute linked data to government datasets. Figure 6 is a snapshot of our demo that links information about the Postal Service Fund (a government account) from two different Data.gov datasets - Dataset-401(Public Budget Database - Budget Authority and offsetting receipts 1976-2014), and Dataset-402(Public

⁸<http://developer.nytimes.com/>



Figure 5: Linking agency budgets to news feed API from New York Times

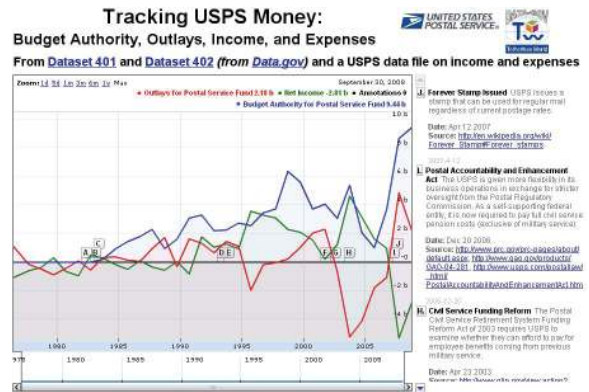


Figure 6: Linking USPS budget data with a user editable news feed from the Data-gov Wiki

Budget Database - Outlays and offsetting receipts 1962-2014). In addition, Dataset-10006 (supplemental USPS income and expense data) is extracted from the USPS website. Moreover, an RSS feed covering user contributed USPS related news events is collected on the Data-gov Wiki, where end users can easily add or modify relevant information related to the USPS via form interface on that wiki. Running this demo, a user may notice the significant budget outlays drop in 2004, and attempt to contribute relevant news articles for future reference by the demo.

Provenance Links in Data-gov Wiki

While datasets can be linked by through URI and namespace-based relations, they can also be linked by knowledge provenance(McGuinness and da Silva 2004).

Provenance Metadata

Provenance annotations are widely used on Data.gov to explain the curation metadata of datasets. For example, Dataset 92 provides provenance annotations for all Data.gov datasets. Here, example properties include: *title* for the

title of a dataset, *agency* for the government agency publishing the dataset, *geographic coverage* for which geographic location(s) were mentioned in the dataset, *category* for the dataset’s topic classification, *date updated* for when the dataset was updated, *data dictionary variable list* for showing how to get a detailed metadata description of the dataset’s content, and *CSV/TXT access point* for defining the URLs from which users may retrieve the CSV or Text based versions of the dataset.

Provenance has also been used for capturing how datasets are related through manipulation processes. For instance, in the conversion process, each converted RDF dataset is related to its original Data.gov dataset via *derivation provenance linking*. During subsequent enhancements, a user-contributed dataset that links an RDF dataset converted from Data.gov raw data to DBpedia can be related to its author via *creation provenance linking*. If archived, any revisions of an RDF dataset can be correlated by *revision provenance linking*. During dataset usage, visualization demos can be linked to the Data.gov Wiki RDF datasets or webpages used via *usage provenance linking*.

Maintaining and Using Provenance Links

To illustrate how provenance links can be created and used in RDF datasets derived from Data.gov, we consider the problem of tracking changes made over time to Data.gov datasets. Currently, no features are provided on the Data.gov website for tracking raw dataset modifications. A manual effort to accomplish this was logged on Twitter⁹, but has been updated sparsely due to the amount of work involved. We therefore adopt a semi-automated approach, as shown in Figure 7, which uses a selection of Semantic Web based technologies. Starting with raw (CSV-based) versions of the Data.gov datasets, we start by using a Web service *CSV2RDF*¹⁰ to extract embedded semantic structuring. *CSV2RDF* takes the following parameters: *URL* for the URL of an online dataset, *xmlbase* for the xmlbase URL of the RDF file to be generated, *propns* for the URL of the namespace of properties derived from the original CSV table header fields, and *output* for the desired output RDF syntax (RDF/XML is used by default). Following this, we use a set of PHP scripts to take snapshots for each observed revision a given dataset¹¹. Next, we use the *SemDiff* web service¹² to compute and summarize the difference between the two consecutive versions of a dataset in RDF. This service currently only processes RDF data without blanknodes and employs a straightforward RDF graph difference computation described in (Carroll 2003). *SemDiff* takes the following parameters: *cur* and *prev* for the URLs of the current and the previous revisions. The output of *SemDiff* summarizes changed instances in RDF or RSS.

