Linked Data Cubes: Research results so far

Areti Karamanou[1,2], Evangelos Kalampokis [1,2], Efthimios Tambouris[1,2], and Konstantinos Tarabanis[1,2]

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

Abstract. During the last years a growing body of literature studied Linked Data Cubes. The objective of this paper is to accumulate this body of knowledge and provide a preliminary analysis of the research results in the area so far. Towards this end, we systematically reviewed the scientific literature to identify relevant studies. These studies were analysed and synthesised in the form of a proposed conceptual framework, which was thereafter applied to further analyse this literature, hence gaining new insights into the field. The framework comprises three dimensions, namely category of contribution, step of data analysis, and application area. The application of the framework resulted in interesting findings. For example, the majority of the contributions that focus on publishing linked data cubes are cases while the majority of the exploitation contributions are software tools. Moreover, integration of data cubes remains largely unexamined in the literature. This paper, however, does not present the final results of the analysis of the literature as this is still an ongoing activity.

Keywords: Data cubes; linked data; statistics; data analytics; review.

1 Introduction

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 data 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 [2].

Linked data has been introduced as a promising technological paradigm that facilitates data integration on the Web [1]. In the case of cubes, linked data has the potential to transform the Web to a platform for performing statistical analysis on top of combined data enabling this way the realisation of innovative data analytics scenarios [3]. As a result, during the last years, a growing number of research contributions in the area of "linked data cubes" have been published in journals and proceedings of conferences and workshops.

The aim of this paper is to consolidate this body of knowledge and provide a preliminary understanding of the research results in the area so far. Towards this end, we systematically reviewed the scientific literature to identify relevant studies. These studies were analysed and synthesised in the form of a proposed conceptual framework, which was thereafter applied to further analyse this literature, hence gaining new insights into the field.

The rest of the paper is structured as follows: Section 2 describes the approach that we followed to achieve the objectives of the paper. Section 3 presents the conceptual framework that structures the area of linked data cubes. Section 4 presents the results of the analysis of the literature based on the conceptual framework. Finally, section 5 draws conclusions.

2 Approach

A systematic literature review [4] was conducted to achieve the objectives of this paper. At first, a systematic search was used to acquire a large number of relevant scientific papers. The authors initially systematically searched Google scholar using “linked data” AND “data cube” as key words. The search resulted in the collection of an initial pool of articles. These articles were used as a starting point to a) go backward by reviewing their citations and b) go forward by using the functionality provided by Google scholar that allows to find articles citing the previously identified articles. The articles identified from the previous process were studied and filtered resulting in a final set of 136 articles. These articles have been published in scientific journals, or presented in scientific conferences and workshops and can be found in the final section of this paper (Literature Review References).

Moreover, a concept-centric analysis was used to synthesize the acquired knowledge. Specifically, the main characteristics of the area were extracted and a conceptual framework for linked data cubes was created in order to structure the area. Finally, the framework was used to classify and further analyse the literature and to extract insights into the area of linked data cubes.

3 Framework

The proposed framework (Fig. 1) comprises three dimensions:

- The type of contributions in the literature. This includes software, architecture, formal theory, vocabulary, and use case. Some of these categories are further divided into sub-categories. For example, the vocabulary type can be related either to a definition of a vocabulary or the application of a vocabulary.
- The linked data cube analysis steps. These include data cube publishing, exploiting, combining, quality assurance, and access control. Two of the steps are further divided. The exploiting step can refer to (a) browsing, (b) OLAP analysis, (c) visualisation, and (d) statistical analysis. Moreover, quality assurance can refer to either instance-level quality or schema-level quality.

– Application areas such as health, finance, and environment.

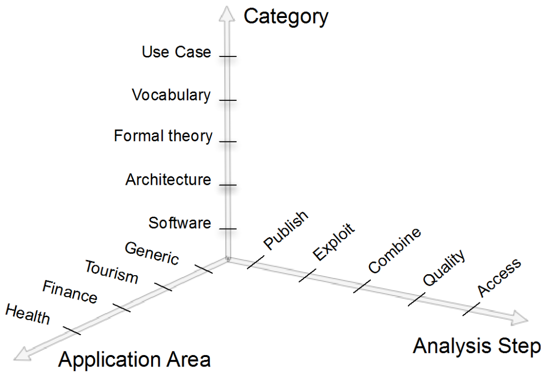

Fig. 1. The linked data cube framework

The dimensions define a three-dimensional space which structures the area and enables the categorisation and further analysis of the literature. For example, a contribution in the literature may refer to a vocabulary that enables access control in linked data cubes in health [65].

4 Results

We now employ the framework in order to gain insight into the research on linked data cubes. We initially categorise the articles based on the three dimensions.

Using only the "Analysis Step" dimension we can categorise the literature into the five steps. The vast majority of the contributions are related to publishing (42%) and exploitation (37%). The rest of the contributions are related to combining (13%), "Quality" (5%), and "Access" (3%). If we drill-down in the "Exploit" step, then we can see that 38% of the contributions refer to visualisation, 29% to OLAP analysis, 26% to statistical analysis, and 7% to browsing.

Using only the "Category" dimension of the framework we can categorise the literature based on the different types of contributions. The vast majority of the contributions describe use cases (38%) and present software tools (38%). Other types of contributions include vocabularies (15%), formal theory (7%), and architectures (2%).

Interestingly, taking into account only the "Application Area" dimension, we can see that 40% of the contributions are domain-specific. These contributions span a wide range of domains including health, policy-making, government, environment, economics, biology, and tourism. The most notable domain is health that characterises 25% of all domain-specific contributions.

