

Towards Interoperable Visualization Applications Over Linked Data

Ghislain Auguste Atemezing and Raphaël Troncy

EURECOM, Campus SophiaTech, France
{auguste.atemezing, raphael.troncy}@eurecom.fr

Abstract. The adoption of Semantic Web technologies in many institutions or government decision-makers are often based on an actual demonstration of their potential via an application. Here, the famous adage “A picture is worth a thousand words” could be rephrased as “*An application is worth a billion triples*”. Hence the importance of having at some point of the Linked Open Data life-cycle, the application layer aiming to provide a way to quickly consume the underlying data being published.

Nowadays, as many government are opening their data, they need a way to rapidly develop smart applications for giving great insights of their data, ranging from exploration applications to more complex ones. Many tools and libraries already exist from the information visualization communities, and many of them are successfully applied to the semantic web. However, many of the great applications developed since the beginning of this era of government Linked Data lack of reusing them, not only the idea based on similar data (e.g. “Where My Money/Tax Goes” applications for tracking government expenses), but also the tools or widgets used to build them. A possible use case could be the following “*What application won the Semantic Web Challenge using facet views for mobile phone?*” or “*What are the tools commonly used for building smart applications in the health domain?*”. We believe that as long as semantic technology is being adopted, there will be always a need for creating smart applications, and developers or consumers will not necessary want to waste time reinventing the wheel, but rather focusing on the added-value of their applications. We propose in this paper an overview of types of applications built with Linked Data, a framework for assessing tools used for building those applications, and a vocabulary to describe visual applications developed on top of LOD for more interoperability and components reusability. Our proposal derives from a study of applications in different domains built in Open Data initiatives in UK, US and France. We identify two benefits of our approach: (i) it could help building a sustainable ecosystem of visual applications making use of 4-5 stars datasets, and (ii) it could make it easier to find application patterns used for creating Linked Data Applications.

Keywords: Visualization tools, mashups, Linked Data, interoperability, consumption, vocabulary

1 Introduction

As many initiatives on Linked Open Data is growing, tools and technologies are getting more and more mature to help pro-consumers to leverage the lifting process of the data. At the same times, standardization bodies such as W3C are helping in providing best practices to publish Open Government Data by using appropriate vocabularies, taking care of stability in the URIs policies, and making links to other datasets. It is the case for example of the Government Linked Data Working Group¹ which is developing standards to help governments publishing their data using Semantic Web technologies. Having a look at different proposals of the Life Cycle of Government LD, one of the last stage is “Publication” where the data is released according to the 4-5 stars principles², with a given access to a SPARQL endpoint. However, for a better understanding of the data, one of the next step is usually to building visualizations through intuitive charts, graphs, etc. that will benefit to citizens, data journalists and other public authorities to improve the quality of their decisions. At the moment, one way of doing is to look around previous initiatives to see what type of application is already there, and make something similar according to a given dataset. Another approach is by organizing *contests* where the challenges are to mash up unexpected datasets with clear and beautiful visualizations. Such approach is harder as developers also try figure out

¹ <http://www.w3.org/2011/gld/>

² <http://5stardata.info/>

which tool and library is used for different applications. What if we describe applications according to the facets/views, datasets, visual tools etc.? In this paper, we propose a small vocabulary aiming at describing applications developed on top of LOD for more interoperability and components reusability. This paper is organized as follow: Section 3 defines the types of applications on LOD, followed by a framework for assessing tools for building such applications (cf. Section 3.2). Section 3.3 deals with the facts which are common to many visual applications, and Section 4 presents the vocabulary for describing the semantics of mashups. We make some references to similar works in Section 2 and provide some outlook in Section 5.

2 Related Work

The Open Data Service at the University of Southampton³ has a register of all the applications developed using their datasets. A catalog of applications using the data is available at <http://id.southampton.ac.uk/dataset/apps>. Each application is described by giving a name, the type of the application, date of creation, the creator and the datasets used. There is also a flag to specify if the application is “official” or not. This initiative seems to be isolated and by having a common layer of vocabulary for the applications, we could add more informations as it is intended in the DVIA vocabulary.

Another approach at Rensseler Institute⁴ is to put at the bottom of the static page of the demo/application showcasing the benefits of Open Data for data.gov. is to put information, containing also a link to the sparql query used for generating the application. As this information is human-readable and can help, the main drawback is the lack of a machine-readable version, using semantics. DVIA can leverage the issue aggregating many such descriptions in other Open Data initiatives as well.

