

# Towards a Linked Open Statistical Data Innovation Ecosystem

Efthimios Tambouris (tambouris@uom.gr)<sup>1</sup>, Evangelos Kalampokis<sup>1</sup> and Konstantinos Tarabanis<sup>1</sup>

**Keywords:** linked statistical data, public services co-creation, W3C data qb vocabulary.

## 1. INTRODUCTION

An increasing number of governments, organisations and enterprises publish huge amounts of data as open data. The major part of this data concern statistics, e.g. data in the Open Data Portal of the European Commission (open-data.europa.eu). This data has the potential for innovative uses including performing advanced data analytics and visualizations on top of combined data that were previously isolated. From a technological point of view, statistical data can be combined by employing linked open data technologies and, specifically, the W3C RDF Data Cube vocabulary [1]. This has the potential to realise the vision of performing data analytics on top of integrated but previously isolated statistical data across the Web [2] [3].

Although several practical solutions have been developed during the last years for creating and exploiting Linked Open Statistical Data (e.g. [4] [5]), these solutions are mainly technology-driven and are not able to address the complexity and dynamics of public sector organisations and public-private collaboration with regards to (a) opening up statistical data and (b) co-producing data-driven public service.

The aim of this paper is to present the Linked Open Statistical Data (LOSD) Innovation Ecosystem that enables exploiting the full potential of LOSD. This ecosystem comprises of a framework, which includes processes, roadmaps, best practices etc., and a number of open specifications and software tools for publishing and reusing LOSD.

## 2. METHODS

In order to investigate the potential of LOSD in public service co-creation the following steps were followed.

First, we reviewed the relevant literature in the fields of open data, statistical data, linked data, data warehouses, open innovation, public service co-creation etc.

Second, we identified European cases where LOSD could provide clear benefits. Some of these cases were related to public service co-creation. These are used as the study's pilots.

Third, we conducted interviews and surveys in these pilots in order to identify needs and opportunities.

Finally, based on the above, we formulated the outline of a LOSD innovation ecosystem including principles, a framework and relevant tools.

The above-mentioned steps were performed in an iterative and incremental manner.

Future steps include deploying the ecosystem and evaluating it in the six pilots.

---

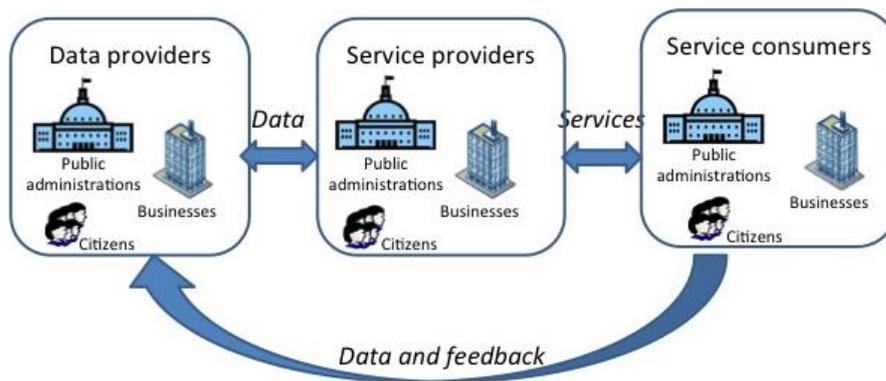
<sup>1</sup> University of Macedonia and ITI-CERTH, Thessaloniki, Greece

### 3. RESULTS

The analysis of the literature and the research conducted in the six pilots suggest there is a need for a holistic Linked Open Statistical Data (LOSD) Innovation Ecosystem. This ecosystem will inherently support the management of LOSD to facilitate public service co-creation. The ecosystem needs to be supported by an appropriate framework (meaning a set of processes, rules, roadmaps, best practices etc.) and software tools for LOSD management.

#### 3.1. LOSD Innovation Ecosystem

The proposed user-centric Linked Open Statistical Data (LOSD) Ecosystem enables the active collaboration of public administration and society (Figure 1).



**Figure 1. LOSD Innovation Ecosystem**

The roles that the stakeholders (PAs, businesses, citizens, NGOs etc.) can have in the ecosystem are: Data provider, Public service provider and Public service consumer. According to our approach, every stakeholder can have any of these roles (Table 1). Public authorities, businesses and citizens could provide data that would potentially support the development of public services. In the case of public authorities these data are public sector information provided as open data through governmental portals. Businesses could provide private data that they collect or produce (e.g. insurance companies could provide data about stolen bikes) for the development of new services, while citizens could provide information about their activities in appropriate platforms that collect such information.

**Table 1. Roles and Stakeholders of the LOSD Innovation Ecosystem**

	<b>Data Provider</b>	<b>Public Service Provider</b>	<b>Service Consumer</b>
<b>PAs</b>	Provision of Open Government Data	Design and deliver of public service Provide public services	In policy making and/or internal decision making
<b>Businesses</b>	Business data (private) to be used in services	Co-design and/or co-deliver of public service	In business intelligence, decision making etc.
<b>Citizens/ NGOs</b>	Citizen provided data	Co-design and/or co-deliver of public service	Information provision, transparency etc.

As regards public service production (i.e. design and delivery) this can be done either by only public administration in a top down manner or by involving businesses or/and citizens in a service co-production process. Finally, the consumers of public services could potentially regard all stakeholders. Towards this end, public services that would facilitate (a) policy-making and internal PA decision-making, (b) business decision-making, and (c) better informed citizens are all possible.