By synthesizing the results, we can further analyse the literature and make some interesting observations regarding the research contributions so far. First, it is very important to combine the "Category" and "Analysis Step" dimensions. By doing so we can get two different views of the literature, (a) the types of contributions in relation to the analysis step and (b) the analysis steps in relation to the types of contributions. The vast majority (64%) of the use cases in the literature are about publishing and only 20% and 13% respectively describe cases of exploiting and combining. Software contributions mainly focus on requirements related to exploitation (56%). Out of these exploitation software contributions 55% enables creating visualisations, 24% enables performing statistical analysis, 13% enables browsing, and 8% enables performing OLAP operations on top of linked data cubes. The rest of the software contributions facilitate publishing (25%), support combining (19%), ensure cubes' quality (6%), and enable controlling access (3%) to linked data cubes. Interestingly, half of the contributions that introduce vocabularies support the publishing of linked data cubes (such as the RDF data cube vocabulary). The rest of the vocabulary related contributions focus on (a) facilitating the exploitation (36%) of linked data cubes, (b) supporting quality assurance (11%), and (c) enabling access control (3%). Finally, formal theory related contributions focus only on issues related to linked data cube integration and exploitation.

If we study the analysis steps according to the types of contribution, we can come up with some interesting results. First, the majority of the publishing contributions are use cases (58%), while software (22%) and vocabularies (18%) follow. The majority of the exploitation contributions are software (57%) with use cases (21%), vocabularies (15%), and formal theories (7%) coming next. Finally, regarding the contributions about combining linked data cubes, 38% are use cases, 29% are software tools, and 25% are theoretical contributions.

We can also combine the "Application Area" dimension with the "Category" dimension. If we do so, we will see that 77% of all software tools are domain independent. The same metric is 82% and 79% in the case of software tools for publishing and exploitation respectively. Interestingly, however, this percentage goes down to 57% in the specific case of software tools that support combining of linked data cubes. On the other hand, 78% of all vocabulary contributions are domain independent. More specifically, domain-specific vocabularies have been developed only for publishing and access control.

We now focus on two specific types of contributions of major importance, namely vocabularies and software. For these two specific categories we present in more detail the contributions so far.

4.1 Vocabularies

Several RDF vocabularies have been proposed to model statistical data as RDF. Two of the first vocabularies introduced were the Statistical Data and Metadata eXchange (SDMX) information model that was proposed to represent statistical data and make them available using web services [21] and the Statistical Core Vocabulary (SCOVO) for modelling and publishing statistical data [49].

Other vocabularies proposed in literature include Open Cubes [34], the vocabulary that models Data Warehouses' cubes [27], the SCOVOLink ontology that extends SCOVO in order to allow the creation of links between the data and the described entities [133] and the LOIUS ontology that also extends SCOVO to describe statistics about University student activities [106]. However the mostly used vocabulary is the RDF Data Cube vocabulary (QB) [23] that models statistical data as RDF. The first version of QB was released on April 2012 but the vocabulary has been a W3C standard since January 2014.

As QB vocabulary became a standard, a number of studies extended it in order to overcome some of the QB's limitations or satisfy specific domains' needs and requirements. For example, the Linked Clinical Data Cube (LCDC) is a vocabulary that combines QB and DDI-RDF in order to allow the publication of clinical data as linked data [75]. Moreover, Bandholtz et al.'s vocabulary is also a SCOVO-based vocabulary that models German Environmental Specimen Bank (ESB), data that describe the accumulation of pollutants/substances in test subjects at specific places over time [9].

Recently, however, the focus has been moved from the definition to the application of vocabularies. For example, Becker et al. [12] performed a quantitative survey on how the QB vocabulary is applied to model multidimensional data. Kalampokis et al. [62] also investigated the challenges related to the different practices that can be followed in applying the QB vocabulary.

Finally, a number of vocabularies facilitate the exploitation of linked data cubes such as the REA (Resources, events, agents) - based model for OLAP cubes [120] and Prat et al.'s model that represents OLAP cubes as an OWL-DL ontology [107, 108]. Another example is the QB4OLAP vocabulary, that extends QB in order to allow implementing OLAP operations on cubes such as rollup, slice, dice and drill-across using SPARQL queries [33, 34, 36]. Finally, Follenfant et al.'s model [38] is an extension to the QB vocabulary that enables the description of analytical processes (e.g. data analysis) that can be performed within reporting tools.

4.2 Software

A number of software solutions have been developed that aim to facilitate the publishing of linked data cubes. For example, the OpenCube toolkit [60, 61, 64] includes a number of open source software components that have been developed to enable data cubes publishing. Specifically, the OpenCube toolkit includes the TARQL extension for the conversion of legacy tabular data to RDF, the D2RQ data cubes extension for the conversion of relational databases to RDF, the JSON-stat2qb data cubes extension for the conversion of JSON-stat files into RDF and the R2RML data cubes extension for data transformation of cubes structured in tabular sources to linked data cubes. Another example is the LOD2 Statistical Workbench [58] that includes a set of tools for accessing, manipulating, exploring and publishing statistical data. Specifically, LOD2 includes tools for importing and editing cubes, and managing their dimensions and code lists.

Moreover, LODStats [32] is a web-based tool that collects and publishes statistics about the LOD cloud. At the same time, OLAP2DataCube [116, 117] and the CSV2DataCube [116] are both plug-ins for the Ontowiki tool for extracting and publishing statistical data in RDF. The two plug-ins enable the publishing of OLAP databases to RDF and CSV data to RDF respectively. Finally, TabLinker [95] is also a tool for publishing excel data as data cubes.