Regarding the tools for visualizing Linked Data, [6] analyses in detail the current approaches used to browse and visualize Linked Data, by identifying requirements for users classify into two groups: tech-savvy and lay-users. As the authors extensively surveyed more generic Linked Data browsers, with text-based presentation and visualization options, they provide some recommendations according to the size of the data such as fine-grained analysis among others. However, they do not target their study on tools that can easily help building visual SW-based applications. However, our approach is to study the tools used to build innovative applications for detecting the components that could be reusable across different domain y/o scope.

3 Linked Data Applications

According to [5], Visualization is “*the use of computer-supported, interactive visual representations to amplify cognition*”. So the unique object of visualization is developing insights from collected data. That justify why each time a new dataset is released, users always expected some showcases to play with the underlying datasets. It is true that many public open initiatives uses incentives actions like *challenges*, *datahackday* or *contest*, etc. to find innovative applications that actually exhibit the benefits of datasets published. Visualizations play crucial role as they can easily find errors in a large collection; detect patterns in a dataset or help navigate through the dataset.

3.1 Typology

Jeni Tennison⁵ defines in her blog⁶ three categories of applications using online data:

- (i) *data-specific applications*, which are constructed around particular data sets that are known to the developer of the application; hence the visualizations obtained are of data-specific applications. Examples are the famous applications of “Where does my money go” in Greece⁷ or UK⁸. Those applications are also called “mashups”.

³ <http://data.southampton.ac.uk/apps.htm>

⁴ <http://data-gov.tw.rpi.edu>

⁵ <http://www.theodi.org/people/jeni>

⁶ <http://www.jenitennison.com/blog/node/126>

⁷ <http://publicspending.medialab.ntua.gr/en/index.php>

⁸ <http://wheredoesmymoneygo.org/>

- (ii) *vocabulary-specific applications*, which are constructed around particular vocabularies, wherever the data might be found that uses them. Examples here are FaceBook’s Social Graph API , IsaViz⁹, among others
- (iii) *generic applications*, which are constructed to navigate through any RDF that they find; e.g. Tabulator[3], OpenLink Data Explorer¹⁰

Because most mashups are data-specific applications, it is important and necessary to know what information the dataset contains. This could be achieved by giving the meaning of some properties or classes of the vocabularies used to create the dataset. Hence, what the data publisher needs to do very often is to make sure that the data they publish is documented. However, what is seen in practice, is to consider using an intuitive visualization self-descriptive to both show the added-value of the data and its documentation.

3.2 Libraries

We have surveyed a number of visualization tools - fifteen to be precise- used to build applications over data in general (raw or structured) based on the following features:

- *Data Formats* for the format of data taken as input by the tool;
- *Data Access*, for the way to access the data from the tool, such as web service, sparql endpoint, etc.
- *Language code*, the programming language used to develop the tool;
- *Type of Views*, the different views potentially accessible when using the tool;
- *Imported Libraries*, the external libraries available within the tool,
- *License* for the Intellectual Properties rights of the tool,
- *SemWeb Compliant*, whether the tool can be easily transposed or compliant with structured data; and
- *Creator*, organizations or persons who developed the tool.

The outcome of this state-of-the-art can then be used to assess in the choice of a given visualization tool, according to some criteria, such as (i) usability, (ii) visualization capabilities, (iii) data accessibility, (iv) deployment and (v) extensibility. For more details on this survey, the readers are encouraged to read [1]¹¹. Table 1 gives an overview of the selected tools studied.

⁹ <http://www.w3.org/2001/11/IsaViz>

¹⁰ <http://ode.openlinksw.com/>

¹¹ Note: The survey does not contain Visual Box as it was not released at the time of writing the survey

