Visualizing Linked Open Statistical Data to Support Public Administration

Efthimios Tambouris
University of Macedonia and ITI-CERTH
156 Egnatia str., 54600
Thessaloniki, Greece
tambouris@uom.gr

Evangelos Kalampokis
University of Macedonia and ITI-CERTH
156 Egnatia str., 54600
Thessaloniki, Greece
ekal@uom.edu.gr

Konstantinos Tarabanis
University of Macedonia and ITI-CERTH
156 Egnatia str., 54600
Thessaloniki, Greece
kat@uom.gr

ABSTRACT

Open data have tremendous potential which however remains unexploited. A large part of open data is numerical and highly structured. We concentrate on those data and capitalize on linked open data (LOD) as the underlying technology. In this paper, we present a number of tools to facilitate publishing and reusing of linked open statistical data. We propose an architecture and implementation that allows developing custom visualization and analysis tools without knowledge of LOD technologies. We further present work towards deploying relevant tools in six different countries to improve decision-making and transparency and thus support public administration.

CCS CONCEPTS

•Information systems → Information systems applications;

KEYWORDS

Linked Open Data, Linked Open Statistics, Policy-Making

ACM Reference format:
Efthimios Tambouris, Evangelos Kalampokis, and Konstantinos Tarabanis. 2017. Visualizing Linked Open Statistical Data to Support Public Administration. In *Proceedings of dg.o '17, Staten Island, NY, USA, June 07-09, 2017,* 6 pages.
DOI: http://dx.doi.org/10.1145/3085228.3085304

1 INTRODUCTION

Opening up governmental data is a political priority in many countries to stimulate among others innovation and economic growth (e.g. [1]). As a result, a large number of public authorities have launched and maintain relevant portals [4]. The expected benefits of opening data are multifaceted and range from transparency to economic growth. As an example, the global annual economic potential value of Open Data has been estimated to $3 trillion [7]. However, the potential of Open Government Data (OGD) has been unrealized to a large extend [9].

The difficulty in exploiting open data seems surprising if we consider the huge importance data have in modern societies. Indeed, during the last years, businesses, academia and government employ various data analytics methods on their own data with great success. For example, business intelligence methods are employed by enterprises to help them survive in the global economy [13]. In addition, evidence based policy-making relies on data analytics to assist policy makers in producing better policies [10].

This difficulty could be explained by a number of barriers (legal, political, institutional, social, and technical) that hamper the interaction between public administration and society (citizens and enterprises). Public authorities publish open government data in an ad-hoc manner based on existing processes, according to their mandate, and often under unclear licenses. They also design and deliver services in a top-down manner. On the other hand, society has needs and data-driven public services, not raw data, can address these needs. As a result, society should be involved in service co-production to ensure that public services address their needs [12]. Societyfis data (e.g. citizen produced data, business data etc.) should also be re-used and combined with OGD in order to enable innovative public services.

In this paper, we present how Linked Open Statistical Data (LOSD) can be practically employed to overcome some of the abovementioned barriers. We concentrate here in the technological barriers in open data use.

The rest of this paper is organized as follows. In section 2, we outline LOSD. In section 3, we present a proposed software architecture as well as a number of tools we have developed to exploit LOSD. In section 4, we present six real-life scenarios where our tools have been deployed to support policy-making and the expected results. Finally, in section 5 a discussion is provided along with the conclusions and future work.

2 LINKED OPEN STATISTICAL DATA (LOSD)

A large part of OGD is of a statistical nature, meaning that consists of numeric values that are highly structured [5]. Moreover, Linked Data has been introduced as a promising technological paradigm for opening up data because it facilitates data integration on the Web. In the case of statistical data, Linked Data has the potential to realize the vision of performing data analytics on top of integrated but previously isolated statistical data across the Web [6] [8]. A fundamental step towards this vision is the RDF Data Cube vocabulary [11].

This emerging field (aka Open Statistics [3]) studies aspects of open data, data warehouses and business intelligence as well as

dg.o '17, Staten Island, NY, USA

978-1-4503-5317-5/17/06...$15.00
DOI: http://dx.doi.org/10.1145/3085228.3085304

linked open data. It provides pre-existing business intelligence capabilities on the Web but also propose new analysis capabilities that were not possible before. An example area is policy-making where integration of distributed data sources is needed to inform decision-making.