A number of software solutions also enable the exploitation of cubes through browsing. For example, LSD Dimensions [91] is a web-based tool for monitoring dimensions and codes (i.e. variables and values) of data cubes. It provides users with a list of the dimensions of cubes stores in Datahub.io and allows users to browse them. Linked Data Cubes Explorer (LDCX) [69] is another tool that enables the browsing of data cubes. It allows users to select a number of datasets and show and explore their dimensions and measures. The OpenCube Browser [64] is also a tool that provides functionalities for exploring linked open statistical data cubes.

In addition, a number of tools have been developed aiming at producing meaningful charts, maps and other visualisations out of statistical data cubes. Visualisations vary from charts to maps. For example, the OpenCube MapView [64], part of the OpenCube toolkit, enables visualisations of linked data cubes with a geo-spatial dimension on maps. The Interactive chart visualisation widget, also part of the OpenCube toolkit, allows the visualisation of RDF data cubes, in particular, time series data using charts. Another example is the LOD2 Statistical Workbench [57] that provides CubeViz [82, 83], a tool for visualizing a data cube's observations with suitable charts. Moreover, Map4rdf [25] is a browsing tool that allows the exploring and visualisation of RDF data cubes that include geospatial information using maps. Map4rdf supports Google Maps and OpenStreetMap. Map4rdf has also a mobile compatible counterpart, Map4RDiOS app [80] which can be installed in mobile devices with iOS and offers similar functionalities with Map4rdf. Another visualisation tool is the Linked Data Visualisation Model [50] that also includes a plug-in to facilitate the publication of non-data cubes as cubes. The CODE Visualisation Wizard [98, 99, 128] is also a platform that imports statistical data, transforms them to data cubes and suggests and creates visualizations (bar charts, lines, pies etc.) on top of them. Moreover, ETIHQ visual dashboard [115] is a software that allows the visualisation of tourism indicators modelled as data cubes to enhance the decision making of Destination Marketing Organizations. Finally, qb.js [87] is another tool that enables the creation of visualizations from data cubes without requiring knowledge of linked data or semantic tools.

Software related to performing OLAP operations on top of data cubes include the OpenCube OLAP Browser [64], a tool that enables using linked data cubes to perform OLAP operations (such as drill-down, roll-up and pivot) and Saad et al.'s [113] prototype that enables performing OLAP operations on top of data cubes.

Software solutions described in literature also allow performing various types of statistical analysis on top of data cubes. For example, the Linked Open Data

Extension of Rapidminer adds to Rapidminer the Data Cube Importer operator [109]. The operator enables the importing of data model using the QB vocabulary so as to perform a plethora of statistical analyses and predictive analytics functions.

Only a few tools enable the integration of data cubes. The OpenCube Compatibility Explorer of the OpenCube toolkit allows users to identify compatible cubes for potential merge and then establish typed links to facilitate discovery. The OpenCube Expander creates new expanded cubes by merging two compatible cubes. Another example is the LOD2 Statistical Workbench [58] that includes tools for interlinking the dimensions of two data cubes and for enriching data cubes with external data (e.g. data from dbpedia). Bacon [10] is also an open source software that enables the fusion of semantically associated cubes and the integration of related cubes into a single cube. The discovery of relative cubes is based on the structure as well as the content of the cubes. The efficiency of the integration is increased by allowing the modification of the cubes' structure and also by detecting duplicate information in the integrated cube.

Relevant studies propose data quality related tools such as the RDF Data Cube Validation tool, part of the LOD2 Statistical Workbench, that can be used to validate the integrity constraints defined in the QB vocabulary specification and to automatically repair identified errors [56] and Vital [24] that allows the detections of bugs in data cubes (e.g. insufficient documentation, wrong data types, syntax errors in URIs, inconsistencies between data structure specifications and observations). A relevant service for assessing the quality of data cubes is Computex [41] that allows the validation of statistical index data represented as data cubes.

Finally, LiMDAC [65] was introduced as a platform that enables controlling the access to medical data cubes.

5 Conclusions

Statistical data is often organised in a multidimensional manner where a measured fact is described based on a number of dimensions structuring this way a data cube. Linked data technologies have the potential to realize the vision of combining and performing analytics on top of previously isolated cubes across the Web.

In this paper we consolidated the literature in the area of linked data cubes and we analysed and synthesised this body of knowledge in the form of a proposed conceptual framework. This framework was then applied to further analyse the literature and, as a result, we come up with interesting results about this research area. For example, the majority of the contributions that focus on publishing of linked data cubes are use cases while the majority of the exploitation contributions are software tools. Moreover, we identified that the integration of data cubes is still a unexploited research topic.

We should however note that this paper does not present the final results of the analysis of the literature as this is still an ongoing activity.

Acknowledgments. Part of this work was funded by the European Commission within the H2020 Programme in the context of the OpenGovIntelligence project (`http://OpenGovIntelligence.eu`) under grant agreement no. 693849.

References

1. 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.
2. Datta, A. Thomas, H.: The cube data model: A conceptual model and algebra for on-line analytical processing in data warehouses. *Decis. Support Syst.*, 27(3):289–301, December 1999.
3. Kalampokis, E., Tambouris, E., Tarabanis, K.: Linked open cube analytics systems: Potential and challenges. *IEEE Intelligent Systems*, 31(5), 2016.
4. Webster, J. Watson, R. T.: Analyzing the past to prepare for the future: Writing a literature review. *MIS quarterly*, pages xiii–xxiii, 2002.

