This is a pre-print version of the following paper:
E. Tambouris, E. Kalampokis and K. Tarabanis (2015) Processing Linked Data Cubes, E. Tambouris et al. (Eds.): EGOV 2015, LNCS 9248. IFIP

Processing Linked Open Data Cubes

Efthimios Tambouris[1,2], Evangelos Kalampokis[1,2], Konstantinos Tarabanis[1,2]

[1] University of Macedonia, Thessaloniki, Greece
[2] Information Technologies Institute, Centre for Research & Technology – Hellas, Greece
{tambouris, ekal, kat}@uom.gr

Abstract. A significant part of Open Data provided by governments and international organizations concerns statistics such as demographics and economic indicators. The real value, however, of open statistical data will unveil from performing analytics on top of combined datasets from disparate sources. Linked data provide the most promising technological paradigm to enable such analytics across the Web. Currently, however, relevant processes and tools do not fully exploit the distinctive characteristics of statistical data. The aim of this paper is to present a process that enables publishing statistical raw data as linked data, combining statistics from multiple sources, and exploiting them in data analytics and visualizations. Moreover, the capability of existing software tools to support the vision of linked statistical data analytics is evaluated. We anticipate that the proposed process will contribute to the development of a roadmap for future research and development in the area.

Keywords: Linked data; data cubes; open data; statistics; data analytics.

1 Introduction

Open Data refer to the idea that certain data should be freely available for re-use for purposes foreseen or not foreseen by the original creator [1]. This data can be an important primary material for added value services and products, which can increase government transparency, contribute to economic growth and provide social value to citizens. As a result, governments and organisations launch portals that operate as single points of access for data they produce or collect [2]. A major part of this Open Data concerns statistics such as demographics and economic indicators [3]. For example, the vast majority of the datasets published on the open data portal[1] of the European Commission are of statistical nature.

Statistical data is often organised in a multidimensional manner where a measured fact is described based on a number of dimensions, e.g. unemployment rate could be described based on geographic area, time and gender. In this case, statistical data is compared to a cube, where each cell contains a measure or a set of measures, and thus we onwards refer to statistical multidimensional data as data cubes or just cubes [4].

[1] http://open-data.europa.eu

Linked Data has been introduced as a promising paradigm for opening up data because it facilitates data integration on the Web [5]. In the case of statistical data, Linked Data has the potential to create value to society, enterprises, and public administration through combining statistics from various sources and performing analytics on top of integrated statistical data [6][7].

During the last years various processes have been introduced to enable an understanding of how (a) governments open up their data for others to reuse (e.g. [8]) and (b) data providers publish their data according to the Linked Data principles to facilitate data integration on the Web (e.g. [9]). However, recent developments suggest that statistical data present distinctive characteristics and thus the vision of linked data cube analytics requires the introduction of new processes and software tools [10][11].

The aim of this paper is to present a process that enables publishing statistical raw data as linked data cubes, combining cubes from multiple sources, and exploiting them in data analytics and visualisations. Towards this end, we interviewed employees from public and private organizations that (a) produce and open up statistical data, (b) publish linked data cubes, or (c) consume statistical data to make decisions. The proposed linked data cubes process is employed to evaluate the capability of existing software tools to support the vision of linked data cube analytics. The evaluation results provide interesting insights into the software tools that are required so that to make possible the vision of linked data cube analytics.

The rest of the paper is structured as follows: Section 2 sets the background of our work and explains the need for a linked data cubes process. Section 3 presents related work regarding processes for linked and open data. In section 4 the approach we followed is described while in section 5 the proposed process is presented. In section 6 we apply the process to evaluate the capacity of existing tools to support the linked data analytics vision. Finally, section 7 draws conclusions.

2 Motivation

A data cube is specified by a set of dimensions and a set of measures. The dimensions create a structure that comprises a number of cells, while each cell includes a numeric value for each measure of the cube. Let us consider as an example a cube from Eurostat with three dimensions, namely time in years, geography in countries, and sex, that measures the percentage of population that is involved in lifelong learning. An example of a cell in this cube would define the percentage of males involved in lifelong learning in Denmark in 2014.

