Modeling, Fusion and Exploration of Regional Statistics and Indicators with Linked Data tools

Valentina Janev[1], Vuk Mijović[1], Dejan Paunović[1], Uroš Milošević[1]

[1] Mihajlo Pupin Institute, University of Belgrade, Serbia
{valentina.janev, vuk.mijovic, dejan.paunovic, uros.milosevic}@pupin.rs

Abstract. This paper contributes to the understanding of challenges related to publishing and consuming public sector information using Linked Data tools. Linked Data paradigm has opened new possibilities and perspectives for the process of collecting and monitoring socio-economic indicators. Due to multi-dimensionality of the statistical data, in order to ensure efficient exploration and analysis, hierarchical data structures are needed for modeling the space and time dimensions. This paper presents several illustrative examples of modeling, analyzing and visualization of Linked Data from Serbian government bodies. The approach utilizes tools from the *Linked Data* stack, as well as the first prototype of the *Exploratory Spatio-Temporal Analysis* component that has been developed in the GeoKnow project framework.

Keywords. Linked Data, integration, modelling, statistics, transparency

1 Introduction

In the last five years, the global Open Government Data (OGD) initiatives, such as the Open Government Partnership[1], have helped to open up governmental data for the public, by insisting on non-sensitive information, such as core public data on transport, education, infrastructure, health, environment, etc. Additionally, the European Commission (EC) has made considerable investments to improve efficiency in the provision of public services, increase transparency, define better strategies for delivering large amounts of trusted data to the public and improve interoperability. As part of alignment of national government activities with the EC 'Interoperability Solutions for European Public Administrations' (ISA) program[2] for the period from 2010-2015 [1], a lot of government portals have been established across Europe that allows different stakeholders to publish standards, guidelines and interoperability assets im-

[1] http://www.opengovpartnership.org/
[2] http://ec.europa.eu/isa/

portant at the national level. This openness, on one hand, will strengthen European democracy and promote efficiency and effectiveness in G2C, G2G, G2B services, while, on the other, it will expand creative use of data beyond its walls by encouraging application of innovative technologies such as semantic mash-ups and widgets.

This paper aims at discussing the challenges related to the possibilities of integrating open data for advanced analyses of socio-economic indicators in official government statistical systems. Furthermore, by using an illustrative case study of the Register of the Regional Development Measures and Incentives[3], maintained by the Serbian Business Registers Agency (SBRA), it demonstrates the adaption of the *Linked Data* stack for analysis and dissemination of official statistics.

The article is organized as follows. First, Section 2 introduces the main challenges of integration and analysis of open data. Sections 3 and 4 present the Linked Open Data (LOD) approach to modeling, fusion and analysis of spatial-temporal data and statistics. Some illustrative examples of visualization of statistics with the *Exploratory spatio-temporal analysis for Linked Data* (ESTA-LD)[4] component are given in Section 5. Finally, Section 6 concludes the paper and outlines the future work.

2 Sharing and reusing public datasets across Europe

2.1 Related work

Linked Data paradigm [2] has been utilized recently in order to achieve linking of datasets together through references to common concepts. The standard for the representation of the information that describes those entities and concepts is RDF [3]. Several projects have been financed within the EU FP7 research program devoted to publishing and consuming data in Linked Data format. As a result, several repositories of open source toolkits, as well as platforms for building Linked Data applications have emerged recently, as presented in Table 1.

In the last three years, in the framework of the LOD2 and GeoKnow projects, the Institute Mihajlo Pupin (IMP) has been involved in maintaining the *Linked Data* stack[5], an integrated set of tools for managing the Linked Data life-cycle. All tools are RDF based and enable developers to build custom applications on the top of the public sector data. For instance, IMP was involved in development of a specific web interface (*Statistical Workbench*[6]) that aggregated several components of the stack organized in an intuitive way to support the specific business context of the statistical office. The *Workbench* contains several dedicated extensions for manipulating RDF data according to the Data Cube vocabulary: validation, merging and slicing of cubes.

[3] http://www.apr.gov.rs/eng/Registers/RegionalDevelopmentMeasuresandIncentives.aspx

[4] *ESTA-LD* is an exploratory spatial-temporal analysis tool, see early prototype http://fraunhofer2.imp.bg.ac.rs/esta-ld/

[5] http://stack.linkeddata.org/

[6] http://fraunhofer2.imp.bg.ac.rs/lod2statworkbench

In the GeoKnow framework, the *GeoKnow Generator*[7], an integrated solution for managing geospatial data, has been developed. Within the ESTA-LD development, IMP will further extend the capabilities of the *GeoKnow Generator* for managing spatio-temporal data.