Currently, a number of pioneering organizations provide their open data as LOSD. These include national statistics offices in many countries such as the UK[1], Scotland[2], Italy[3], Japan[4], and France[5].

Eurostat also aims at addressing Open Statistics challenges. For example, DIGICOM project aims, amongst others, to facilitate automated access to European aggregate data for heavy users or re-disseminators and to improve access to micro data through linked open data.

3 LOSD TOOLS

We have developed a number of ICT tools to support publishing and reusing of LOSD. Tools for reuse are based on a proposed architecture to facilitate software development without LOSD programming skills.

3.1 LOSD publishing

The developed tools for LOSD publishing are based on existing open source tools, mainly Grafter[6]. This is a command-line application that enables creating LOSD from various file formats, such as CSV. We have developed two tools that enable creating LOSD from various file formats.

3.1.1 Table2qb and Grafter. The Table2qb tool[7], implemented with Grafter, takes data in a specific tabular structure, either as a CSV or Excel file, and converts it into an RDF Data Cube. Its functionality includes representing the data as a series of observations with dimensions, attributes and measures, and generating the associated Data Structure Definition.

3.1.2 Data Cube Builder. Data Cube Builder[8] is a tool for transforming non-RDF data sources to RDF Data Cube. It is built on top of the pre-existing tool TARQL. Data Cube Builder can be used through multiple interfaces such as desktop UI, command line, web user interface and as a web service.

3.2 LOSD reusing

LOSD tools can be used for data exploration and visualization. This includes performing classic OLAP operations, such as slicing and dicing.

In order to compare our approach to previous developments we adopt a simple software architecture that consists of three layers. The *human interface* layer is handling human-computer interaction, the *business logical* layer contains the main logic of the software and the *data access* layer is handling the communication with the data store [4]. n our case, the data store is an RDF management system,

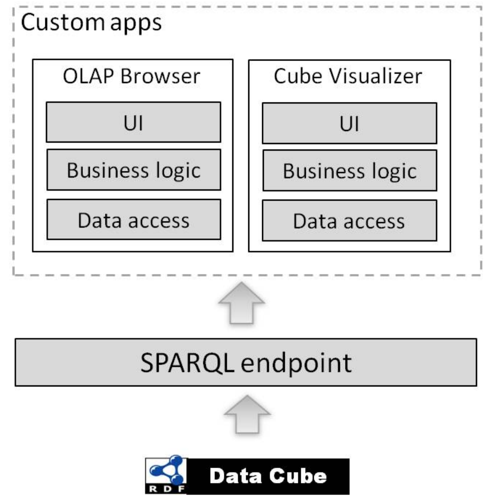

Figure 1: Traditional Architecture of LOSD tools

such as Virtuoso or Sesami, where LOSD reside. The communication with the data store is performed through a SPARQL endpoint using SPARQL, a language for managing RDF data.

In Figure 1, the traditional architecture of two tools, namely OLAP browser and Cube Visualizer, is presented. These are monolithic, vertical applications where software is developed for all three layers. These human interface and business logic layers are different in these tools. However, the data access layers are similar in both tools providing common data managing functionality. Nevertheless, these had to be coded separately leading to additional costs. More importantly, the development of the data access layers requires significant programming expertise in LOSD, a skill that is not widely available between programmers. This is a significant barrier for the development of LOSD reusing tools and the exploitation of LOSD in general.

In Figure 2, we present the same tools using an alternative layered architecture we have developed. Here, we have developed a JSON API for accessing RDF Stores in an easy and uniform way. This is based on a relevant specification that we devised. Thus, there is no more a need to implement a different data access layer for each tool as the same functionality is available through reusing the API.

As a result, it is now much easier to build software applications for reusing LOSD. Actually, all LOSD-related programming is now hidden and thus tools can be developed using only common Web programming skills, e.g. CSS and JavaScript. More technical details are outside the scope of this paper. The interested reader is advised to consult the relevant technical project reports and publications.