A fundamental step towards the exploitation of data cubes in Linked Data is the RDF data cube (QB) vocabulary, which enables modelling data cubes as RDF [12]. Centric class in the vocabulary is *qb:DataSet* that defines a cube. A cube has a *qb:DataStructureDefinition* that defines the structure of the cube and multiple *qb:Observation* that describe each cell of the cube. The structure is specified by the abstract *qb:ComponentProperty* class, which has three sub-classes, namely *qb:DimensionProperty*, *qb:MeasureProperty*, and *qb:AttributeProperty*. The first one

defines the dimensions of the cube, the second the measured variables, while the third structural metadata such as the unit of measurement.

At the moment, a number of statistical datasets are freely available on the Web as linked data cubes. For example, the European Commission's Digital Agenda[2] provides its Scoreboard as linked data cubes. An unofficial linked data transformation[3] of Eurostat's data, created in the course of a research project, includes more than 5,000 linked data cubes. Few statistical datasets from the European Central Bank, World Bank, UNESCO and other international organisations have been also transformed to linked data in a third party activity [15]. Census data of 2011 from Ireland and Greece and historical censuses from the Netherlands have been also published as linked data cubes [13][14]. Finally, the Department for Communities and Local Government (DCLG) in the UK also provides local statistics as linked data[4].

The real value, however, of linked data cubes is revealed in the case of combining statistics from disparate sources and performing analytics on top of them in an easy way. Following the previous example, let us consider a cube from Digital Agenda measuring internet usage that is structured based on the same three dimensions, i.e. time in years, countries, and sex. If we combine these two cubes from Eurostat and Digital Agenda, we can perform a regression analysis and derive some interesting results like the plot of Figure 1. In this case, the value is present when all needed steps can be easily performed using relevant online tools.

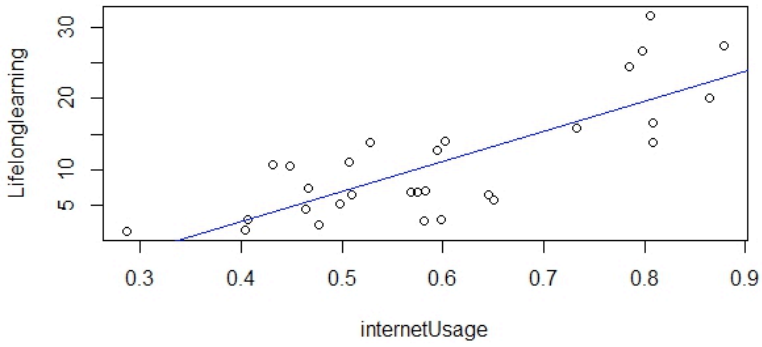

Fig. 1 Combining and analysing data cubes from Eurostat and Digital Agenda

During the last years, a few research endeavors focused on performing statistical analyses on top of combined linked data cubes [16]-[18]. These endeavors mainly proposed ad-hoc solutions that use specific datasets in order to prove the applicability of the approach. We believe, however, that a common understanding of the whole process of creating and exploiting linked data cubes is required in order to apply the concept of linked data analytics at a Web scale.

[2] http://digital-agenda-data.eu/data
[3] http://eurostat.linked-statistics.org
[4] http://opendatacommunities.org/data

3 Related Work

According to the literature, open data processes specify the steps that governments should follow to set their data free for others to reuse [8] [19]-[21]. For example, the process introduced by Janssen & Zuiderwijk [8] involves five steps, namely creating data, opening data, finding open data, using open data, and discussing and providing feedback on open data. Moreover, a few processes have been recently proposed in the literature to describe the steps that are followed in publishing and consuming Linked Data [9] [22]-[25]. For example, Auer et al. [9] present a process comprising eight steps: (i) transform data to RDF, which includes the extraction of data from sources (structured or unstructured) and its mapping to an RDF data model, (ii) store and index data efficiently and using appropriate mechanisms, (iii) manual revise, extend and create new structured information according to the initial data, (iv) establish links to different sources that regard the same entities but are published by different data publishers (v) enrich data with high-level structures so as to be more efficiently aggregated and queried (vi) assess data quality using data quality metrics available for structured information such as accuracy of facts and completeness, (vii) repair data so as to encounter data quality problems identified in the previous step, and (viii) search, browse and explore the data in a fast and user friendly manner.