Literature Review References

5. Álvarez-Rodríguez, J. M., Labra-Gayo, J. E., Ordoñez de Pablos, P. *Leveraging Semantics to Represent and Compute Quantitative Indexes: The RDFIndex Approach*, pages 175–187. Springer International Publishing, Cham, 2013.
6. Aracri, R., De Francisci, S., Pagano, A., Scannapieco, M.: Official statistics meets the semantic web: How sdmx and rdf can live together. In *Proceedings of the NTTS Conference*, 2015.
7. Aracri, R., De Francisci, S., Pagano, A., Scannapieco, M., Tosco, L., Valentino, L.: Publishing the 15th italian population and housing census in linked open data. In *SemStats2014. Springer, Riva del Garda, Italy*, 2014.
8. Auer, S., Demter, J., Martin, M., Lehmann, J. *LODStats – An Extensible Framework for High-Performance Dataset Analytics*, pages 353–362. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.
9. Bandholtz, T., Schulte-Coerne, T., Rüther, M.: Linked environment data: Scovofying the environment specimen bank. In *ESWC*, 2009.
10. Bayerl, S. Granitzer, M.: Bacon: Linked data integration based on the rdf data cube vocabulary. In *Proceedings of the 5th International Conference on Web Intelligence, Mining and Semantics*, WIMS '15, pages 14:1–14:6, New York, NY, USA, 2015. ACM.
11. Bayerl, S. Granitzer, M.: Data-transformation on historical data using the rdf data cube vocabulary. In *Proceedings of the 15th International Conference on Knowledge Technologies and Data-driven Business*, i-KNOW '15, pages 15:1–15:8, New York, NY, USA, 2015. ACM.
12. Becker, K., Jahangiri, S., Knoblock, C. A.: A quantitative survey on the use of the cube vocabulary in the linked open data cloud. In *Proceedings of 3rd International Workshop on Semantic Statistics (SemStats 2015)*, 2015.
13. Becker, K., Tan, X., Jahangiri, S., Knoblock, C. A.: Finding, assessing, and integrating statistical sources for data mining. In *KNOW@ LOD*, 2015.
14. Bosch, T., Cyganiak, R., Gregory, A., Wackerow, J.: Ddi-rdf discovery vocabulary: A metadata vocabulary for documenting research and survey data. In *LDOW*, 2013.

15. Brasoveanua, A. M., Saboub, M., Scharla, A., Hubmann-Haidvogela, A., Fischla, D.: Visualizing statistical linked knowledge for decision support: *Semantic Web*, pages 1–25, 2016.
16. Capadisli, S., Auer, S., Ngonga Ngomo, A.-C.: Linked sdmx data: Path to high fidelity statistical linked data for oecd, bfs, fao, and ecb: *Semantic Web*, 2013.
17. Capadisli, S., Auer, S., Riedl, R.: Linked statistical data analysis: *Semantic Web Challenge*, 2013.
18. 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.
19. Celino, I. Calegari, G. R.: Geo-statistical exploration of milano datasets. In *Second International Workshop for Semantic Statistics SemStats*, 2014.
20. Ceolin, D., Nottamkandath, A., Fokkink, W. *Automated Evaluation of Annotators for Museum Collections Using Subjective Logic*, pages 232–239. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.
21. Cyganiak, R., Field, S., Gregory, A., Halb, W., Tennison, J.: Semantic statistics: Bringing together sdmx and scovo.: *LDOW*, 628, 2010.
22. Cyganiak, R., Hausenblas, M., McCuirc, E. *Official Statistics and the Practice of Data Fidelity*, pages 135–151. Springer New York, New York, NY, 2011.
23. Cyganiak, R., Reynolds, D., Tennison, J.: The rdf data cube vocabulary: *W3C Recommendation (January 2014)*, 2014.
24. Daga, E., d'Aquin, M., Gangemi, A., Motta, E.: Early analysis and debugging of linked open data cubes. In *Second International Workshop on Semantic Statistics*, 2014.
25. De Leon, A., Wisniewki, F., Villazón-Terrazas, B., Corcho, O.: Map4rdf-faceted browser for geospatial datasets. In *PMOD workshop*. Informatica, 2012.
26. Debattista, J., Lange, C., Auer, S.: Representing dataset quality metadata using multi-dimensional views. In *Proceedings of the 10th International Conference on Semantic Systems*, SEM '14, pages 92–99, New York, NY, USA, 2014. ACM.
27. Diamantini, C. Potena, D.: Semantic enrichment of strategic datacubes. In *Proceedings of the ACM 11th International Workshop on Data Warehousing and OLAP*, DOLAP '08, pages 81–88, New York, NY, USA, 2008. ACM.
28. Do, B.-L., Aryan, P. R., Trinh, T.-D., Wetz, P., Kiesling, E., Tjoa, A. M.: Toward a framework for statistical data integration. In *Proceedings of the 3rd International Workshop on Semantic Statistics co-located with 14th International Semantic Web Conference (ISWC 2015)*, 2015.
29. Do, B.-L., Trinh, T.-D., Aryan, P. R., Wetz, P., Kiesling, E., Tjoa, A. M.: Toward a statistical data integration environment: The role of semantic metadata. In *Proceedings of the 11th International Conference on Semantic Systems*, SEMANTICS '15, pages 25–32, New York, NY, USA, 2015. ACM.
30. Do, B.-L., Trinh, T.-D., Wetz, P., Kiesling, E., Anjomshoaa, A., Tjoa, A. M.: Multiscale exploration of spatial statistical datasets: A linked data mashup approach. In *The Second International Workshop on Semantic Statistics (SemStats 2014)*, 2014.
31. Ermilov, I., Auer, S., Stadler, C.: Csv2rdf: User-driven csv to rdf mass conversion framework. In *Proceedings of the ISEM*, volume 13, pages 04–06, 2013.
32. Ermilov, I., Martin, M., Lehmann, J., Auer, S. *Linked Open Data Statistics: Collection and Exploitation*, pages 242–249. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.
33. Etcheverry, L., Gomez, S. S., Vaisman, A.: Modeling and querying data cubes on the semantic web: *arXiv preprint arXiv:1512.06080*, 2015.

