Towards Interoperable Visualization Applications Over Linked Data

Ghislain Atemezing / @gatemezing
Raphaël Troncy / @rtroncy
Goal and Agenda

- Goal: Describe common patterns of visualization applications from Government Data Portals to make them reusable and interoperable

- The importance of visualization in Linked Data
  - Show some use cases
  - Survey of Apps in Government Data Portals

- On Reusable Applications
  - A survey of visualization tools used by Open Data Apps
  - Search for common and reusable features

- DVIA: a vocabulary for Describing Visualization Apps

- Conclusion and Future Work
An application is worth a billion triples

- **Use case 1**: “Give me applications in Government Linked Data platforms in the domain of taxes”

- **Use case 2**: “Find me apps running on Android in data.uk.gov for finding the nearest drugstores”

- **Use case 3**: “How many free or inexpensive applications are available using US datasets in Health and Insurance domains”
Visualization Apps in Government Portals

- Study of applications consuming Open Data
  - Countries: UK (7), USA (3) and France (3)
  - Number of applications: (13)
  - Domain: education, health, transport, government, city, housing, criminality, foreign aid

- Classification dimensions
  - Platform (web, mobile), data sources, which views are available (maps, charts, timeline, etc.)
  - URL policy for identifying data objects
  - License for the application / for the data
  - Commercial / non-commercial
## Visualization Apps in Government Portals

<table>
<thead>
<tr>
<th>Applications</th>
<th>Domain</th>
<th>Country</th>
<th>Use RDF data</th>
<th>Platform</th>
<th>Views</th>
</tr>
</thead>
<tbody>
<tr>
<td>UK Crime</td>
<td>Criminality</td>
<td>UK</td>
<td>no</td>
<td>Web</td>
<td>Map, chart, pie chart, tabular</td>
</tr>
<tr>
<td>UK Pharmacy</td>
<td>Drugstore</td>
<td>UK</td>
<td>no</td>
<td>iPhone, Android</td>
<td>Map</td>
</tr>
<tr>
<td>Numberhood</td>
<td>Local area dynamics</td>
<td>UK</td>
<td>no</td>
<td>IPhone</td>
<td>chart, histogram, tabular views</td>
</tr>
<tr>
<td>BUSIt London</td>
<td>Transport</td>
<td>UK</td>
<td>no</td>
<td>Web, Mobile</td>
<td>Map</td>
</tr>
<tr>
<td>UK School Finder</td>
<td>Education</td>
<td>UK</td>
<td>no</td>
<td>Web</td>
<td>Map, tabular, histogram</td>
</tr>
<tr>
<td>Where-can-I-Live</td>
<td>Housing</td>
<td>UK</td>
<td>no</td>
<td>Web</td>
<td>Map, chart</td>
</tr>
<tr>
<td>Open Data Communities</td>
<td>Local government</td>
<td>UK</td>
<td>yes</td>
<td>Web</td>
<td>Graph, Map</td>
</tr>
<tr>
<td>FlyOnTime</td>
<td>Transport</td>
<td>US</td>
<td>no</td>
<td>Web</td>
<td>Pie chart, histogram, tabular</td>
</tr>
<tr>
<td>White House Visitor Search</td>
<td>Government</td>
<td>US</td>
<td>yes</td>
<td>Web</td>
<td>Bar chart, pie chart, tabular</td>
</tr>
<tr>
<td>Comparing US-USAID vs UK-DFID Global Foreign Aid</td>
<td>Foreign Aid</td>
<td>US</td>
<td>yes</td>
<td>Web</td>
<td>Map, pie chart</td>
</tr>
<tr>
<td>Fourmisante</td>
<td>Health</td>
<td>France</td>
<td>no</td>
<td>Web</td>
<td>Map</td>
</tr>
<tr>
<td>MaVilleVuDuCiel</td>
<td>City</td>
<td>France</td>
<td>no</td>
<td>Web</td>
<td>Map, facet views</td>
</tr>
<tr>
<td>Home' n' Go</td>
<td>Housing</td>
<td>France</td>
<td>no</td>
<td>Web</td>
<td>Map, tabular views</td>
</tr>
</tbody>
</table>
Linked Data Applications Categories (*)

- **Generic applications**
  - Used to browse any RDF graph
    (e.g. Tabulator, OpenLink Data Explorer)

- **Vocabulary-specific applications**
  - Built around particular vocabularies
    (e.g.: FaceBook Social Graph API, IsaViz, CubeViz)

- **Data-specific applications (mashups)**
  - Built around particular datasets
    (e.g. Where does my money go, IOGDS: International Open Government Dataset Search)

(*) Classification according to: Jeni Tennison: [http://www.jenitennison.com/blog/node/126](http://www.jenitennison.com/blog/node/126)
Relevant Features in Visualization Tools

- Data format given as input (csv, xml, shp, rdf, etc.)
- Data access (API, dump, etc.)
- Language code
- Type of view
- External Libraries
- License
- Metadata: author, organisation
Classes of Visualization Tools

- 2 Classes of visualization tools
  - Those operating over RDF data
  - Those operating over structured data