Tools	Data Formats	Data Access	Language Code	Type of Views	Imported Libraries	License	SemWeb Compliant	Creator
Chooasel	xls, csv	API	GWT	Text, Map, Bar chart	Time (Simile), Protovis, charts, Flexvis	Open	No	Lars Grammel
Fresnel	RDF	–	RDF	Propertie, Labels	Welkin, IsaViz, Haystack, CSS	Open	Yes	Emmanuel Pietriga et al.
Spark	RDF-JSON	SPARQL	PHP	Date chart, Pie chart, simple table	–	Open	Yes	AIFB-KIT
LDA	RDF	SPARQL	Java, PHP	–	–	Open	Yes	Talis, Epimorphis
Semantic Web Import	RDF	SPARQL CONSTRUCT	Netbeans	Graph Node Views	–	CECILL-B	Yes	Wimmics (INRIA)
Many Eyes	xls, plain text and HTML	API	Java, Flash	charts, trees, graphs, maps	–	IBM	No	IBM research
D3.js	CSV, SVG, GeoJSON	API	JavaScript	charts, trees, graphs, maps	Jquery, sizzle, colorbrewer	Open	Possible	Mike Bostock
Facet Spatial Semantic Browsing Widgets	RDF-JSON	SPARQL	JavaScript	Map, Facet view	Jquery, dynatree	Open	Yes	AKSW
Sgvizler	RDF-JSON	SPARQL SELECT	JavaScript	Map, Line chart, timeline, sparkline	Google visualization API	Open	Yes	Martin G. Skjæland
Visual Box	RDF	SPARQL SELECT	PHP, Django template	Map, charts, timeline, graphs	Google charts, TimeKnots, d3.js	Open	Yes	Alvaro Graves (RPI)
Map4rdf	RDF-JSON	SPARQL	Java, GWT	Facet, Map	OSM Layers, Google Maps	Open	Yes	OEG (UPM)
Exhibit	JSON Exhibit	Data dump	JavaScript	Map, Tile, Thumbnail, Tabular and Timeline	–	Open	Yes	MIT
Google Visualization API	JSON, CSV	API	JavaScript	Many charts, controls and dashboard	AJAX API	Open	Possible	Google
Data Publica	DSPL	API	JavaScript	Map, graph, histogram, table	Google charts, highchart.js	Proprietary	No	Data Publica
GeoAPI	GML, KML, GPX	API	JavaScript	Map views	OpenLayers, Prototype.js	Free for non commercial use	Yes	IGN (France)

Table 1. Survey of some tools used for making mashups on the Web.

3.3 On Reusable Applications

We first review the numerous applications that have been developed on top of datasets opened by governments (UK, USA, France) and public local authorities. We made a random survey of thirteen (13) applications [2] in various domains such as of security, health, finance, transportation, housing, city, foreign aid and education. The main template used was to find out:

- the name of the application;
- the scope or target domain of the application;
- a small and concise description;
- the platform on which the application can be deployed and view;
- the policy used for creating the URL of the application;
- the legacy data used to build the application, and a mention of the process of the lifting process of the raw data to RDF if available;
- the different views available of the application;
- comments or relevant drawback to mention;
- the license of the application

Table 2 provides the information extracted from `openspending` in Greece using the aforementioned template.

Features	Value
Name	Where My Money Goes in Greece
Access Url	http://publicspending.medialab.ntua.gr/
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
URL Policy	http://publicspending.medialab.ntua.gr/en/NAME-CHART.php . e.g: http://publicspending.medialab.ntua.gr/en/toppayersday.php
Data Source	http://opendata.diavglia.gov.gr ; Greek Tax data (TAXIS)
Type of views	Bubble tree, column and bar charts
Visualization tools	HighchartsJS, Bubble TreeJS JQueryJS ; RaphaelJS
License	Open
Business Value	Not Commercial (Free)

Table 2. Getting reusable information from `openspending` in Greece Application

4 DVIA: A vocabulary Describing Visualization Application

DVIA¹² is a small vocabulary aims at describing any applications developed to consume datasets in 4-5 stars, using visual tools to showcase the benefits of Linked Data. It reuses four existing vocabularies: Dublin Core terms¹³, dataset catalogue (DCAT)¹⁴, Dublin Core Metadata Initiative¹⁵ and Organization vocabulary¹⁶. It is composed of three main classes :

- **Application**: This class represents the application or the mashup developed for demo-ing or consuming data in LD fashion. It is subclass of **dctype:Software**

¹² <http://bit.ly/Vb4L8k>

¹³ <http://purl.org/dc/terms/>

¹⁴ <http://www.w3.org/ns/dcat#>

¹⁵ <http://purl.org/dc/dcmitype>

¹⁶ <http://www.w3.org/ns/org#>

- **Platform:** The platform where to host or use the application, could be on the Web (Firefox, Chrome, IE, etc..) or mobile (android, iOS, mobile) or even desktop
- **VisualTool:** Represents the tool or library used to build the application.