The ESTA-LD development is driven by the GeoKnow requirements and by the features currently not supported by the GeoKnow components Facete [4] and CubeViz [5] such as visualization of statistics on a geographical map.

Topic	**Project**	**Tool**
Establishing and maintaining data and tool catalogues / repositories	LATC	LATC Data Publication & Consumption Tools Library, http://wifo5-03.informatik.uni-mannheim.de/latc/toollibrary/ and *LATC 24/7 Interlinking Platform*, see http://latc-project.eu/platform
	LOD2	PublicData.eu catalog Linked Data Stack
	PlanetData	PlanetData Tool Catalogue,see http://planet-data.eu/planetdata-tool-catalogue
	ENGAGE	ENGAGE Data, http://www.engagedata.eu/
Monitoring of Open Data Catalogues	OPENDATA-MONITOR	Harvesting framework, http://project.opendatamonitor.eu/
Harmonizing open data	HOMER	HOMER Federated Index, http://homerproject.eu/docs/Federation_technical_specific_CSI_V04.pdf
Managing Linked statistical data	LOD2	LOD2 Statistical Workbench
	OpenCube	http://opencube-project.eu/
	CEDAR	Census dataset, http://www.cedar-project.nl/
Managing Linked geo-spatial data	GeoKnow	GeoKnow Generator, http://geoknow.eu

2.2 Problem statement

In order to share datasets between users and platforms, the datasets need to be accessible (regulated by licence), discoverable (described with metadata) and retrievable (modelled and stored in a recognizable format). According to the *World Bank Group* definition "Data is open if it is technically open (available in a machine-readable standard format, which means it can be retrieved and meaningfully processed by a computer application) and legally open (explicitly licensed in a way that permits commercial and non-commercial use and re-use without restrictions)."

As a part of OGD initiatives, a large number of data catalogues have been established across Europe that store descriptions of the public datasets. A *data catalogue* can be viewed as an electronic library index that structures descriptions (meta-data)

[7] http://generator.geoknow.eu

about the actual data. An example of the data catalogues is the *PublicData.eu* (http://PublicData.eu), a Pan European data portal that provides access to open, freely reusable datasets from local, regional and national public bodies across Europe. Although OGD policies have spread fast (the number of the open data initiatives has grown from two to over 300)[6], the availability of truly open data remains low, with less than 7% of the dataset surveyed in the Open Data Barometer published both in bulk machine-readable forms, and under open licenses [7].

In order to illustrate the possibilities of sharing and re-using public datasets across Europe, in Figure 1, we have presented three types of the open data portals:

- Open Data Portal in Country X contains data and metadata descriptions, but does not provide DCAT[8] support for harmonization of portal/catalog with similar data portals/catalogues. Publishers can use the cross country portals for publishing data, e.g. the Engage portal[9],
- Open Data Portal in Country Y contains data and metadata descriptions, as well as a Linked Data SPARQL endpoint, but it is isolated , that is, it is not integrated at the international level
- Open Data Portal in Country Z is CKAN based, meaning that it can be easily harvested by other metadata catalogs at the international level.