However, these processes are general and need to be specialised for accommodating statistical data modelled using linked data technologies. In particular, these generic processes present the following limitations when applied to linked data cubes:

- They focus on the publishing part of Linked Data and they do not provide details on the exploitation, which is usually summarised at the last step of the process. In our case, however, the possible statistical analyses are well defined in the literature (e.g. OLAP analysis, statistical learning etc.) and thus should be further elaborated particularly as they can also provide feedback to the publishing steps of the process.
- Typically, data integration in the Web of Linked Data is facilitated by establishing *owl:sameAs* links [26], which indicate that two URI references refer to the same thing. However, in the case of cubes these links are applicable only at the metadata level that define the structure of the cube and not at the observation level. As a result, integration of data cubes is not currently properly accommodated in existing Linked Data processes.
- The use of the QB vocabulary introduces considerable complexity that calls for specific requirements in the publishing steps.

4 Approach

In order to understand the requirements of a linked data cube process we interviewed employees from public and private organisations that work with open data, linked data, and statistical analysis. More specifically, the following appointments were made per area:

- Open data: The head of the Open Data team of the Flemish government.
- Linked Data: Two employees from an international Swiss Bank.
- Statistical Data: 16 employees of the Irish Central Statistics Office (CSO). The interviewees were chosen as a cross-section of CSO staff from different functional areas and different levels of seniority, with particular focus on staff involved in data dissemination and IT operations. One of CSO's major statistical datasets is Census 2011, which has been already published as Linked Data[5].
- Open/Linked Data: Three employees from the Research Centre of the Government of Flanders; a government having as mission statement to conduct research in the fields of demographics, macroeconomics and social-cultural developments.
- Open/Linked Data: Three employees from the Assistant Deputy Director of Strategic Statistics in the UK Department for Communities and Local Government (DCLG). DCLG currently produces 53 main statistical datasests and is commited to routiely release its data as linked open data. It also maintains a data portal that currently contains more than 150 datasets[6].

The interviews along with the relevant literature resulted in the series of consecutive steps that structure the linked data cube process presented in section 5.

5 A Process for Linked Data Cubes

This section presents the proposed process for creating value through linked data cubes. This process comprises three phases, namely (a) Creating Cubes, (b) Expanding Cubes, and (c) Exploiting Cubes. The first phase involves creating linked data cubes from raw data, the second supports the expansion of a cube by linking it with other cubes on the Web, and the last one enables the exploitation of the cubes in data analytics and visualisations. The three phases further split up into a number of steps. A depiction of this process is presented in Figure 2. In the rest of the section the steps of each phase are outlined.

5.1 Step1.1: Discover & pre-process raw data

This step enables stakeholders to discover, access, view and process raw data cubes. At this step, data cubes come in various data formats such as CSV files, XLS files, RDBMS or RDF files. In addition, cubes can be formatted in various structures such as rectangular data, tree data and graph data.

In this step, stakeholders are able to browse raw data and perform activities aiming to improve the quality of raw data (e.g. data sorting, filtering, cleansing, transformation). This step could also include raw file or raw data storage in a local

[5] http://data.cso.ie
[6] http://opendatacommunities.org/

repository or database system. In this case, metadata regarding the provenance of raw data may be also stored along with the actual data.