A short description of an indicative list of tools that we developed follows. Before that, we present some details on the JSON API Data Cube Access Specification and Implementation.

3.2.1 JSON API for Data Cube. JSON-QB API enables accessing data stored as RDF Data Cubes in a way that could be easily used by typical Web developers, i.e. programmers with JavaScript skills but without knowledge of Linked Data.

[1]http://statistics.data.gov.uk
[2]http://statistics.gov.scot
[3]http://datiopen.istat.it
[4]http://data.e-stat.go.jp/lodw/
[5]http://rdf.insee.fr
[6]https://github.com/Swirrl/grafter
[7]https://github.com/OpenGovIntelligence/table2qb
[8]https://github.com/OpenGovIntelligence/data-cube-builder

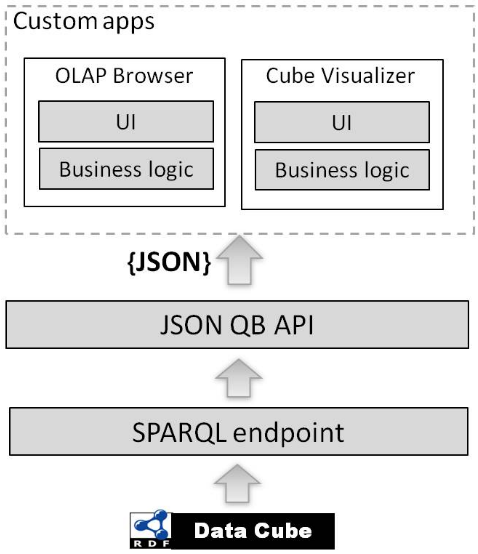

Figure 2: Proposed Architecture of LOSD tools

There is a specification[9] and an implementation[10] of the JSON-QB API. The API implementation can be installed on top of any RDF repository and offers basic and advanced operations on RDF Data cubes.

3.2.2 Data Cube Explorer. Data Cube Explorer[11] is a web-based tool that catalogues and presents details of available data cubes to the users. It also enables users to preview cube data using pivot table, cube browser and other visualization widgets.

3.2.3 OLAP Browser. The OLAP Browser[12] enables the exploration of RDF data cubes by presenting each time a two-dimensional slice of the cube as a table. OLAP Browser is based on the JSON-QB API for Data Cube implementation.

3.2.4 Cube Visualizer. The Cube Visualizer[13] is a web application that creates and presents to the user graphical representations of an RDF data cubefis one-dimensional slices. It is built as a client of the JSON-QB API implementation.

4 LOSD SUPPORT OF PUBLIC ADMINISTRATION

We have started deploying LOSD tools in six different settings across Europe, as shown in Table 1. In this section, we describe the results of the activities performed so far. These include problem definition, expected project/service and progress so far.

4.1 The Flemish Government

Citizens and companies want to compare reported and permitted emissions or emission ratios between geographical regions or companies. They also want to link emission data with population density or integrate the emission data with health data and look for correlations.

The target users of this service will be citizens as they would see what pollutants in what amounts are emitted near where they live. Finally, government is also target user of the final service as it can be assisted in revising permitted pollution levels in a certain area based on existing levels and reported emissions. Additional LOSD, R-statistics and visualization libraries are needed to develop the final service.

Figure 3 presents emission points across Flanders on a map. The user can explore locations, organizations, and detailed pollution data. These data come from five disparate datasets that are connected following the linked data principles. The Flemish service is available online at https://ontwikkel.milieuinfo.be/emissiepunten/.