The diagram of the main classes and properties is depicted in Figure 1. The actual version of the vocabulary in Turtle format can be found at <http://www.eurecom.fr/~atemezin/datalift/visumodel/visu-vocab.ttl>. Listing 1.1 is a snapshot of the description of the application which won which won the Semantic Web Challenge¹⁷ in 2012, the *EventMedia Live* application, described using DVIA vocabulary. It depicts apart from some metadata about the application (dct:title, dct:name, dct:issued, dct:creator and dct:license), the different visualization libraries integrated for building Eventedia Live (e.g.: Google API, Backbone, etc), as well as the operating systems where it is designed for, the different views/facets available in the application (map, charts, graphs, force-directed layout, ect) and the heterogeneous datasets used to implement it.

```

visuapp:eventMedia01
  a dvia:Application ;
  dct:title "EventMedia Live"@en;
  dvia:description "An application for reconciling Live events with media" ;
  dvia:url <http://eventmedia.eurecom.fr> ;
  dct:issued "2012-11-10"^^xsd:date ;
  dvia:businessValue "not commercial" ;
  dvia:keyword "events, media"^^xsd:string ;
  dct:license <http://www.opendatacommons.org/licenses/pddl/1.0/> ;
  dvia:platform [ a dvia:Platform ;
    dct:title "Desktop" ;
    dvia:preferredNavigator "Google Chrome" ;
    dvia:alternativeNavigator "FireFox" ;
    dvia:system "Mac OS, Windows, Linux"^^string ] ;
  dvia:usesTool [ a dvia:visualTool; dct:title "Google visualization Tool" ;
  dct:description "Google visualization API" ;
  dvia:accessUrl <https://developers.google.com/chart/interactive/docs/reference> ;
  dvia:downloadUrl <http://www.google.com/uds/modules/gviz/gviz-api.js/> ] ;
  dvia:usesTool visuapp:visualTool02 ;
  dvia:consumes [ a dcat:Dataset; dct:title "BBC dump" ] ;
  dvia:consumes [ a dcat:Dataset; dct:title "last.fm scrapped dataset" ] ;
  dvia:consumes [ a dcat:Dataset; dct:title "upcoming scrapped dataset" ] ;
  dvia:consumes [ a dcat:Dataset; dct:title "eventful scrapped dataset" ] ;
  dvia:consumes [ a dcat:Dataset; dct:title "Flickr scrapped dataset" ] ;
  dvia:consumes [ a dcat:Dataset; dct:title "Music Brainz " ] ;
  dvia:consumes [ a dcat:Dataset; dct:title "Foursquare Json file" ] ;
  dvia:consumes [ a dcat:Dataset; dct:title "DBpedia" ] ;
  dct:creator [ foaf:mbox "khrouf@eurecom.fr"; foaf:name "Houda Khrouf" ];
  dct:creator [ foaf:mbox "vuk@eurecom.fr"; foaf:name "Vuk Milicik" ];
  dct:creator [ foaf:mbox "raphael.troncy@eurecom.fr"; foaf:name "Raphael Troncy" ];
  dvia:view "map, chart, graph, force-directed layout" ;
  ...

```

Listing 1.1. Snapshot in Turtle of the description of Event Media Live Application

The full version of this sample description is available at <http://www.eurecom.fr/~atemezin/datalift/visumodel/eventMedia-sample.ttl>. The actual version of the DVIA intends to be small enough to cover the concepts that are needed to reuse partial or full parts of applications.

5 Conclusion and Future Work

We have presented in this paper an approach that could help reusing tools and libraries in the domain of applications built on top of datasets in Linked Open Data. We first presented our motivation as to demonstrate to the end users or decision makers, there is always a need to develop a showcase, hence sometimes looking for analogous or similar applications in the area to ease the process of creation the mashup. We proceeded by surveying some applications to extract relevant common facts worth reusable, from which we proposed a small vocabulary (DVIA) to semantically describe visual applications.

We need to validate the DVIA by scrapping and reconciling data from all innovative applications websites and exposed them in RDF, starting with some of those we studied and others applications submitted at the Semantic Web Challenge the past three years. We strongly believe that having such a hub describing the tools, datasets, etc. used to build applications consuming 4-5 stars datasets will

¹⁷ <http://challenge.semanticweb.org/2012/winners.html>

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#>.