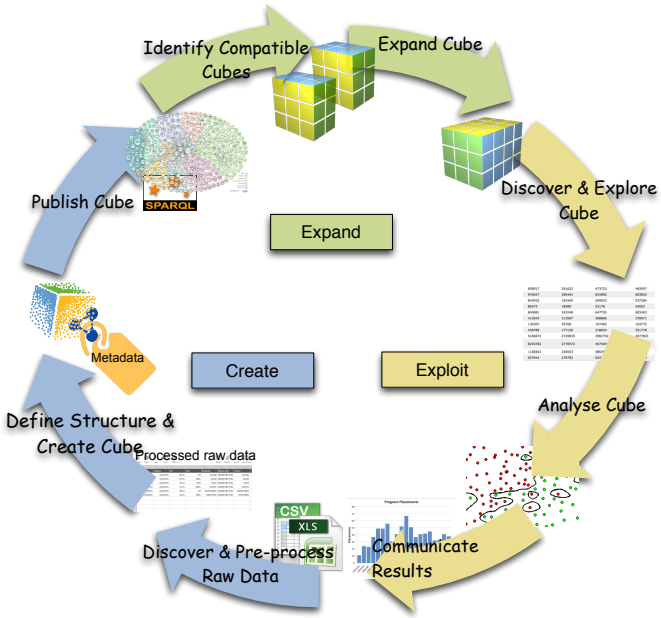

Fig. 2. The Linked Data Cubes process

5.2 Step 1.2: Define structure & create cube

An important step in Linked Data creation regards the definition of the structure of a model that the data will be mapped to. Initially, a conceptual model that drives the development of the structure of the Linked Data Cube is created. This specifies:

- The dimensions of the cube, which define what the observation applies to.
- The measured variables (i.e. what has been measured) along with details on the unit of measure or how the observations are expressed.

As reusing widely accepted vocabularies is considered to be of high importance in linked data, defining the structure of the model also requires importing and reusing existing linked data vocabularies. In the case of data cubes, the *RDF Data Cube (QB) vocabulary* constitutes the main framework to model data cubes as RDF graphs. In addition, other linked data vocabularies can be also used to define the values of the dimensions, measures and attributes of the cube. Common statistical concepts can be reused across datasets e.g. dimensions regarding age, location, time, sex etc. or the values of specific dimension (e.g. the countries of Europe). These concepts are defined in linked data vocabularies that standardise dimensions, attributes and code lists. The most widely accepted is the SDMX-RDF vocabulary[7], which is based on the statistical encoding standard SDMX.

[7] https://code.google.com/p/publishing-statistical-data/

As a result, publishing linked data cubes mainly requires discovery and reuse of controlled vocabularies. We should also note that reusing controlled vocabularies could be considered as reconciling against such collections. This peculiarity of data cubes introduces an extra need that is related to the management of controlled vocabularies that could be reused across different datasets. This includes the creation, store, search, discovery and reuse of existing controlled vocabularies.

This step also includes the creation of the actual RDF data out of the raw data based on the structure definition that was created at the previous step. This step includes the following activities: (a) URI design, (b) Definition of mapping between raw and RDF data, (c) Data storage to an RDF store, and (d) Validation for compliance with schema or values constraints.

Finally, this step also includes the enrichment of RDF data cubes with metadata to facilitate discovery and reuse. Sources of metadata include raw data files, the cube's structure and/or standard thesaurus of statistical concepts.

5.3 Step 1.3: Publish cube

In this step, the generated data cubes are made available to the public through different interfaces e.g. Linked Data API, SPARQL endpoint, downloadable dump etc. In addition, during this step the datasets are publicised in data catalogues such as Europe's public data portal[8] or other national portals (e.g. data.gov.uk or data.gov.gr), the datahub platform[9] or the Linking Open Data cloud[10].

Metadata that describe the dataset should be also published along with the actual data. The produced metadata are usually shared across multiple platforms and implementations. As a result, stakeholders need to be able to import or export metadata related to data cubes.

5.4 Step 2.1: Identify compatible cubes

This step supports the identification of compatible to join cubes in order to enable expanding linked data cubes. The identification of compatible cubes is performed through two processes:

- Search on an existing collection of linked data cubes and evaluate the compatibility of a cube at hand with every cube in the collection. The compatibility evaluation is based on (a) the structure of the cubes i.e. dimensions, measures, levels and hierarchies, and (b) the desired type of join. For example, a cube is compatible to join in order to add a new measure to an original cube if: *i)* both cubes have the same dimensions, *ii)* the second cube has at least the same values in each dimension of the original cube, and *iii)* the second cube has at least one measure that does not exist at the original cube.

[8] http://publicdata.eu
[9] http://datahub.io
[10] http://lod-cloud.net

- Create a set of compatible cubes from an initial linked data cube by computing aggregations across a dimension or a hierarchy. In the case of aggregating data across a dimension, 2^n new cubes are created where n is the number of the dimensions of the cube. In the case of aggregating data across a hierarchy, a new cube is created that contains observations for all values of a dimension at every level. Special attention should be paid on the types of measures and dimensions and the aggregation function (i.e. sum, count, min, max etc.) that can be used.