34. Etcheverry, L., Vaisman, A., Zimányi, E. *Modeling and Querying Data Warehouses on the Semantic Web Using QB4OLAP*, pages 45–56. Springer International Publishing, Cham, 2014.
35. Etcheverry, L. Vaisman, A. A. *Enhancing OLAP Analysis with Web Cubes*, pages 469–483. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.
36. Etcheverry, L. Vaisman, A. A.: Qb4olap: a new vocabulary for olap cubes on the semantic web. In *Proceedings of the Third International Conference on Consuming Linked Data-Volume 905*, pages 27–38. CEUR-WS. org, 2012.
37. Fernández, A. V. Zarrabeitia, A. S.: Implementation of a linked open data solution for the statistics agency of cantabria's metadata and data bank. In *Proceedings of the 2013 International Conference on Dublin Core and Metadata Applications*, pages 1–8. Citeseer, 2013.
38. Follenfant, C., Trastour, D., Corby, O.: A model for assisting business users along analytical processes. In *SPIM-2nd Workshop on Semantic Personalized Information Management: Retrieval and Recommendation-2012*, volume 781, pages 38–41. CEUR-WS, 2011.
39. Frischmuth, P., Martin, M., Tramp, S., Riechert, T., Auer, S.: Ontowiki–an authoring, publication and visualization interface for the data web: *Semantic Web*, 6(3):215–240, 2015.
40. Gayo, J. E. L., Farhan, H., Fernández, J. C., Rodráguez, J.: Representing verifiable statistical index computations as linked data. In *Second International Workshop for Semantic Statistics SemStats*, 2014.
41. Gayo, J. E. L. Rodrıguez, J. M. A.: Validating statistical index data represented in rdf using sparql queries. In *RDF Validation Workshop. Practical Assurances for Quality RDF Data, Cambridge, Ma, Boston*. Citeseer, 2013.
42. Gottron, T., Hachenberg, C., Harth, A., Zapilko, B.: Towards a semantic data library for the social sciences. In *SDA*, pages 48–59. Citeseer, 2011.
43. Gupta, K., Lambhate, P., Emmanual, M.: Processing linked multidimensional data on the semantic web. In *International Conference on Computing, Communication and Energy Systems*, 2016.
44. Hallo, M., Luján-Mora, S., Maté, A.: Publishing a scorecard for evaluating the use of open-access journals using linked data technologies. In *Computer, Information and Telecommunication Systems (CITS), 2015 International Conference on*, pages 1–5. IEEE, 2015.
45. Hallo, M., Luján-Mora, S., Maté, A.: Evaluating open access journals using semantic web technologies and scorecards: *Journal of Information Science*, page 0165551515624353, 2016.
46. Hallo, M., Luján Mora, S., Trujillo Mondéjar, J. C., et al.: An approach to publish statistics from open-access journals using linked data technologies. In *9th International Technology, Education and Development Conference*. IATED, International Association of Technology, Education and Development, 2015.
47. Hartmann, T., Zapilko, B., Wackerow, J., Eckert, K.: Constraints to validate rdf data quality on common vocabularies in the social, behavioral, and economic sciences: *arXiv preprint arXiv:1504.04479*, 2015.
48. Hartmann, T., Zapilko, B., Wackerow, J., Eckert, K.: Evaluating the quality of rdf data sets on common vocabularies in the social, behavioral, and economic sciences: *arXiv preprint arXiv:1504.04478*, 2015.
49. Hausenblas, M., Halb, W., Raimond, Y., Feigenbaum, L., Ayers, D. *SCOVO: Using Statistics on the Web of Data*, pages 708–722. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.

