Integrating Open Government Data for Transparency

Open Government Data is meant to provide transparency, but the inconsistent structure inhibits querying across multiple sources. GovWILD integrates data about politicians, companies, and funds into a clean and interlinked data set. We provide a Web interface to uncover hidden connections and to drill-down into the network of politics and industry.

Features

- Search 150,000 persons, 270,000 legal entities and 1,100,000 funds
- Explore the data interactively in different views (graph, wiki, table)
- Query the data using keyword search or SPARQL queries
- Download the integrated data for further analysis as RDF or JSON

Data

USA: USASpending, Earmarks, Congress
Europe: Financial Transparency System, Parliament Members
Germany: Party Donations, Agricultural Subsidies, Commercial Register, Bundestag
Other: Freebase, New York Times Articles tagged with entities

The data set enriches the LOD cloud with sameAs links to Freebase, DBpedia and New York Times.

Integration Process

Individual components transform data to a common structure, identify common entities across multiple sources, and finally fuse conflicting data into a value-added and rich data set.

- JAQL and SystemT: IBM
- Duplicate Detection (DuDe) toolkit: HPI
- Information Workbench: fluid Operations

WWW2012 Demo: Integrating Open Government Data for Transparency

Christoph Böhm, Markus Freitag, Arvid Heise, Claudia Lehmann, Andrina Mascher, Felix Naumann
Information Systems Group at Hasso Plattner Institute University of Potsdam, Germany
team@govwild.org

Peter Haase, Michael Schmidt
fluid Operations, Walldorf, Germany

Vuk Ercegovac, Mauricio Hernandez
IBM Research - Almaden, California, USA