This step can also include the establishment of typed links between compatible to join cubes. These links will enable, at a later stage, identifying linked data cubes that can be combined in order to perform enhanced analytics on top of multiple linked data cubes. For this reason, it is important to define compatibility of cubes and develop tools that could search on large collections of cubes and discover cubes that can potentially be combined.

5.5 Step 2.2: Expand cube

Expanding cubes enables adding more data into a cube. We assume that a cube can be expanded by increasing the size of one of the sets that defines it. Therefore, a cube can be expanded by adding one or more elements into the set of measures, the set of concepts in a dimension and the set of dimensions. This can be done by merging a cube with a second one, which is compatible with the initial cube. The links that have been established at the previous step can be exploited towards this end.

Following the same example, we see that we have two cubes that describe two different measures (i.e. unemployed people and crime incidents) based on the same dimensions (i.e. time and geography), and with the same concepts (i.e. 2010 for time and the European countries for geography). These two cubes are compatible to merge and thus a new cube with two measures can be created out of the initial ones.

We should note that the expanded cube could be either created and stored or just conceptually defined in order to be used along with a data analytics tool.

5.6 Step 3.1: Discover & explore cubes

At this step, stakeholders aiming to consume data exploit the mechanisms set up at the previous step in order to discover the appropriate cubes for a task at hand. For example, we consider a researcher that needs to study the relation between unemployment and criminality and thus needs to analyse data that describe unemployment and criminality in different geographic areas or time periods.

In general, the discovery of linked data cubes could be done through:

- A data catalogue that allows exploring the available data cubes based on (a) generic metadata records stored inside the catalogue platform that describe the cube as a whole, and (b) Cube-specific metadata that provide information about the concepts that formulate the cube i.e. dimensions and measures.
- Full-text search that enables discovery of data cubes not only by metadata but also by the actual content of the cubes.

In our example, we suppose that the researcher identifies two cubes:

- A cube presenting the number of unemployed people in three dimensions i.e. countries, years and age groups.
- A cube presenting crime incidents in two dimensions i.e. countries and time (quarters of the year).

At this stage, we consider that the researcher is also able to browse the cube in order to better understand the data and proceed with further analysis. This enables the researcher to view data based on different dimensions or measures. For example, if the data describes the unemployment rate at different European countries in different years then stakeholders could view either the unemployment rates of a particular country throughout the years or the unemployment rates of a specific year across different countries. This would enable stakeholders also to sort or filter the data based on the values of the dimensions or the actual values of the observations.

5.7 Step 3.2: Analyse cube

In this step the data cubes that were resulted from the previous step are employed in order to perform analytics through (a) OLAP operations, (b) computing simple summaries of the data, and (c) creating statistical learning models.

The transformation of linked data cubes at the previous step will enable stakeholders to perform the following OLAP operations:

- *Dimension reduction:* This would enable users to select part of a data cube by removing one of the dimensions. In the unemployment rate example this would enable, for example, removing the age group dimension and thus keeping only the time and location dimensions.
- *Roll-up* and *drill-down* operation: These OLAP operations allow stakeholders to navigate among levels of data cube by stepping down or up a concept hierarchy. Following the previous example, stepping down a concept hierarchy for the dimension time could perform this OLAP operation. If we consider the concept hierarchy "month<quarter<year" then drill down would present unemployment rate of different age groups at different countries for every quarter.

The stakeholder could also select to produce either quantitative (i.e. summary statistics) or visual (i.e. simple to understand graphs) summaries. As regards the quantitative summaries, a stakeholder in this step will be able to describe the observations across a dimension using descriptive statistics. For example, this step would enable the calculation of the mean and standard deviation of the unemployment rate of European countries in a particular year. Moreover, stakeholders would be able to calculate statistics (e.g. Pearson's correlation coefficient) that estimate dependences between paired measures described in disparate but compatible cubes. Paired here is used to denote that the measures share at least one common dimension and thus can be compared.

Finally, the types of visualisation charts that can be used in this step include scatter plots, bar charts, pie charts, histograms, geo charts, timelines etc.