50. Helmich, J., Klímek, J., Nečaský, M. *Visualizing RDF Data Cubes Using the Linked Data Visualization Model*, pages 368–373. Springer International Publishing, Cham, 2014.
51. Hoefler, P., Granitzer, M., Veas, E. E., Seifert, C.: Linked data query wizard: A novel interface for accessing sparql endpoints. In *LDOW*, 2014.
52. Höffner, K. Lehmann, J.: Towards question answering on statistical linked data. In *Proceedings of the 10th International Conference on Semantic Systems*, SEM '14, pages 61–64, New York, NY, USA, 2014. ACM.
53. Höffner, K., Martin, M., Lehmann, J.: Linkedspending: Openspending becomes linked open data: *Semantic Web*, 7(1):95–104, 2015.
54. Ibragimov, D., Hose, K., Pedersen, T. B., Zimányi, E. *Towards Exploratory OLAP Over Linked Open Data – A Case Study*, pages 114–132. Springer Berlin Heidelberg, Berlin, Heidelberg, 2015.
55. Jakobsen, K. A., Andersen, A. B., Hose, K., Pedersen, T. B.: Optimizing rdf data cubes for efficient processing of analytical queries. In *Ceur Workshop Proceedings*. CEUR-WS. org, 2015.
56. Janev, V., Mijovic, V., MiloSevic, U., Vranes, S.: Supporting the linked data publication process with the lod2 statistical workbench: *Semantic Web Journal (Submitted) http://www. semantic-web-journal. net/content/supporting-linked-datapublication-process-lod2-statistical-workbench*, 2014.
57. Janev, V., Mijović, V., Paunović, D., Milošević, U. *Modeling, Fusion and Exploration of Regional Statistics and Indicators with Linked Data Tools*, pages 208–221. Springer International Publishing, Cham, 2014.
58. Janev, V., Mijović, V., Vraneš, S.: Lod2 tool for validating rdf data cube models. In *Web Proceedings of the 5th ICT Innovations Conference*, pages 12–15, 2013.
59. Janev, V., Milošević, U., Spasić, M., Vraneš, S., Milojković, J., Jireček, B.: Integrating serbian public data into the lod cloud. In *Proceedings of the Fifth Balkan Conference in Informatics*, BCI '12, pages 94–99, New York, NY, USA, 2012. ACM.
60. Kalampokis, E., Karamanou, A., Nikolov, A., Haase, P., Cyganiak, R., Roberts, B., Hermans, P., Tambouris, E., Tarabanis, K.: Creating and utilizing linked open statistical data for the development of advanced analytics services. In *Second international workshop for semantic statistics, SemStats2014. CEUR-WS. org*, 2014.
61. 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. In *Proceedings of the 2014 International Conference on Posters & Demonstrations Track - Volume 1272*, ISWC-PD'14, pages 137–140, Aachen, Germany, Germany, 2014. CEUR-WS.org.
62. Kalampokis, E., Roberts, B., Karamanou, A., Tambouris, E., Tarabanis, K.: Challenges on developing tools for exploiting linked open data cubes. In *Proceedings of the 3rd International Workshop on Semantic Statistics (SemStats2015) within the 14th International Semantic Web Conference (ISWC2015)*, volume 1551, 2015.
63. Kalampokis, E., Tambouris, E., Tarabanis, K. *Linked Open Government Data Analytics*, pages 99–110. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.
64. Kalampokis, E., Tambouris, E., Tarabanis, K.: Ict tools for creating, expanding, and exploiting statistical linked open data: *Statistical Journal of the IAOS*, 2016.
65. Kamateri, E., Kalampokis, E., Tambouris, E., Tarabanis, K.: The linked medical data access control framework: *Journal of biomedical informatics*, 50:213–225, 2014.

66. Kämpgen, B. *DC Proposal: Online Analytical Processing of Statistical Linked Data*, pages 301–308. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.
67. Kämpgen, B. Harth, A.: Transforming statistical linked data for use in olap systems. In *Proceedings of the 7th International Conference on Semantic Systems*, I-Semantics '11, pages 33–40, New York, NY, USA, 2011. ACM.
68. Kämpgen, B. Harth, A. *No Size Fits All – Running the Star Schema Benchmark with SPARQL and RDF Aggregate Views*, pages 290–304. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.
69. Kämpgen, B. Harth, A. *OLAP4LD – A Framework for Building Analysis Applications Over Governmental Statistics*, pages 389–394. Springer International Publishing, Cham, 2014.
70. Kämpgen, B., ORiain, S., Harth, A.: Interacting with statistical linked data via olap operations. In *Extended Semantic Web Conference*, pages 87–101. Springer, 2012.
71. Kämpgen, B., Stadtmüller, S., Harth, A. *Querying the Global Cube: Integration of Multidimensional Datasets from the Web*, pages 250–265. Springer International Publishing, Cham, 2014.
72. Kämpgen, B., Weller, T., O'Riain, S., Weber, C., Harth, A. *Accepting the XBRL Challenge with Linked Data for Financial Data Integration*, pages 595–610. Springer International Publishing, Cham, 2014.
73. Khan, Y., Saleem, M., Iqbal, A., Mehdi, M., Hogan, A., Ngomo, A.-C. N., Decker, S., Sahay, R.: Safe: Policy aware sparql query federation over rdf data cubes. In *SWAT4LS*, 2014.
74. Koho, M., Hyvönen, E., Lehikoinen, A. *Ornithology Based on Linking Bird Observations with Weather Data*, pages 75–85. Springer International Publishing, Cham, 2014.
75. Lefort, L., Bobruk, J., Haller, A., Taylor, K., Woolf, A.: A linked sensor data cube for a 100 year homogenised daily temperature dataset. In *Proceedings of the 5th International Conference on Semantic Sensor Networks-Volume 904*, pages 1–16. CEUR-WS. org, 2012.
76. Lefort, L. Leroux, H.: Design and generation of linked clinical data cubes. In *Proceedings of SemStats 2013. Sydney, Australia*, 2013.
77. Leforta, L., Hallera, A., Taylora, K., Woolfb, A.: The acorn-sat linked climate dataset: *Semantic Web*, 2013.
78. Leroux, H. Lefort, L.: Using cdisc odm and the rdf data cube for the semantic enrichment of longitudinal clinical trial data. In *SWAT4LS*. Citeseer, 2012.
79. Leroux, H. Lefort, L.: Semantic enrichment of longitudinal clinical study data using the cdisc standards and the semantic statistics vocabularies: *Journal of biomedical semantics*, 6(1):1, 2015.
80. Llaves, A., Corcho, O., Fernandez-Carrera, A.: Map4rdf-ios: a tool for exploring linked geospatial data. In *Proceedings of Workshop on Linked Geospatial Data*, 2014.
81. Lodi, G., Maccioni, A., Scannapieco, M., Scanu, M., Tosco, L.: Publishing official classifications in linked open data. In *SemStats 2014. Springer, Riva del Garda, Italy*, 2014.
82. Mader, C., Martin, M., Stadler, C. *Facilitating the Exploration and Visualization of Linked Data*, pages 90–107. Springer International Publishing, Cham, 2014.
83. Martin, M., Abicht, K., Stadler, C., Ngonga Ngomo, A.-C., Soru, T., Auer, S.: Cubeviz: Exploration and visualization of statistical linked data. In *Proceedings of the 24th International Conference on World Wide Web*, WWW '15 Companion, pages 219–222, New York, NY, USA, 2015. ACM.