Table 1: Pilot country and description

Country	Description
Belgium	Environmental permits and inspections
Estonia	Supporting data-driven real estate decisions
Greece	Management of Government Vehicles
Ireland	Marine research , technology and innovation
Lithuania	Enhancing one-stop-shop for businesses
UK	Rethinking Worklessness in Trafford

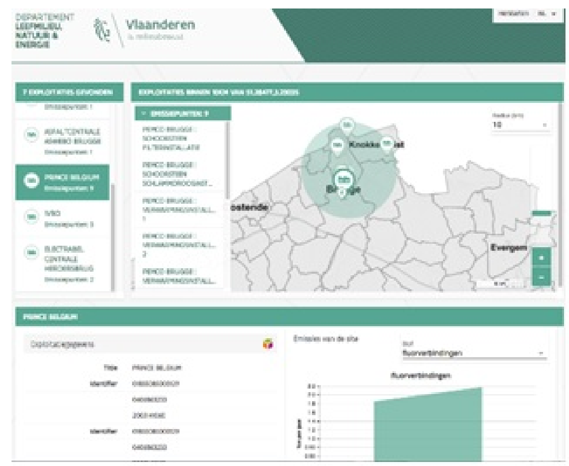

Figure 3: Map chart of Flanders

4.2 The Estonian Ministry of Economic Affairs

If someone wants to buy or rent a flat, a house, office premises or a land in Estonia, they usually need to go to real estate websites to get the general information (total area, built year, building material, ownership, etc.). If they want to know specific details or learn about restrictions concerning the flat, house, office premises or land as well as the area in which the real estate is located, they need to visit numerous different databases owned by different public authorities.

The foreseen service in this pilot will present a large amount of useful information about the flat, house, office premises or land that different types of users are interested in. This information

[9]https://github.com/OpenGovIntelligence/json-qb
[10]https://github.com/OpenGovIntelligence/json-qb-api-implementation
[11]https://github.com/OpenGovIntelligence/data-cube-explorer
[12]https://github.com/OpenGovIntelligence/qb-olap-browser
[13]https://github.com/OpenGovIntelligence/CubeVisualizer

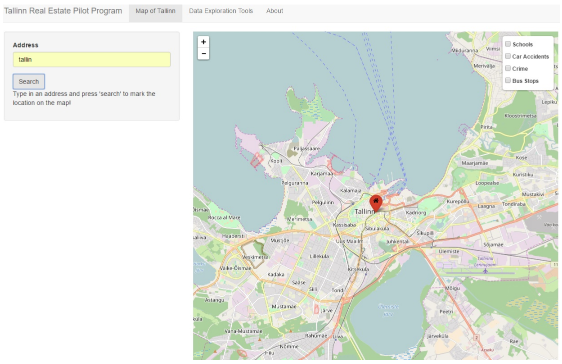

Figure 4: Map visualization of the Estonian Service

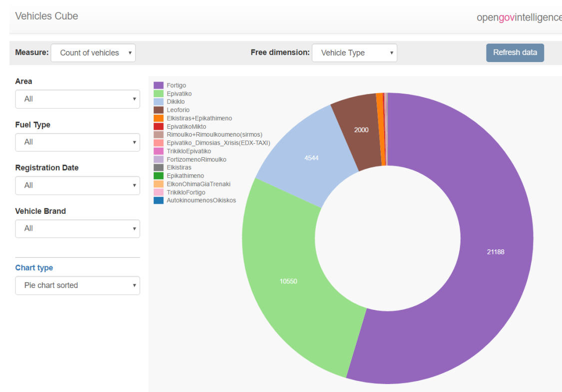

Figure 5: Pie chart showing percentage of Vehicles per Vehicle Type

would normally have to be searched for from different registers, so the main benefit would be substantial timesaving for users. The final service will show different information on a map, providing visualizations and information on one screen. The target users of this service will be all citizens who need to rent or purchase new real estate or land, real estate broker, real estate developers, investors, notaries and government officials who are responsible for urban planning and "long range developments/plans" who need information on trends to improve the area. The platform should foremost be able to visualize statistical data in one map solution. The development of this service requires opening up and linking data from disparate sources.

So far, data from businesses registry and car crashes have been transformed to LOSD. Visualization tools have been used to better understand that data. Fig. 4 presents a map of Tallinn where users can explore car accident incidents.

4.3 The Greek Ministry of Interior

The Ministry of Interior and Administrative Reconstruction is in charge of monitoring and managing an approximate number of 11.500 government vehicles, which are used by all Greek Public Agencies. The datasets it possesses originate from different sources and have not yet been properly structured and combined in order to be converted into meaningful information, which will facilitate internal decision-making and increase transparency towards the public.