Following the example of the previous steps, the researcher use the cubes created after the last step in order to perform the following:

- Create a scatter-plot presenting unemployed people against crime incidents across European countries.

- Calculate Pearson's correlation coefficient between number of unemployed and number of crimes.

In this step, the cubes that were created in the previous steps could be also used in machine learning and predictive analytics in order to produce learning or predictive models. At the same step, the models that were created could also be published into the Linked Data Web and thus feedback the lifecycle at the first step.

Following the example of unemployment and criminality, we consider that the researcher now wants to create a model in order to be able to estimate future crime rates based on unemployment rates. Towards this end, the researcher exploits the results of the previous step and the data cubes in order to select an appropriate data mining method (e.g. Support Vector Machines) and build a model. The researcher goes back to the previous steps in order to also identify data to evaluate the model.

5.8 Step 3.3: Communicate results

This step involves the visualisation of results. This step may feed back to the first step of the process if the results of the analyses performed in the previous steps indicate a need for further analyses requiring additional data. Towards this end, the analysis proceeds with the first step of the process in order to discover new raw data, transform them to RDF and eventually perform a comparative analysis with existing RDF data cubes.

6 Tools

The exploitation of Linked Data in statistics requires specialised software tools that (a) are generic and thus applicable to all datasets that use the QB vocabulary, and (b) support each step of the Linked Data Cube process. Therefore, existing linked data tools should be evaluated to determine their capability to fully support the process steps.

In this section, we evaluate nine widely-used open data, linked data, and statistical analysis tools, namely:

1. OpenRefine[11]
2. PoolParty[12]
3. CSVImport (LOD2 project)[13]
4. TabLinker[14]
5. SILK[15]
6. Pubby[16]
7. CubeViz (LOD2 project)[17]

[11] http://openrefine.org
[12] http://www.poolparty.biz
[13] https://github.com/AKSW/csvimport.ontowiki
[14] https://github.com/Data2Semantics/TabLinker
[15] http://wifo5-03.informatik.uni-mannheim.de/bizer/silk/
[16] http://wifo5-03.informatik.uni-mannheim.de/pubby/

8. SPARQL R[18]
9. RapidMiner (LOD)[19]

Table 1 Evaluating the capacity of 9 tools to support the 8 steps of the process

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Step 1	*F*	*N*	*P*	*P*	*N*	*N*	*N*	*N*	*N*
Step 2	*P*	*P*	*P*	*P*	*P*	*N*	*N*	*N*	*N*
Step 3	*N*	*N*	*P*	*P*	*N*	*P*	*N*	*N*	*N*
Step 4	*N*	*N*	*N*	*N*	*P*	*N*	*N*	*N*	*N*
Step 5	*N*	*N*	*N*	*N*	*N*	*N*	*N*	*N*	*N*
Step 6	*N*	*N*	*N*	*N*	*N*	*N*	*N*	*N*	*N*
Step 7	*N*	*N*	*N*	*N*	*N*	*N*	*P*	*P*	*P*
Step 8	*N*	*N*	*N*	*N*	*N*	*N*	*P*	*P*	*P*

In Table 1 the results of our analysis are presented. The horizontal axis presents the tools while the vertical the process steps. In each cell a letter indicates whether the tool (F)ully, (P)artially or N(ot) covers the functionality required by a step of the process. The analysis that we performed revealed that the following important functionalities are not currently supported by existing tools:

- Transform raw data to linked data cubes (as existing tools for RDF creation are difficult to use in the case of the QB vocabulary).
- Materialise cubes by computing aggregations across dimensions and hierarchies. This functionality is important for enabling OLAP browsing.
- Identify cubes with similar structure that could potential integrate.
- Create integrated views of multiple linked cubes on the Web. This will enable performing analytics on top of multiple cubes at a Web scale.
- Browse a linked data cube and perform advanced OLAP operations such as drill-down and roll-up.

Some tools, such as R and RapidMiner, with their extensions for importing RDF data can be also used for enabling performing data analytics on top of linked data cubes. We should, however, note that these generic RDF importers are difficult to use in the case of cubes because of the complexity that the QB vocabulary introduces. Our analysis revealed that linked data cube specific extensions are needed.