84. Martin, M., van Nuffelen, B., Abruzzini, S., Auer, S.: The digital agenda scoreboard: A statistical anatomy of europes way into the information age: *Semantic Web Jorunal*, 2012.
85. Maté, A., Llorens, H., de Gregorio, E. *An Integrated Multidimensional Modeling Approach to Access Big Data in Business Intelligence Platforms*, pages 111–120. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.
86. Matei, A., Chao, K.-M., Godwin, N. *OLAP for Multidimensional Semantic Web Databases*, pages 81–96. Springer Berlin Heidelberg, Berlin, Heidelberg, 2015.
87. McCusker, J. P., McGuinness, D. L., Lee, J., Thomas, C., Courtney, P., Tatalovich, Z., Contractor, N., Morgan, G., Shaikh, A.: Towards next generation health data exploration: a data cube-based investigation into population statistics for tobacco. In *System Sciences (HICSS), 2013 46th Hawaii International Conference on*, pages 2725–2732. IEEE, 2013.
88. Mehdi, M., Sahay, R., Derguech, W., Curry, E.: On-the-fly generation of multidimensional data cubes for web of things. In *Proceedings of the 17th International Database Engineering & Applications Symposium*, IDEAS '13, pages 28–37, New York, NY, USA, 2013. ACM.
89. Meimaris, M. Papastefanatos, G.: Containment and complementarity relationships in multidimensional linked open data: *Semantic Statistics (SEMSTATS)*, 2014.
90. Meimaris, M., Papastefanatos, G., Vassiliadis, P., Anagnostopoulos, I.: Efficient computation of containment and complementarity in rdf data cubes. In *EDBT 2016*, 2016.
91. Meroño-Peñuela, A. *LSD Dimensions: Use and Reuse of Linked Statistical Data*, pages 159–163. Springer International Publishing, Cham, 2015.
92. Meroño-Peñuela, A., Ashkpour, A., Guéret, C.: From flat lists to taxonomies: Bottom-up concept scheme generation in linked statistical data. In *Proceedings of the 2nd International Workshop on Semantic Statistics, Riva del Garda, Italy*, 2014.
93. Merono-Penuela, A., Ashkpour, A., Rietveld, L., Hoekstra, R.: Linked humanities data: The next frontier? a case-study in historical census data. In *CEUR Workshop Proceedings*, volume 951, 2012.
94. Meroño-Peñuela, A., Guéret, C., Ashkpour, A., Schlobach, S.: Cedar: The dutch historical censuses as linked open data: *Semantic Web–Interoperability, Usability, Applicability*, 2015.
95. Meroño-Peñuela, A., Guéret, C., Hoekstra, R., Schlobach, S.: Detecting and reporting extensional concept drift in statistical linked data. In *1st International Workshop on Semantic Statistics (SemStats 2013), ISWC. CEUR*, 2013.
96. Meroño-Peñuela, A., Guéret, C., Schlobach, S.: Linked edit rules: A web friendly way of checking quality of rdf data cubes. In *Proceedings of the Third International Workshop on Semantic Statistics (SemStats 15) co-located with Fourteenth International Semantic Web Conference (ISWC15), Bethlehem, PA, USA*, pages 1–12, 2015.
97. Miloevi, U., Janev, V., Spasi, M., Milojkovi, J., Vrane, S.: Publishing statistical data as linked open data: *Proceedings of the 2nd International Conference on Information Society Technology, Information Society of the Republic of Serbia*, 2012.
98. Mutlu, B., Hoefler, P., Sabol, V., Tschinkel, G., Granitzer, M.: Automated visualization support for linked research data.: *I-SEMANTICS (Posters & Demos)*, 1026:40–44, 2013.

