

SEMANTIC WEB FOR OPEN GOVERNMENTS

Semantic Web technologies offer great benefits in helping aggregate and publish big data. One such area of particular importance is government data. Not only is the scope in question extensive, but the variety of kinds of information and its methods of storing and publishing demand technologies which are efficient, accessible and transparent.

This gains even more significance in the face of **open data** initiatives of several countries launched in recent years. One such initiative is the American government's portal Data.gov, whose aim is *"to increase public access to high value, machine readable datasets generated by the Executive Branch of the Federal Government"*. The site stores and publishes all data collected by the government (with the exception of private and security-related information), which constitutes a tremendous resource of knowledge about the country.

LINKING OPEN GOVERNMENT DATA (LOGD)

One project that aims to demonstrate the benefits of semantic technologies in open government data is **Linking Open Governemnet Data** (LOGD), led by Professor Jim Hendler (one of the pioneers of the Semantic Web). It is a research initiative whose goal is to investigate how semantic technologies can be used to make government data more transparent, accessible and capable of being processed.

*"Data.gov mandates that all information is accessible from the same place, but **the data is still in a hodgepodge of different formats using differing terms, and therefore challenging at best to analyze and take advantage of**"* James Hendler

The team translates government-published datasets into RDF, enabling to combine information which so far could not be compared due to inconsistencies in formats and technology. This enables to generate **interactive mash-ups of data from any given source**, providing new insight into phenomena from all areas of government data – population statistics, income rates, weather data, pollution measures, car ownership, education and many others. An example of how liberating data from structure can result in interesting findings relates to the Clean Air Status and Trades Network (CastNet):

CastNet measures ground-level ozone and other pollutants at stations all over the country, but CastNet doesn't give the location of the monitoring sites, only the readings from the

sites. The Rensselaer team located a different data set that described the location of every site. **By linking the two along with historic data from the sites, using RDF, a semantic Web language, the team generated a map that combines data from all the sets and makes them easily visible.** (source: <http://cacm.acm.org/news/101857-reusselaer-team-shows-how-to-analyze-raw-government-data/fulltext>)

The fundamental goal behind the initiative is to enable everyone, regardless of their knowledge and skill in Web programming and IT, to make use of such mechanisms and create new ones themselves. Given the enormous amount of information already stored, it is easy to realize how potent such a tool might be, especially in the fields of research, education and journalism.

EXAMPLES OF MASH UPS:

- Trends in Smoking Prevalence, Tobacco Policy Coverage and Tobacco Prices (1991-2007)
http://logd.tw.rpi.edu/demo/trends_in_smoking_prevalence_tobacco_policy_coverage_and_tobacco_prices
- Data.gov Mashathon 2010: an Energy Data Mashup
<http://www.data.gov/developers/node/datagov-mashathon-2010-energy-data-mashup>
- US Global Foreign Aid, 1947-2008 <http://data.gov.tw.rpi.edu/demo/stable/usaid2008/demo-1554.html>

RESOURCES

LOGD official website: <http://logd.tw.rpi.edu/>

Data.gov Semantic Web section <http://www.data.gov/developers/page/semantic-web>

Rensselaer Team Shows How To Analyze Raw Government Data

<http://cacm.acm.org/news/101857-reusselaer-team-shows-how-to-analyze-raw-government-data/fulltext>