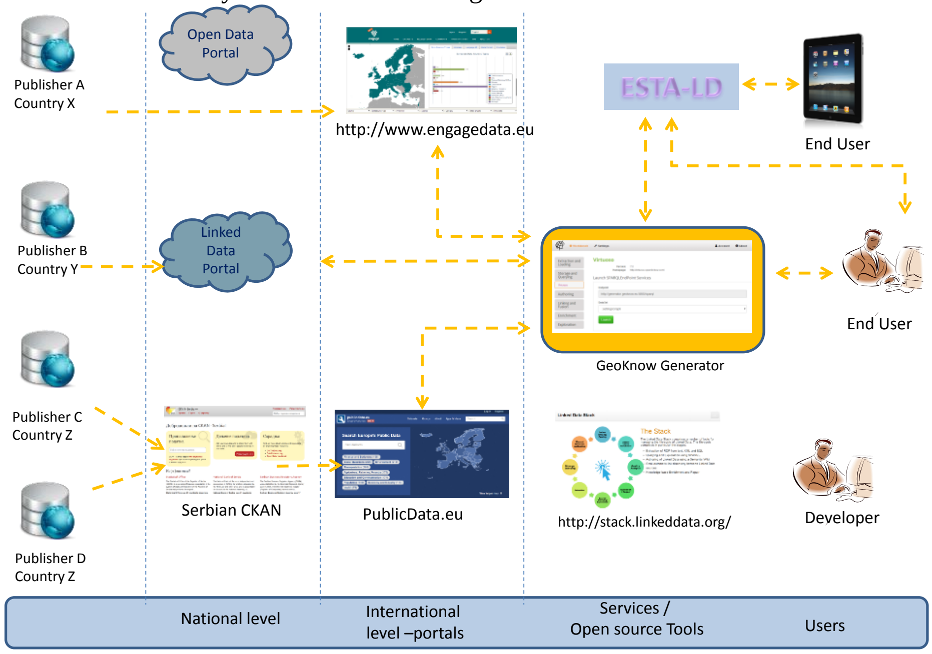

Fig. 1. Exploration and consumption of open data

---

[8] DCAT, http://www.w3.org/TR/vocab-dcat/, is an RDF vocabulary designed to facilitate interoperability between data catalogs published on the Web.

[9] http://www.engagedata.eu/

The data can be exposed for exploration in different ways, e.g. via a SPARQL endpoint or downloadable dump file. Although there are best practices for publishing open (and linked) data[10], the metadata can be of low quality leading to the questions such as:

- Is the open data ready for exploration? Is the metadata complete? What about the granularity? Do we have enough information about the domain/region the data is describing?
- Is it possible to fuse heterogeneous data and formats from different publishers and what are the necessary steps? Are there standard services for querying government portals?

Our initial analyses have shown that automatic processing and consuming open data for building smart public service is hampered by many factors including

- lack of standard approaches for querying government portals;
- quality of metadata (e.g. incomplete description of the public datasets, changes on schema level);
- reliability and completeness of the content of public datasets (e.g. uptodateness of open data);
- heterogeneity of formats used by different publishers.

The vision for ICT-driven public sector innovation refers to the use of technologies for the creation and implementation of new and improved processes, products, services and methods of delivery in the public sector [8], see right side of Figure 1. However, in the light of the above statements, innovative application on top of open data are needed that will provide some insights about the meaningfulness of integrating data from different domains. Smart public services are needed that make the use of open data more efficient and less time-consuming.

2.3 Case study: Register of the Regional Development Measures and Incentives

Statistical data are often used as foundations for policy prediction, planning and adjustments, having a significant impact on the society (from citizens to businesses to governments). In the GeoKnow framework, we have studied the possibilities for adoption of Linked Data technologies for automating the integration and enhancing the analysis of data from the Serbian Register of the Regional Development Measures and Incentives. The Register is a unique, centralized electronic database of the taken measures and implemented incentives that are of significance for regional development. In the case of regional development policies complex indicator systems are required due to the multiplicity of objectives, large number of projects (investments) in different domains and involvement of different level of government (local, regional, national, European) [9], [10]. Instead of the current way of collecting and visualizing data, innovative approaches could be applied that:

[10] http://www.w3.org/TR/ld-bp/

- extend the existing public service provided by SBRA and standardize the data collection process;
- ensure interoperability (e.g. integration of data from the Register with data from the Dissemination database of the Statistical Office of the Republic of Serbia, SORS) and improve transparency of data on European level;
- allow advanced spatio-temporal analysis of available indicators;
- offer user-friendly services in a secure and flexible manner allowing personalisation for different types of users (public administration, businesses and citizens).

Table below compares the basic characteristics of several example datasets published in Linked Data format by SORS and SBRA. Datasets can be explored via the Statistical Workbench interface [11].