### 3.2. LOSD Framework

The proposed framework takes a holistic view on LOSD-driven public service innovation looking at three sides of innovation: first the content (what is a data-driven/LOSD-driven public service, what role does co-creation play in developing this service, what are the core components of the data and service architecture), then the context in which innovation takes place (what factors drive or constrain this innovation, who are the actors) and finally the process of innovation (what steps and how should be taken in order to realize a (new) LOSD-driven public service). In the full version of the paper, the main elements of the LOSD framework will be presented.

### 3.3. LOSD Specifications and Tools

The LOSD Innovation Ecosystem should be supported by a number of open specifications and tools that capitalise on linked open data technologies and the W3C datacube vocabulary. The most important include:

JSON QB API specification for Data Cube: This aims to facilitate the reuse of data stored as RDF Data Cube in a way that would be easy for developers to use. The main aim is people with moderate Javascript skills, but no knowledge of Linked Data.

JSON QB API implementation: This tool implements the JSON QB API specification. It aims to provide an easy to use API for developers that reuse statistical data stored in the form of RDF Data cubes. The API implementation can be installed on top of any RDF repository and offer basic and advanced operations on RDF Data Cubes.

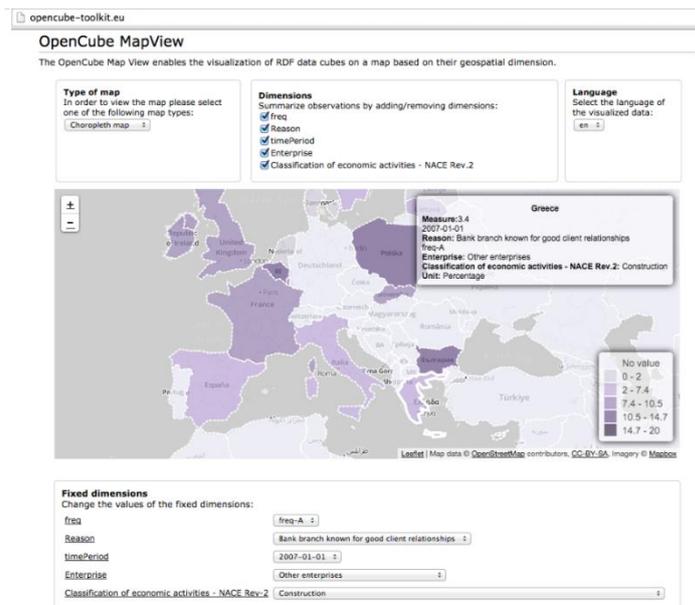


Figure 2. Screenshot of a Visualisation tool

Spreadsheet to Data Cube convertor: This aims to develop a spreadsheet-based template and corresponding transformation to allow data publishers to publish data in RDF Data Cube format. This tool extends the open source Grafter tool (grafter.org).

Data Cube Builder: This is built on top of TARQL and enables data publishers to publish RDF Data cubes by transforming non RDF formats to RDF Data cubes.

Data Cube Explorer: This is a Web-based tool that allows data analysts to reuse existing data stored as RDF Data cubes by presenting visualizations and metadata (e.g. available data cubes characteristics) and executing basic pivot operations.

OLAP Browser: This is a Web-based tool that enables data analysts to reuse existing data stored as RDF Data cubes by presenting a two-dimensional slice of the cube as a table and enabling OLAP operations (roll-up, drill-down, pivot, dimension reduction etc.)

During the conference's paper presentation some of these tools will be demonstrated to the participants (Figure 2). A relevant workshop will be also planned.

#### **4. CONCLUSIONS**

A large number of open data concern statistics. Open statistical data from multiple sources (e.g. open government data portals) can be combined by employing linked data technologies. The resulting Linked Open Statistical Data (LOSD) can be used for many different purposes. An increasingly important aim for the public sector is public service co-creation. In this paper, based on reviewing the literature and surveying six European cases, we present a holistic ecosystem for LOSD innovation. This ecosystem comprises of a framework, which includes processes, roadmaps, best practices etc., and a number of software tools for managing LOSD.

#### **ACKNOWLEDGEMENTS**

This work 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] R. Cyganiak and D. Reynolds, The RDF data cube vocabulary: W3C recommendation, W3C, Tech. Rep., January 2014.
- [2] A. Abello, J. Darmont, L. Etcheverry, M. Golfarelli, J.-N. Mazón, F. Naumann, T. Pedersen, S. B. Rizzi, J. Trujillo, P. Vassiliadis, and G. Vossen, Fusion cubes: Towards self- service business intelligence, 9, 2, (2013), 66–88.
- [3] J. Perez, R. Berlanga, M. Aramburu and T. Pedersen, Integrating data warehouses with web data: A survey, IEEE Transactions on Knowledge and Data Engineering, 20, 7, (2008) 940–955.
- [4] E. Kalampokis, E. Tambouris and K. Tarabanis, Linked Open Cube Analytics Systems: Potential and Challenges, IEEE Intelligent Systems, 31, 5, (2016) 89-92.
- [5] P. E. R. Salas, F. M. Da Mota, K. K. Breitman, M. A. Casanova, M. Martin and S. Auer, Publishing statistical data on the web, International Journal of Semantic Computing, 6, 4, (2012), 373–388.