In the situation above, various provenance links are created and used. For each web service call, we can embed

⁹http://twitter.com/DataGov_Tweets

¹⁰<http://data-gov.tw.rpi.edu/ws/csv2rdf.html>

¹¹both the last-modified field in the HTTP header and the CSV file length were used to determine if a dataset had been changed

¹²<http://onto.rpi.edu/sw4j/diff.html>

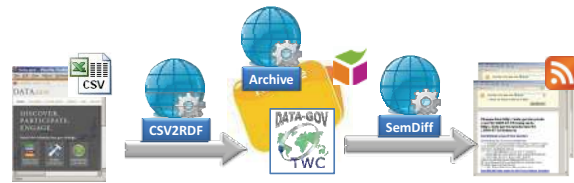


Figure 7: RESTful web services based architecture for creating and computing revision provenance links

provenance metadata into the service result to declaratively establish *derivation provenance links*. For this, the RDF data created by *CSV2RDF* service is originally derived from a CSV file; in turn, corresponding visualization demos derive information from RDF datasets from the Data.gov Wiki. With the above provenance metadata, users may compute dependencies transitively to, for instance, establish which original Data.gov datasets were used in a demo.

The *revision provenance links* are the key input to the *SemDiff* service. By linking the latest two revisions of Dataset-92, we can derive and summarize their differences (e.g. what types of revision operations were performed on which entries) in RDF. We then can use this information to further derive statistics as well as generate explanations. For example, we can infer that most of the revisions (74%) on the Data.gov catalog are adding new datasets, and that other revisions are updating existing entries (e.g. changing the publishing agency).

Furthermore, we can derive detailed explanations of update operations using SPARQL querying on the revision provenance links and the archived original dataset. Here are some examples (changes of Dataset-92 from 2009-07-03 to 2009-07-10) generated by the SPARQL query below: (i) Dataset-93’s *title* was changed to “USGS Global Visualization Viewer for Aerial and Satellite Data” from “Landsat Satellite data: land-surface images of the entire Earth, ... from 1972 to the present.,”, and (ii) Dataset 101’s *agency* was change to “National Archives and Records Administration” from “Government Printing Office”.

```
SELECT DISTINCT ?url ?title ?revision_type ?revision_details ?graph
FROM NAMED <http://data-gov.tw.rpi.edu/raw/92/2009-07-03/diff-data-92.rdf>
FROM NAMED <http://data-gov.tw.rpi.edu/raw/92/2009-07-10/diff-data-92.rdf>
FROM <http://data-gov.tw.rpi.edu/raw/92/2009-07-10/data-92.rdf>
FROM <http://inference-web.org/2.0/pml-owl>
WHERE {
  ?o <http://www.w3.org/2000/01/rdf-schema#label> ?revision_type .
  ?s1 <http://data-gov.tw.rpi.edu/vocab/p/92/url> ?url .
  ?s1 <http://data-gov.tw.rpi.edu/vocab/p/92/title> ?title .
  GRAPH ?graph {
    ?s <http://inference-web.org/2.0/pml-owl#hasDiffRelation> ?o .
    ?s <http://www.w3.org/2000/01/rdf-schema#comment> ?revision_details.
    ?s <http://inference-web.org/2.0/pml-owl#hasInstanceReference> ?s1.
  }
  FILTER ( !regex(str(?o), "hasDiffAddInstance") . )
ORDER BY ?url
```

Related work

Increasingly, the Web community has become engaged with the idea of open government data. *Apps for Democracy*¹³ (US) and *Show Us a Better Way*¹⁴ (UK) are promoting

¹³<http://www.appsfordemocracy.org/>

¹⁴<http://www.showusabetterway.co.uk/>

technologies that make government data useful for citizens, businesses and government agencies. Efforts from the *Sunlight Foundation*¹⁵ and *OMBWatch*¹⁶ have yielded web-based databases and tools to make government information accessible online about members of Congress, their staff, legislation, federal spending and lobbyists.

Semantic web technologies have been used in some work in facilitating users to access government data. There are some Semantic Web powered services tracking and publishing government data in RDF. *GovTrack*¹⁷ is a civic project that collects data about the U.S. Congress and provides both XML and RDF APIs (including SPARQL query endpoints) for others to reuse its data. The *OpenPSI Project*¹⁸ is now collecting RDF based catalog information about UK's government datasets to support government-based information publishers, research communities, and web developers.