Table 1. Basic characteristics of spatio-temporal data published by two Serbian government agencies

Publisher	Domain	1st dimension: Space Granularity levels	2nd dimension: Time period, granularity	Other dimensions
SBRA	Regional development measures and incentives by purpose	Country-region-area-municipality	2010-2013, yearly	Purpose type
SBRA	Regional development measures and incentives by financial type	Country-region-area-municipality	2010-2013, yearly	Financial type
SORS	Demographic statistics, Population projections	Country-region	2009-2010, yearly	Sex, age
SORS	Labour market	Country-region	2008-2013, yearly-monthly	Sex, age, occupation, type of settlement
SORS	Structural business statistics	Country-region	2007-1011, yearly	Enterprise type, NACE activity
SORS	Tourism (overnight stay)	Country-region	2010-2013, yearly-monthly	Tourist type

All indicators mentioned in the examples above have attributes related to both space and time, thus imposing challenges for visualizing both dimensions on a geographical map. Moreover, these datasets are often multi-dimensional originally, meaning that the information can be represented on different granularity levels in space and time, as well as the type of information (different attributes, see last column). In Figure 2, for instance, the distribution of incentives in different geographical entities in Serbia on different granularity levels is presented.

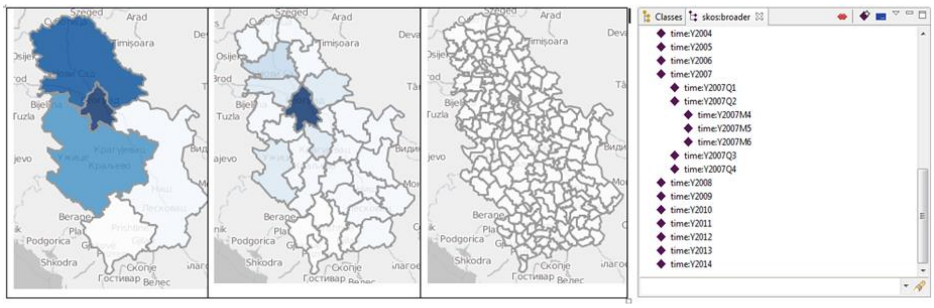

Fig. 2. Exploration and consumption of open data

3 Linked Data approach to modeling spatio-temporal data

A statistical data set comprises a collection of observations made at some points across some logical space. The collection can be characterized by a set of dimensions that define what the observation applies to (e.g. time, country) along with metadata describing what was measured (e.g. economic activity, prices), how it was measured and how the observations were expressed (e.g. units, multipliers, status). We can think of the statistical data set as a multi-dimensional space, or hyper-cube, indexed by those dimensions. This space is commonly referred to as a *cube* for short; though the name shouldn't be taken literally, it is not meant to imply that there are exactly three dimensions (there can be more or fewer) nor that all the dimensions are somehow similar in size.

3.1 Modeling statistical data (vocabulary RDF Data Cube)

In January 2014, W3C recommended the *RDFData Cube* vocabulary [12], a standard vocabulary for modeling statistical data, see http://www.w3.org/TR/vocab-data-cube/. The vocabulary focuses purely on the publication of multi-dimensional data on the Web. The model builds upon the core of the *SDMX 2.0 Information Model* [13]. In 2001, the Statistical Data and Metadata Exchange (SDMX[11]) Initiative was organised by seven international organizations (BIS, ECB, Eurostat, IMF, OECD, World Bank and the UN) to release greater efficiencies in statistical practice. An example how to use the vocabulary to represent one single observation is given in Figure 4. The observed socio-economic phenomenon is described using several dimensions modelled with the SKOS vocabulary.

In this example we have a *coarse-grained representation of the indicator* “Tourists arrivals” for the territory of the country of Serbia (`geo:RS`) and for year 2005 (`time:Y2005`). Additionally, the indicator represents the “Total” number of tourists including “Domestic” and “Foreign”.

[11] http://www.sdmx.org/

```
http://elpo.stat.gov.rs/lod2/RS-DATA/Tourism/Tourists_arrivals/data/obs1>
        a qb:Observation ;

        rs:geo geo:RS ;
        rs:time time:Y2005 ;                              SKOS Dimensions
        rs:dataType datatype:number ;
        rs:obsIndicator "Tourists arrivals" ;
        rs:obsTurists "Total" ;

qb:dataSet <http://elpo.stat.gov.rs/lod2/RS-DATA/Tourism/Tourists_arrivals/data> ;

sdmx-measure:obsValue "1988469" .
```