The target users of the final product will be Greek Public Agencies who use government vehicles. These will use the service to obtain measures and reports on their use of government vehicles. In addition, the ministry will be able to take relevant decisions to the management of vehicles more quickly and accurately. Towards this end, organizational, cultural, institutional and legal challenges should be also identified and addressed [2].

So far, a spreadsheet with vehicles has been cleaned and transformed into LOSD while tools have been used to obtain some initial visualizations and insights. Fig. 5 presents a pie chart visualization where users can explore the different types of vehicles in relation to registration date, vehicle brand, and fuel type.

4.4 The Irish Marine Institute

In Ireland, a search and rescue operation problem is perceived as a cross-country problem of public administration, business and citizens. The rescue team wants to know the current conditions in the waters around the coastline. A member of the team wants to return information, such as geo-located photographs, to teamfis coordinator so he can be kept up to date of the search teamfis location and conditions. In addition to the public authorities, the public is involved in searching the coastline. The volunteers want to have access to the same apps and much of the same data as the authorities, but some information may not be available to them. The teamfis coordinator review the information collected by the app after each rescue to build up dataset which allows him to develop local search and rescue policies.

The final service will provide a tool in which search and rescue personnel can identify the key areas to search for a casualty in the water for rescue or recovery. This may also include onshore search parties who can provide their location and coastal imagery. The target users of this service are search and rescue services. Statistical analysis of data on entry and rescue/recovery locations, visualization of forecast model outputs, visualization of traffic situation in the bay and predictive analysis of particle tracking are the tools needed to develop the final service.

The data needed in this scenario were already in LOSD format hence this pilot concentrated on reusing tools (see Fig. 6).

4.5 The Lithuanian Ministry of Economy

Market research is a national business problem in Lithuania. Entrepreneurs in Vilnius city have no information about the opportunities and competition in the areas they want to start their businesses. They need to invest a lot of resources in order to find out if their idea has any potential. Linked Open Statistical Data (LOSD) can be used to simplify market research and decision-making process during the business planning stage.

The final service will let users navigate the Vilnius city map and see all active businesses from up to the five most popular business areas in the city. The target users of this service will be

Figure 6: Average Bar chart of Irish National Tide Gauge Network data set

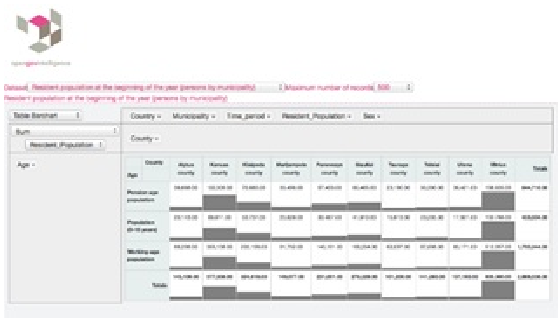

Figure 7: Bar chart showing age-related information per county

entrepreneurs, who are planning to start or expand their businesses in Vilnius city.

So far, data from various databases have been transformed to LOSD and different visualizations are possible (see Fig. 7).

4.6 UK Trafford Council

The Department for Work and Pensions (DWP) is a central government department, who maintains around 800 Job Centre Plus in England where people can claim out of work benefits, receive advice on CV-writing and interviews, and apply for jobs. The DWP is reviewing the distribution of these Job Centers and wants open statistical data to help this job done.

The final service will allow exploration of data, especially from a spatial point of view - allowing people to see where needs are greater and where there are available assets or groups who could support an alternative model of delivering Job Centre Plus services. The tool will also provide a dashboard for decision-makers to get the most up-to-date information about worklessness that they can explore, drill-through, etc. The target users of this service will be local DWP teams responsible for reconfiguring the Job Centers, in conjunction with Local Authority leads for worklessness. Ultimately, the public are the users of the Job Centre Plus.

5 DISCUSSION AND CONCLUSION

Linked Open Statistical Data (LOSD) is emerging to overcome some of the Open Data barriers. LOSD forces data structure and quality and enables data integration across the Web. Therefore, it can significant improve adoption and support in decision-making. LOSD however is not without its challenges. One of these is the lack of LOSD tools along with the significant programming expertise required for their development. This is mainly related to the need to access data stored in RDF using SPARQL.