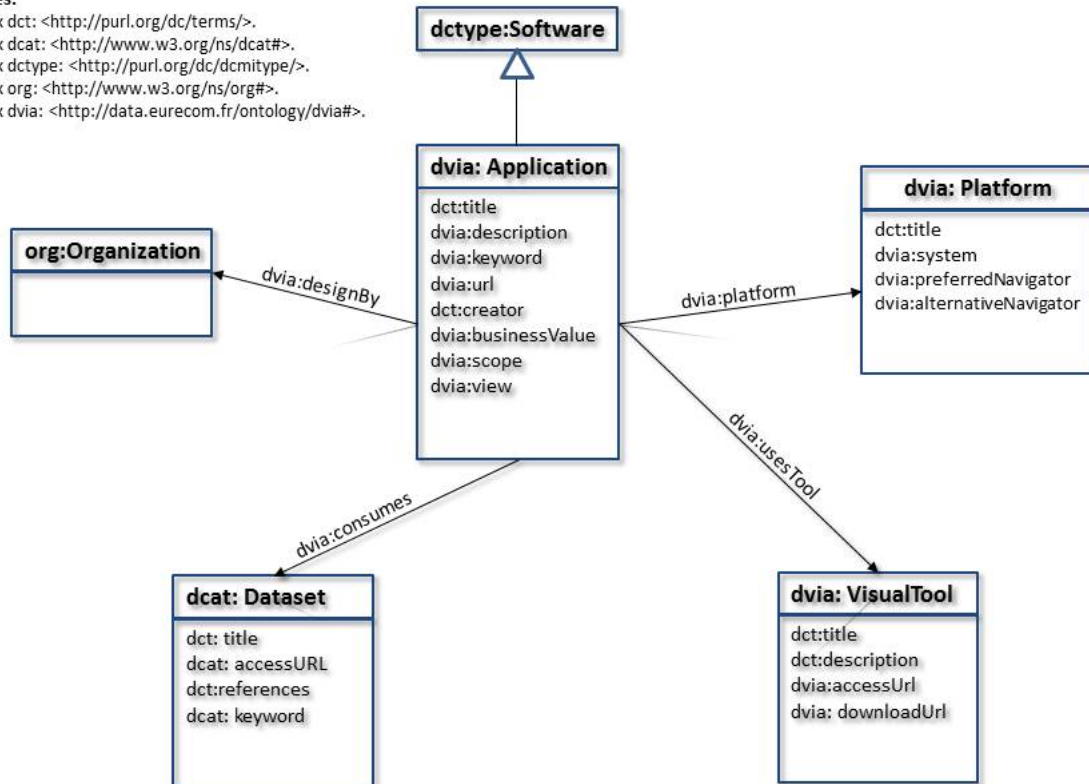


Fig. 1. Conceptual Model of the DVIA vocabulary

contribute to interoperable applications leading to a Linked Open Visualizations (LOVIZ) on the LOD cloud.

Our ultimate goal is to detect patterns used for creating Linked Data visual applications on the web using RDF data. For this challenge, we will need to perform some mining tasks of the underlying data (list of properties, number of triples, categories, etc.), the ontologies used, the templates or libraries for visualizations (Exhibit, GeoAPI, LDA, Sparkl, d3.js, etc.) and finally evaluate the efforts for a user to build the application. This could lead to a framework for generating automatically visualizations from heterogeneous datasets by extracting relevant features combined to suitable templates tools for visualizations.

Acknowledgments

This work has been partially supported by the French National Research Agency (ANR) within the Datalift Project, under grant number ANR-10-CORD-009.

References

1. G. Ateazing and R. Troncy. Tools for Visualization (v.1.2), 2012.
2. G. Ateazing and R. Troncy. Usage scenarii for Applications (v.1.1), 2012.
3. T. Berners-lee, Y. Chen, L. Chilton, D. Connolly, R. Dhanaraj, J. Hollenbach, A. Lerer, and D. Sheets. Tabulator: Exploring and analyzing linked data on the semantic web. In *3rd International Semantic Web User Interaction Workshop (SWUI)*, 2006.
4. J. M. Brunetti, S. Auer, and R. Garcia. The linked data visualization model. In *11th International Semantic Web Conference (ISWC)*, 2012.
5. S. K. Card, J. D. Mackinlay, and B. Shneiderman, editors. *Readings in information visualization: using vision to think*. Morgan Kaufmann Publishers Inc., 1999.
6. A.-S. Dadzie and M. Rowe. Approaches to visualising Linked Data: A survey. *Semantic Web*, 2(2):89–124, 2011.
7. V. Geroimenko and C. Chen. *Visualising the Semantic Web*. Springer-Verlag, London, 2006.