Fig. 3. Example of using the RDF Data Cube vocabulary

3.2 Modeling spatial and time dimensions (using SKOS vocabulary)

In order to formalize the conceptualization of hierarchical dimensions (space, time), we can use the *Simple Knowledge Organization System* (SKOS), see http://www.w3.org/TR/2005/WD-swbp-skos-core-spec-20051102/. SKOS Core is a model and an RDF vocabulary for expressing the basic structure and content of concept schemes such as thesauri, classification schemes, subject heading lists, taxonomies, 'folksonomies', other types of controlled vocabulary, and also concept schemes embedded in glossaries and terminologies.
Concepts represented as `skos:Concept` are grouped in concept schemes (`skos:ConceptScheme`) that serve as code lists from which the dataset dimensions draw on their values. Semantic relation used to link a concept to a concept scheme is `skos:hasTopConcept`. Herein, we will present an example of coding the space and time dimension in RDF. SKOS properties `skos:broader` and `skos:narrower` can be used for relating concepts of the same type, in our case, geographical area (`geo:Region`). However, if the concepts are not of the same type (e.g. to regions and municipalities), the `skos:related` alignment can be applied.

```
geo:RS21
    rdf:typegeo:Region ;
    owl:sameAs
      <http://dbpedia.org/page/%C5%A0umadija_and_Western_Serbia
> ;
    skos:broadergeo:RS ;
    skos:narrower geo:RS212 , geo:RS216 , geo:RS211 , geo:RS215
, geo:RS213 ;
    skos:narrower geo:RS218 , geo:RS214 , geo:RS217 ;
    skos:notation "RS21"^^xsd:string ;
    skos:prefLabel "Region of Sumadija and Western Serbia"@en ,
"REGION ŠUMADIJE I ZAPADNE SRBIJE"@sr-rs .
```

The observed data can be described with time information using different formats (e.g. seconds from the begging of an event, day-time, day, month, year). One way to specify the data frequency (or time granularity) in a dataset is to use the SDMX Content-oriented guidelines[12]. A specific time period (e.g. January in year 1980) can be coded as following:

```
time:Y1980M1
rdf:type time:P1M ;
skos:broader time:Y1980Q1 ;
skos:notation "Y1980M1"^^xsd:string ;
skos:prefLabel "1980/january"@en .
```

3.3 Fusing statistical data (RDF Data Cubes)

Data fusion is the process of integration of multiple data and knowledge representing the same/similar real-world object (e.g. socio-economic indicators for a region) into a consistent, accurate, and useful representation. Data fusion makes sense if the observations are on the same granularity level (described with the same number of dimensions), and are disjoint, meaning there are not two observations with the same values for all dimensions. However, if the tool for exploring statistical data integrates drill-down functionalities, datasets with different granularity levels can be used for merging.

Depending on the data sources and the alignment of concept schemes used in input datasets, different fusing algorithms are needed, e.g.:

- Same concept schemes are used in all input datasets.
  In this case the merging approach is rather straightforward: the user selects a reference DSD and the observations from all merged cubes are copied directly into the new cube. After the fusion step, check for validation of the integrity constraints can be run [14].
- Some input datasets do not use concept schemes.
  In this case, it is often necessary to perform additional harmonization of the concept schemes. One alignment approach is to compare the values of the properties that are used with identified concept schemes, and to consider that properties draw values from the existing concept schemes.
- Input datasets use different concept schemes.
  In this case, vocabulary matching algorithm for alignment of concept schemes is needed as a prerequisite for the fusion.

4 Linked Data approach to analysis of spatial-temporal data

Exploratory data analysis (EDA) is an important first step in any data analysis that focuses on identifying general patterns in the data, and identifying outliers and features of the data that might not have been anticipated. The goal of the spatial-temporal

[12] http://sdmx.org/wp-content/uploads/2009/01/02_sdmx_cog_annex_2_cl_2009.pdf

analysis is to derive information from data using visualization techniques and standard EDA techniques including histogram, run chart, or scatter plot / scatter graph.

The selection of techniques for exploratory analysis of spatio-temporal Linked Data depends on the characteristics of data and the goals of analysis. Figure 4 presents a semi-automatic analysis process of socio-economic indicators that is currently supported by the Linked Data tools in the following way:

1. The user uploads a graph into the local RDF data store (Virtuoso);
2. The user checks for validation of the integrity constraints [14]. If errors are identified, automatic or semi-automatic repair of multidimensional model is available;
3. The user can proceed with the exploration using Facete, CubeViz or ESTA-LD. Facete provides faceted navigation of the model and visualization of points of interest (described with *WGS84*[13] vocabulary) on a geographical map. CubeViz is an *RDF Data Cube*[14] model exploration tool without possibilities to relate the indicators with geographical entities.
4. When launching the ESTA-LD component, the user specifies the SPARQL endpoint, and selects the graph that contains the data to be explored. The data is then retrieved from the specified SPARQL endpoint and visualized on the choropleth map. The choropleth map provides an easy way to visualize how measurement varies across a geographic area. It is an ideal way to communicate spatial information quickly and easily, since the data is aggregated or generalized into classes or categories that are represented on the map by grades of colour. The ranges of data values for different colors are recalculated every time a new set of data is retrieved from the SPARQL endpoint. After the data is retrieved, the user can utilize different filtering options that are currently implemented (see Fig. 5):

- For selecting values from the time dimension,
- For selecting the indicator under study,
- For selecting the granularity level for the space dimension,
- Interactive selection of the area of interest on the geographical map (for the selected area a bar-chart or histogram (see Fig. 6) representation of the indicator is displayed).

[13] http://www.w3.org/2003/01/geo/
[14] http://www.w3.org/TR/vocab-data-cube/

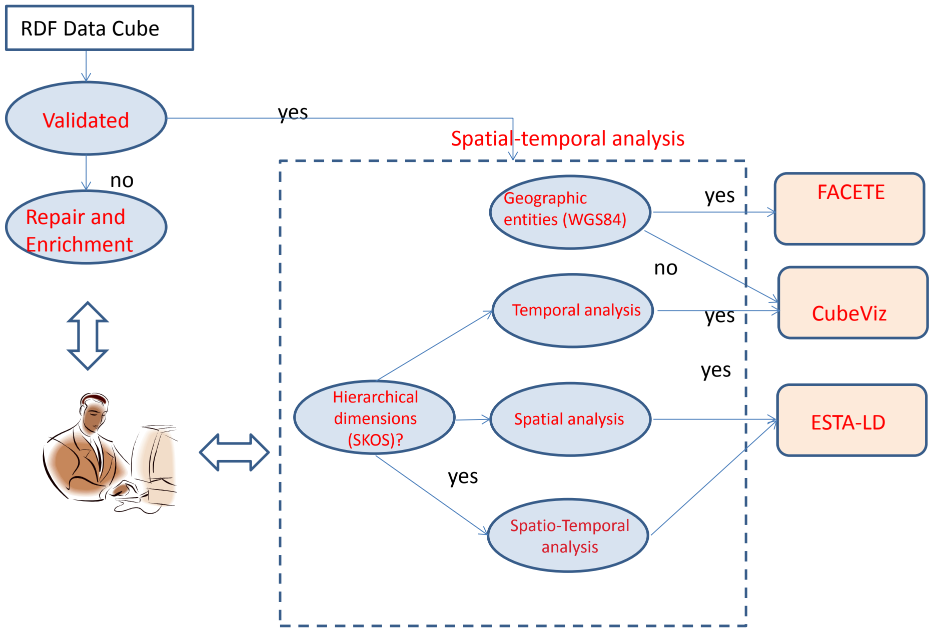

Fig. 4. Guided analysis of spatio-temporal data

5 ESTA-LD Implementation details and validation

ESTA-LD aims at overcoming the limitations of CubeViz and Facete and offering generic interface for exploring spatio-temporal data. In order to enable ease of integration with the *GeoKnow Generator*, ESTA-LD component was developed using HTML5 and JavaScript. Representation and interaction with geographic information was implemented using *Leaflet*, an open source JavaScript library for mobile-friendly interactive maps, see http://leafletjs.com/. The geographic data (such as region borders), originally created from shape files, is stored in GeoJSON format. It is brought in and programmed with JavaScript and added to maps to create interactive visualizations. On the other side, different statistical indicators, which are the subjects of the spatiotemporal analysis, are stored in the RDF Data Store. This data repository is accessed and queried using SPARQL query language. The actual retrieval of data from the SPARQL Endpoint is implemented using the jQuery library and its standard getJSON function. Finally, the results of the spatiotemporal analysis are visualized using *Highcharts*, a charting library written in pure HTML5/JavaScript, offering intuitive, interactive charts to a web site or web application see http://www.highcharts.com/.

To validate our approach, we used data from two different institutions - the National Statistical Office and the Business Registers Agency of Serbia. Data (aggregated, interlinked with public datasets and fused) appear at different levels of granularity.