In this paper, we propose a new architecture based on standardizing the LOSD Data Access layer of these tools. We propose and implement a specification enabling uniform and easy access to LOSD using standard APIs and JSON. Using this API, it is now easy for any Web programmer to develop custom visualizing and analysis tools for LOSD. We believe this is a significant step towards wide adoption of LOSD. Future work in this area is to standardize this specification.

In this paper, we further present a number of tools that we have used for LOSD publishing and reuse. Tools for publishing enable developing LOSD from CSV and Excel files. Tools for reuse provide browsing capabilities and OLAP operations and capitalize on the API mentioned above. Future work in this area includes improvement of existing tools (e.g. in terms of user experience) and development of more tools to cover additional analysis needs.

Finally, we present six scenarios where LOSD tools have been deployed under real-life situations. The aim is to improve internal decision-making for Public Authorities leading to better policy-making as well as to increase transparency and information provision to the public and businesses. Future work in the area includes the use of tools in actual settings as well as the evaluation of the employment of LOSD in terms of tangible and intangible benefits and problems.

ACKNOWLEDGMENTS

The work reported this paper is funded by the European Commission within the H2020 Programme in the context of the project OpenGovIntelligence (http://OpenGovIntelligence.eu) under grand agreement No. 693849.

REFERENCES

[1] Dennis A., Wixom B. H., and Tegarden D. 2002. *Systems analysis and design: An object-oriented approach with UML*. John Wiley & Sons.
[2] Chaniotaki E., Kalampokis E., Tambouris E., Tarabanis K., and Stasis A. 2017. Exploiting Linked Statistical Data in Public Administration: The Case of the Greek Ministry of Administrative Reconstruction. In *Twenty-third Americas Conference on Information Systems (AMCIS2017)*. AIS.
[3] Kalampokis E, Tambouris E, Karamanou A, and Tarabanis K. 2016. Open Statistics: The Rise of a new Era for Open Data? *Springer* EGOV2016, LNCS 9820 (2016), 31–43.
[4] Kalampokis E, Tambouris E, and Tarabanis K. 2011. A Classification Scheme for Open Government Data: Towards Linking Decentralised Data. *International Journal of Web Engineering and Technology* 6, 3 (2011), 266–285.
[5] Kalampokis E, Tambouris E, and Tarabanis K. 2016. ICT Tools for Creating, Expanding, and Exploiting Statistical Linked Open Data. *Statistical Journal of the IAOS* Preprint (2016), 1–12.
[6] Kalampokis E, Tambouris E, and Tarabanis K. 2016. Linked Open Cube Analytics Systems: Potential and Challenges. *IEEE Intelligent Systems* 31, 5 (2016), 89–92.
[7] Manyika J, Chui M, Groves P, Farrell D, Kuiken S. V, and Doshi E. A. 2013. Open data: Unlocking innovation and performance with liquid information. *Technical report, McKinsey & Company* (2013).
[8] Perez J, Berlanga R, Aramburu M, and Pedersen T. 2008. Integrating data warehouses with web data: A survey. *IEEE Transactions on Knowledge and Data Engineering* 20, 7 (2008), 940–955.
[9] Janssen M, Charalabidis Y, and Zuiderwijk A. 2012. Benefits, adoption barriers and myths of open data and open government. *Inf. Syst. Manag* 29, 4 (2012), 258–268.

[10] Davies H. T. O, Nutley S. M, and Smith P. C (Eds). 2000. What works? Evidence-based Policy and Practice in Public Services, UK. (2000).

[11] Cyganiak R and Reynolds D. 2014. The RDF data cube vocabulary: W3C recommendation. *W3C Tech. Rep.* (January 2014).

[12] Bovaird T. 2007. Beyond engagement and participation: User and community coproduction of public services. *Public administration review* 67, 5 (2007), 846–860.

[13] Jourdan Z, Rainer R. K, and Marshall T. E. 2008. Business Intelligence: An Analysis of the Literature. *Information Systems Management* 25, 2 (2008), 121–131.