Work exists on integrating Semantic Web ontologies with geospatial information for e-government. (Wiegand 2007) proposes linked data enhanced search for geospatial data in e-government portals by leveraging Semantic Web ontologies in data browsing and query expansion. (Goodwin, Dolbear, and Hart 2009) showed an example dataset for the administrative geography of Great Britain to demonstrate the benefit of using declarative topological links between geographic entities in spatial queries.

Several Semantic Web vocabularies can be used to annotate linked government data and their provenance. *OE-GOV*¹⁹ is publishing OWL ontologies for eGovernment to enable distributed creation and maintenance of metadata about government data, and thus support provenance and trust in the sources of government data. The *Proof Markup Language* (PML) (McGuinness et al. 2007) is an Semantic Web provenance interlingua which has been used for tracking both provenance annotations and links in many settings such as theorem proving, machine learning, and text analytics. There is a recently proposed semantic web provenance ontology (Hartig 2009) on tracking web access patterns across linked data.

Conclusion

Using the Data.gov Wiki as a presentation medium, we have published 5 billion triples - derived from raw data on Data.gov - and built a number of demos from this converted data. We investigated various strategies for making government data linkable and maintainable. While government datasets can be linked by cross-reference, sharing the same ontological, temporal, or spatial properties, they can also be linked by provenance associations, including derivation-based and version-based. In many cases, links within linked government data can be generated through simple automated heuristics; however, support from human curators is also needed to handle exceptions in linked data. A readable and

writable interface, such as a semantic wiki, is needed to tap the power of social networks to expand our data corpus.

Our ongoing efforts will include converting more raw government data (from Data.gov and elsewhere) into RDF, and developing more applications and demonstrations for leveraging linked government data. In addition, we will maintain our efforts to link converted government datasets to the Linking Open Data cloud, and examine links between potential US (e.g., data.gov) and UK linked government data efforts (Alani et al. 2007). Strategies will also be explored for helping web users with limited technical background contribute to this project (e.g. crowdsourcing data linking and/or application development). We will also be increasing our efforts to capture, propagate, and maintain provenance information related to linked data. As part of this, we will do further research on scaling up the SemDiff service to enable provenance beyond the sameAs links, which will provide a basis for more sophisticated provenance problems.

Acknowledgement

This work is partially supported by NSF Award #0524481, IARPA Award #FA8750-07-2-0031, and DARPA award #FA8750-07-D-0185, #55-002001, #FA8650-06-C-7605.

References

- Alani, H.; Dupplaw, D.; Sheridan, J.; O'Hara, K.; Darlington, J.; Shadbolt, N.; and Tullio, C. 2007. Unlocking the potential of public sector information with semantic web technology. In *ISWC/ASWC*, 708–721.
- Berners-Lee, T. 2007. Linked data. <http://www.w3.org/DesignIssues/LinkedData.html>.
- Berners-Lee, T. 2009. Putting government data online. <http://www.w3.org/DesignIssues/GovData.html>.
- Carroll, J. J. 2003. Signing rdf graphs. In *International Semantic Web Conference*, 369–384.
- Ding, L.; DiFranzo, D.; Magidson, S.; McGuinness, D. L.; and Hendler, J. 2009. The data-gov wiki: a semantic web portal for linked government data. In *ISWC'09 poster*.
- Goodwin, J.; Dolbear, C.; and Hart, G. 2009. Geographical linked data: The administrative geography of great britain on the semantic web. *Transactions in GIS* 12.
- Hartig, O. 2009. Provenance information in the web of data. In *Linked Data on the Web*.
- Krötzsch, M.; Vrandečić, D.; and Völkel, M. 2006. Semantic mediawiki. In *ISWC'06*, 935–942.
- McGuinness, D. L., and da Silva, P. P. 2004. Explaining answers from the semantic web: the inference web approach. *J. Web Sem.* 1(4):397–413.
- McGuinness, D. L.; Ding, L.; da Silva, P. P.; and Chang, C. 2007. Pml 2: A modular explanation interlingua. In *Workshop on Explanation-aware Computing (ExaCt-2007)*.
- Wiegand, N. 2007. Semantic web for geospatial e-government portals. In *Proceedings of the 8th annual international conference on Digital government research*.

¹⁵<http://www.opencongress.org/>

¹⁶<http://www.fedspending.org/>

¹⁷<http://www.govtrack.us/developers/data.xpd>

¹⁸<http://www.openpsi.org/>

¹⁹<http://www.oegov.us/>