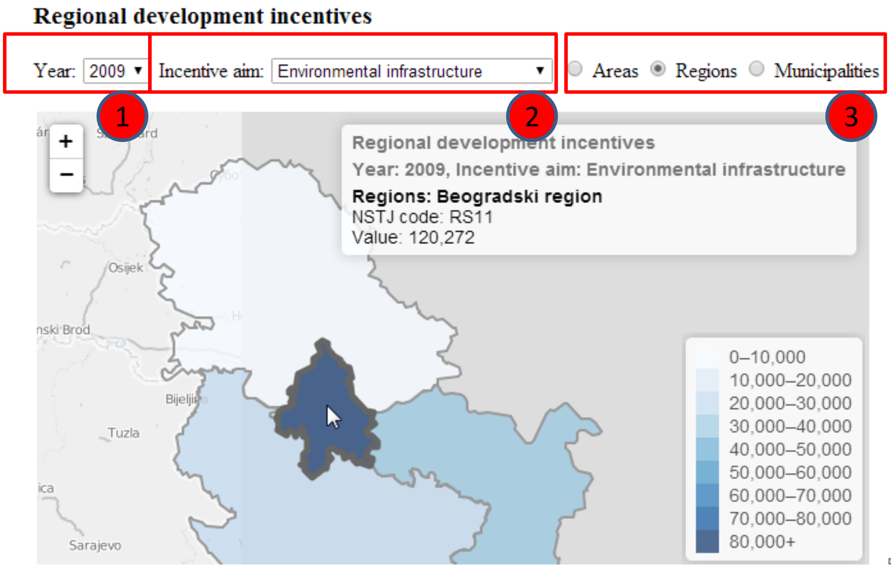

Fig. 5. Spatial-temporal filtering options

Our initial testing has shown that the first prototype of the ESTA-LD tool, although still under development, proved to be a valuable instrument for advanced spatio-temporal analysis of Linked Data. The whole geospatial information life cycle in the resulting first prototype for exploratory spatio-temporal analysis will be further tested with data from different countries and different statistical domains.

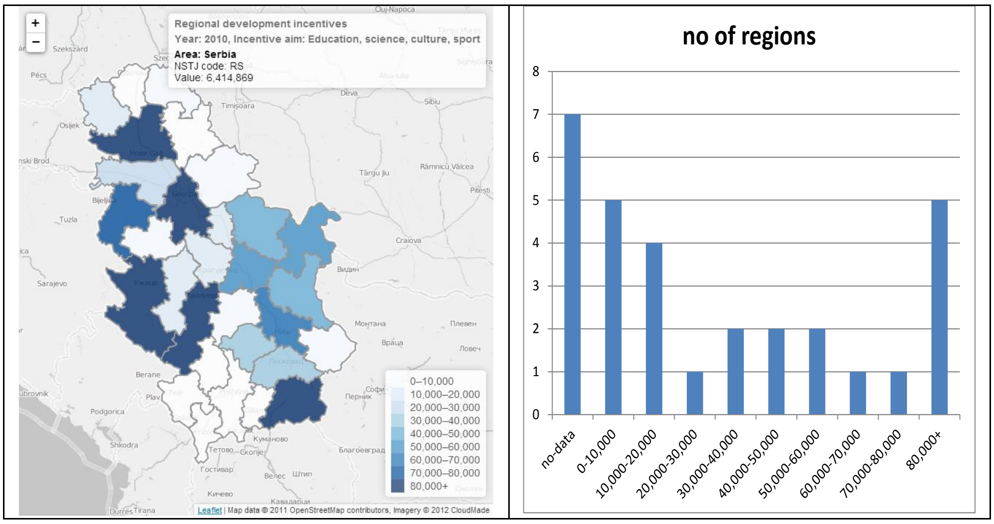

Fig. 6. Analysis of incentives (a. on a geographical map, b. histogram of a continuous variable)

6 Conclusions and further work

The EC ISA program envisions publishing of the public/private datasets in machine readable format, thus, making sharing, using and linking of information easy and efficient. The datasets will be offered through standardized data catalogues while standardized vocabularies (eGovernment core vocabularies, *WGS84*, *RDF Data Cube*) and ontologies will be used for data modelling. Taking into consideration the ISA recommendations, this paper discusses the challenges of integration and analysis of the open data and introduces the Linked Data approach to modeling and fusion of spatial-temporal data and statistics. To validate the approach, the 'Register of the Regional Development Measures and Incentives' case study has been elaborated and a new tool ESTA-LD has been developed.