7 Conclusions

During the last years the Open Data movement has been introduced evangelising the need for certain data to be freely available for re-use. A major part of Open Data concerns statistics that is structured as multi-dimensional data cubes. Linked Data

[17] http://cubeviz.aksw.org
[18] http://cran.r-project.org/web/packages/SPARQL/
[19] http://dws.informatik.uni-mannheim.de/en/research/rapidminer-lod-extension/

technologies have the potential to realise the vision of combining and performing analytics on top of previously isolated cubes at a Web scale. However, a common understanding of the whole process of creating and exploiting linked data cubes is required in order to be able to apply the concept of linked data cube analytics at Web scale. Existing processes for linked and open data are general and need to be specialised for accommodating statistical data modelled using linked data technologies

In this paper, we introduced a process that enables publishing statistical raw data as linked data cubes, combining cubes from multiple sources, and exploiting them in data analytics and visualisations. The process comprises the following eight steps: (i) discover & pre-process raw data, (ii) define structure & create cube, (iii) publish cube, (iv) identify compatible cubes, (v) expand cube, (vi) discover & explore cube, (vii) analyse cube, and (viii) communicate results.

The proposed process was applied to evaluate the capability of existing tools to support the vision of linked data cube analytics. The results revealed that tools need to be specified and developed to support the easy creation of linked data cubes, the identification of linked data cubes with similar structure, the integration of linked data cubes, and the easy statistical analysis (e.g. OLAP analysis) of integrated cubes. We anticipate that the proposed process will contribute to a better understanding of linked data cube analytics and to the development of a roadmap for future research and development in the area.

Acknowledgments

This work is funded by the European Commission within the 7th Framework Programme in the context of the project OpenCube (http://opencube-project.eu) under grand agreement No. 611667. The authors would like to thank the whole OpenCube consortium that interviewed employees from private and public organisations as well as Mr. Konstantinos Tsiftsis who analysed linked data from Eurostat and Digital Agenda and created Figure 1.

References

1. European Commission. Open data: An engine for innovation, growth and transparent governance. Communication from the Commission, COM(2011) 882 final, December 2011
2. Kalampokis, E., Tambouris, E., Tarabanis, K.: A Classification Scheme for Open Government Data: Towards Linking Decentralized Data. International Journal of Web Engineering and Technology, 6(3), 266-285 (2011)
3. Cyganiak, R., Hausenblas, M., McCuirc, E.: Official Statistics and the Practice of Data Fidelity. In: Wood, D. (ed.) Linking Government Data, pp. 135–151. Springer (2011)
4. Datta, A., Thomas, H.: The cube data model: a conceptual model and algebra for on-line analytical processing in data warehouses. Decision Support Systems 27(3), 289 – 301 (1999)