- 5 features to assess a visualization tool
  - Usability
  - Visualization capabilities (type of views)
  - Data accessibility (API, endpoint, dump)
  - Deployment
  - Extensibility
## Visual Tools with API access

<table>
<thead>
<tr>
<th>Tools</th>
<th>Data formats</th>
<th>Type of Views</th>
<th>License</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chooseel</td>
<td>xls, csv</td>
<td>Text, Map, BarChart</td>
<td>Open</td>
</tr>
<tr>
<td>Many Eyes</td>
<td>xls, plain text, html</td>
<td>Charts, trees, graphs, maps</td>
<td>IBM</td>
</tr>
<tr>
<td>D3.js</td>
<td>csv, svg, GeoJson</td>
<td>Charts, trees, graphs, maps</td>
<td>Open</td>
</tr>
<tr>
<td>Google Visualization API</td>
<td>Json, csv</td>
<td>Many charts, control and dashboard</td>
<td>Open</td>
</tr>
<tr>
<td>Data Publica</td>
<td>DSPL</td>
<td>Map, graph, histogram, table</td>
<td>Proprietary</td>
</tr>
<tr>
<td>GeoAPI</td>
<td>gml, kml, gpx</td>
<td>Map views</td>
<td>Free for non commercial use</td>
</tr>
</tbody>
</table>
## Visual Tools using RDF

<table>
<thead>
<tr>
<th>Tools</th>
<th>Type of views</th>
<th>License</th>
<th>Data Access</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fresnel</td>
<td>Properties, labels</td>
<td>Open</td>
<td>-</td>
</tr>
<tr>
<td>Spark</td>
<td>Charts, simple table</td>
<td>Open</td>
<td>SPARQL</td>
</tr>
<tr>
<td>LDA</td>
<td>-</td>
<td>Open</td>
<td>SPARQL</td>
</tr>
<tr>
<td>SemWeb Import</td>
<td>Graph, node views</td>
<td>CECILL-B</td>
<td>SPARQL CONSTRUCT</td>
</tr>
<tr>
<td>Facet Spatial</td>
<td>Map, facet views</td>
<td>Open</td>
<td>SPARQL</td>
</tr>
<tr>
<td>Semantic Browsing</td>
<td>Map, facet views</td>
<td>Open</td>
<td>SPARQL</td>
</tr>
<tr>
<td>Widgets</td>
<td>Map, facet views</td>
<td>Open</td>
<td>SPARQL</td>
</tr>
<tr>
<td>Sgvizler</td>
<td>Map, charts, timeline, sparkline</td>
<td>Open</td>
<td>SPARQL SELECT</td>
</tr>
<tr>
<td>Visual Box</td>
<td>Map, charts, timeline, graphs</td>
<td>Open</td>
<td>SPARQL SELECT</td>
</tr>
<tr>
<td>Map4rdf</td>
<td>Map, facet views</td>
<td>Open</td>
<td>SPARQL</td>
</tr>
</tbody>
</table>
Scope/Domain: Department for Communities and Local Government, datasets access

Description: visualize available datasets (finance, housing, deprivation, geography) by authorities or postcode. On the dashboard, it provides graphs showing the national distribution of a district and how the values for this local authority compare with others in England.

Supported Platform: Web

URL Policy: http://{domain}/id/{...} with redirection to the corresponding document at: http://{domain}/doc/{...}. Hampshire County Council is: http://opendatacommunities.org/id/county-council/hampshire

Data Sources: 36 datasets from DCLG, Administrative Geography and Postcodes from Ordnance Survey.

Type of View: Graph, Map views.

Visualization Tools: google visualization API, raphael.js

License: Open Government license [OGL]

Business Value: Non commercial
Scope/Domain: Public spending, Government

Description: The application helps visualizing the most characteristic facts of the Greek public spending, interconnected to foreign expenditure and other data.

Supported Platform: Web


Data Sources: http://opendata.diavgeia.gov.fr; Greek Tax data (TAXIS).

Type of View: Bubble tree, column and bar charts

Visualization Tools: HighchartsJS, Bubble TreeJS, JqueryJS; RaphaelJS

License: Open Government license [OGL]

Business Value: Non commercial (Free)
A vocab to describe Apps

Prefixes:
@prefix dct: <http://purl.org/dc/terms/>.
@prefix dcat: <http://www.w3.org/ns/dcat#>.
@prefix dctype: <http://purl.org/dc/dcmitype/>.
@prefix org: <http://www.w3.org/ns/org#>.
@prefix dvia: <http://data.eurecom.fr/ontology/dvia#>.

DVIA vocabulary: http://data.eurecom.fr/ontology/dvia#
Conclusion and Future Work

- **Use and evaluate DVIA**
  - Scrap and reconcile data from apps contests: see the upcoming catalog from Apps4Europe
  - Interconnect application descriptions that use heterogeneous datasets in order to reuse tools and views

- **Detect patterns for visualization**
  - Inspect datasets to detect top categories:
    - persons, organizations, events, geo, time series, statistics, etc.
  - Inspect “sameAs” links between datasets:
    - focus visualization on properties attached to interconnected data objects
  - Generate automatically visualizations based on patterns
Thanks for your attention!

Questions?