Although, there are tools for visualization of Linked Data including geo-spatial data, this paper points to the need for guided analysis of the Linked Data retrieved from arbitrary SPARQL endpoint and proposes the ESTA-LD component as a generic interface for exploring spatio-temporal data. The ESTA-LD component contributes to further development and standardization of the Linked Data technologies. It shows the use of emerging W3C standards (RDF Data Cube vocabulary, SKOS vocabulary) in the field of statistical and temporal geospatial information management.

In future, a significant effort will be put into further generalization of the ESTA-LD filtering and exploration and integration of the component into the *GeoKnow Generator*. Challenges that will be addressed are generalization of the ESTA-LD filtering and exploration, so that the component supports different spatial and temporal SKOS based conceptual schemes. Finally, the tool will be further enhanced taking into consideration different aspects, such as scalability, flexibility and ease-of-use/friendliness.

7 References

1. Decision no 922/2009/EC of the European parliament and of the Council of 16 September 2009 on Interoperability Solutions for European Public Administrations (ISA). Official Journal of the European Union, L 260/20 (3.10.2009). `http://ec.europa.eu/isa/documents/isa_lexuriserv_en.pdf`
2. Auer, S., Lehmann, J.: Making the web a data washing machine - creating knowledge out of interlinked data. Semantic Web Journal, vol. 1, no. 12, pp. 97-104. IOS Press (2010)
3. Cyganiak, R., Wood, D., Lanthaler, M.: RDF 1.1 Concepts and Abstract Syntax (February 25, 2014). `http://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/Overview.html`
4. Stadler, C., Martin, M., Auer, S.: Exploring the Web of Spatial Data with Facete. Companion proceedings of 23rd International World Wide Web Conference WWW, page 175--178 (2014)
5. P. E Salas, F. Maia Da Mota, K. Breitman, M. A Casanova, M. Martin, S. Auer. Publishing Statistical Data on the Web. International Journal of Semantic Computing 06(04):373-388 (2012)
6. Alonso, J. M.: Announcing the Global Open Data Initiative (GODI). World Wide Web Foundation, (June 11, 2013).

http://www.webfoundation.org/2013/06/announcing-the-global-open-data-initiative-godi/
7. Open Data Barometer - 2013 Global Report, http://www.opendataresearch.org/dl/odb2013/Open-Data-Barometer-2013-Global-Report.pdf
8. Orientation paper: research and innovation at EU level under Horizon 2020 in support of ICT-driven public sector. EC Digital Agenda news (22/05/2013), http://ec.europa.eu/information_society/newsroom/cf/dae/document.cfm?doc_id=2588
9. Indicators and regional development policies. The Italian position and current practice. Ministry of Economic Development - ITALY (2008) http://www.dps.mef.gov.it/documentazione/docs/all/postion_paper_indicators%2029%2002%2008.pdf
10. Schönthale, K., von Andrian-Werburg, S.: Identification and Selection of Indicators, DIAMONT Work Package Report. http://www.uibk.ac.at/diamont/downloads/workpackages/WP7_finalreport_070514.pdf
11. Janev, V. et al.: Supporting the Linked Data publication process with the LOD2 Statistical Workbench. Semantic Web Journal (under review), http://www.semantic-web-journal.net/content/supporting-linked-data-publication-process-lod2-statistical-workbench
12. Cyganiak R., Reynolds D.: The RDF Data Cube vocabulary (January 16, 2014). http://www.w3.org/TR/vocab-data-cube/
13. SDMX Information model: UML Conceptual Design (version 2.0, November 2005). http://sdmx.org/docs/2_0/SDMX_2_0%20SECTION_02_InformationModel.pdf
14. Janev, V., Mijović, V., Vraneš, S.: LOD2 Tool for Validating RDF Data Cube Models. In V. Trajkovik, A.Mishev (eds) Web Proceedings of the 5^{th} ICT Innovations Conference, Ohrid, Macedonia, September 12-15, 2013. http://ict-act.org/proceedings/2013/htmls/papers/icti2013_submission_01.pdf

Acknowledgements. The research presented in this paper is partly financed by the European Union (FP7 LOD2 project, Pr. No: 257943; FP7 GeoKnow, Pr. No: 318159), and partly by the Ministry of Science and Technological Development of the Republic of Serbia (SOFIA project, Pr. No: TR-32010).