5. Bizer, C., Heath, T., Berners-Lee, T.: Linked data - the story so far. International Journal on Semantic Web and Information Systems 5(3), 1–22 (2009)
6. Kalampokis, E., Tambouris, E., Tarabanis, K.: Linked Open Government Data Analytics. Wimmer, M.A., Janssen, M., Scholl, H.J. (eds.) EGOV2013, LNCS, vol. 8074, pp.99-110. IFIP International Federation for Information Processing (2013)
7. Abello, A., Darmont, J., Etcheverry, L., Golfarelli, M., Mazon, J.-N., Naumann, F., Pedersen, T., Rizzi, S. B., Trujillo, J., Vassiliadis, P., Vossen, G.: Fusion cubes: Towards self-service business intelligence. Data Warehousing and Mining 9(2), 66-88 (2013)
8. Janssen, M., Zuiderwijk, A.: Open Data And Transformational Government. Transforming Government Workshop. Brunel University, United Kingdom (2012)
9. Auer, S., Lehmann, J., Ngomo, A. C. N., Zaveri, A.: Introduction to linked data and its lifecycle on the web. In: Reasoning Web. Semantic Technologies for Intelligent Data Access, pp. 1-90, Springer Berlin Heidelberg (2013)
10. Kalampokis, E., Nikolov, A., Haase, P., Cyganiak, R., Stasiewicz, A., Karamanou, A., Zotou, M., Zeginis, D., Tambouris, E., Tarabanis K.: Exploiting Linked Data Cubes with OpenCube Toolkit, Proc. of the ISWC 2014 Posters and Demos Track a track within 13th International Semantic Web Conference (ISWC2014), 19-23 October 2014, Riva del Garda, Italy, CEUR-WS Vol.1272 (2014).
11. Ermilov, I., Martin, M., Lehmann, J., Auer, S.: Linked Open Data Statistics: Collection and Exploitation. In: Klinov, P., Mouromtsev, D. (eds.) Knowledge Engineering and the Semantic Web, vol. 394, pp. 242-249. Springer Berlin Heidelberg (2013)
12. Cyganiak, R., Reynolds, D.: The RDF Data Cube Vocabulary: W3C Recommendation. W3C, Tech. Rep., (2014)
13. Petrou, I., Papastefanatos, G. Dalamagas, T.: Publishing census as linked open data: A case study. In: Proceedings of the 2Nd International Workshop on Open Data, ser. WOD '13. New York, NY, USA: ACM, 2013, pp. 4:1–4:3.
14. Meroño-Peñuela, A., Ashkpour, A., Rietveld, L., Hoekstra, R., & Schlobach, S. (2012). Linked Humanities Data: The Next Frontier?. In: A Case-study in Historical Census Data. Proceedings of the 2nd International Workshop on Linked Science 2012, vol. 951, 2012.
15. Capadisli, S., Auer, S. Riedl, R.: Linked Statistical Data Analysis, ISWC SemStats (2013), http://csarven.ca/linked-statistical-data-analysis
16. Zapilko, B., Mathiak, B.: Performing statistical methods on linked data. In: International conference on dublin core and metadata applications, pp. 116-125 (2011)
17. Capadisli, S., Meroño-Peñuela, A., Auer, S., Riedl, R.: Semantic similarity and correlation of linked statistical data analysis. In: Proceedings of the 2nd International Workshop on Semantic Statistics (SemStats 2014), ISWC. CEUR (2014)
18. Kämpgen, B., Stadtmüller, S., Harth, A.: Querying the global cube: Integration of multidimensional datasets from the web. In: Knowledge Engineering and Knowledge Management (pp. 250-265). Springer International Publishing (2014)
19. van Veenstra, A. F., van den Broek, T.: A Community-driven Open Data Lifecycle Model Based on Literature and Practice. In *Case Studies in e-Government 2.0* (pp. 183-198). Springer International Publishing (2015)
20. Curtin, G. G.: Free the data!: E-governance for megaregions. Public Works Management & Policy, 14(3), 307–326 (2010)
21. Hyland, B., Wood, D.: The joy of data - a cookbook for publishing linked government data on the web. In Linking government data (pp. 3-26). Springer New York (2011)
22. Edgard, M., Shekarpour, S., Auer, S. and Ngomo, A-CN.: Large-scale RDF dataset slicing. In Semantic Computing (ICSC), 2013 IEEE Seventh International Conference on, pp. 228-235. IEEE (2013)
23. Villazón-Terrazas, B., Luis, M. V-B., Oscar, C., Asunción, G-P.: Methodological guidelines for publishing government linked data. In *Linking government data*, pp. 27-49. Springer New York (2011)

24. Ding, L., Peristeras, V., Hausenblas, M.: Linked Open Government Data [Guest editors' introduction]. Intelligent Systems, IEEE, 27(3), 11-15 (2012)
25. Alani, H., Dupplaw, D., Sheridan, J., O'Hara, K., Darlington, J., Shadbolt, N., Tullo, C.: Unlocking the potential of public sector information with semantic web technology. The 6th International Semantic Web Conference, 11–15 Nov 2007, Busan, Korea (2007)
26. Schmachtenberg, M., Bizer, C., Paulheim, H.: Adoption of the linked data best practices in different topical domains. In: *The Semantic Web – ISWC 2014*, ser. Lecture Notes in Computer Science, P. Mika, T. Tudorache, A. Bernstein, C. Welty, C. Knoblock, D. Vrandec˘ic´, P. Groth, N. Noy, K. Janowicz, and C. Goble, Eds. Springer International Publishing, vol. 8796, pp. 245–260 (2014)