99. Mutlu, B., Hoefler, P., Tschinkel, G., Veas, E., Sabol, V., Stegmaier, F., Granitzer, M.: Suggesting visualisations for published data. In *Information Visualization Theory and Applications (IVAPP), 2014 International Conference on*, pages 267–275. IEEE, 2014.
100. Mynarz, J., Cyganiak, R., Hausenblas, M., Iqbal, A.: Modelling of statistical linked data: *Proceedings of Znalosti 2011*, 2011.
101. Nguyen, T. B. Ngo, S. N.: Semantic cubing platform enabling interoperability analysis among cloud-based linked data cubes. In *Proceedings of the 16th International Conference on Information Integration and Web-based Applications & Services*, iiWAS '14, pages 547–553, New York, NY, USA, 2014. ACM.
102. Patton, E. W., Brown, E., Poegel, M., De Los, H., Santos, C. F., Bennett, K. P., McGuinness, D. L.: Semnext: A framework for semantically integrating and exploring numeric analyses. In *In Proceedings of SemStats 2015*, 2015.
103. Perakis, K., Bouras, T., Ntalaperas, D., Hasapis, P., Georgousopoulos, C., Sahay, R., Beyan, O. D., Potlog, C., Usurelu, D.: Advancing patient record safety and ehr semantic interoperability. In *2013 IEEE International Conference on Systems, Man, and Cybernetics*, pages 3251–3257. IEEE, 2013.
104. Petrou, I., Meimaris, M., Papastefanatos, G.: Towards a methodology for publishing linked open statistical data: *JeDEM-eJournal of eDemocracy and Open Government*, 6(1):97–105, 2014.
105. 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*, WOD '13, pages 4:1–4:3, New York, NY, USA, 2013. ACM.
106. Pirrotta, G.: Linking italian university statistics. In *Proceedings of the 6th International Conference on Semantic Systems*, I-SEMANTICS '10, pages 2:1–2:10, New York, NY, USA, 2010. ACM.
107. Prat, N., Akoka, J., Comyn-Wattiau, I.: Transforming multidimensional models into owl-dl ontologies. In *2012 Sixth International Conference on Research Challenges in Information Science (RCIS)*, pages 1–12. IEEE, 2012.
108. Prat, N., Megdiche, I., Akoka, J.: Multidimensional models meet the semantic web: Defining and reasoning on owl-dl ontologies for olap. In *Proceedings of the Fifteenth International Workshop on Data Warehousing and OLAP*, DOLAP '12, pages 17–24, New York, NY, USA, 2012. ACM.
109. Ristoski, P., Bizer, C., Paulheim, H.: Mining the web of linked data with rapidminer: *Web Semantics: Science, Services and Agents on the World Wide Web*, 35:142–151, 2015.
110. Rodríguez, J. M. A., Clement, J., Gayo, J. E. L., Farhan, H., De Pablos, P. O.: Publishing statistical data following the linked open data principles: The web index project: *IGI Global*, pages 199–226, 2013.
111. Roussakis, Y., Chrysakis, I., Stefanidis, K., Flouris, G., Stavrakas, Y.: A flexible framework for defining, representing and detecting changes on the data web: *CoRR*, abs/1501.02652, 2015.
112. Ruback, L., Pesce, M., Manso, S., Ortiga, S., Salas, P. E. R., Casanova, M. A.: A mediator for statistical linked data. In *Proceedings of the 28th Annual ACM Symposium on Applied Computing*, SAC '13, pages 339–341, New York, NY, USA, 2013. ACM.
113. Saad, R., Teste, O., Trojahn, C.: Olap manipulations on rdf data following a constellation model. In *First International Workshop on Semantic Statistics, collocated with the 12th International Semantic Web Conference, Sydney, page (on line). DataLift*, 2013.

114. Sabol, V., Tschinkel, G., Veas, E., Hoefler, P., Mutlu, B., Granitzer, M. *Discovery and Visual Analysis of Linked Data for Humans*, pages 309–324. Springer International Publishing, Cham, 2014.
115. Sabou, M., Braoveanu, A. M. P., Önder, I. *Linked Data for Cross-Domain Decision-Making in Tourism*, pages 197–210. Springer International Publishing, Cham, 2015.
116. Salas, R. P. E., Martin, M., Da Mota, F. M., Auer, S., Breitman, K., Casanova, M. A.: Publishing statistical data on the web. In *Semantic Computing (ICSC), 2012 IEEE Sixth International Conference on*, pages 285–292. IEEE, 2012.
117. Salas, R. P. E., Martin, M., Da Mota, F. M., Auer, S., Breitman, K. K., Casanova, M. A.: Olap2datacube: An ontowiki plug-in for statistical data publishing. In *Proceedings of the Second International Workshop on Developing Tools as Plug-Ins*, pages 79–83. IEEE Press, 2012.
118. Sato, H. Wen, W.: Towards easy matching between statistical linked data: Dimension patterns. In *International Semantic Web Conference*, 2013.
119. Schlegel, K., Bayerl, S., Zwicklbauer, S., Stegmaier, F., Seifert, C., Granitzer, M., Kosch, H. *Trusted Facts: Triplifying Primary Research Data Enriched with Provenance Information*, pages 268–270. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.
120. Schütz, C., Neumayr, B., Schrefl, M. *Business Model Ontologies in OLAP Cubes*, pages 514–529. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.
121. Seifert, C., Granitzer, M., Höfler, P., Mutlu, B., Sabol, V., Schlegel, K., Bayerl, S., Stegmaier, F., Zwicklbauer, S., Kern, R. *Crowdsourcing Fact Extraction from Scientific Literature*, pages 160–172. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.
122. Southall, H. R. Stoner, M. J.: Creating a spatio-temporal data feed api for a large and diverse library of historical statistics for areas within britain. In *GIS Research UK 2015*, 2015.
123. Stegmaier, F., Seifert, C., Kern, R., Höfler, P., Bayerl, S., Granitzer, M., Kosch, H., Lindstaedt, S., Mutlu, B., Sabol, V., Schlegel, K., Zwicklbauer, S. *Unleashing Semantics of Research Data*, pages 103–112. Springer Berlin Heidelberg, Berlin, Heidelberg, 2014.
124. Tambouris, E., Kalampokis, E., Tarabanis, K.: Processing linked open data cubes. In *International Conference on Electronic Government*, pages 130–143. Springer, 2015.
125. Tarasova, T., Argenti, M., Marx, M. *Semantically-Enabled Environmental Data Discovery and Integration: Demonstration Using the Iceland Volcano Use Case*, pages 289–297. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.
126. Tilahun, B., Kauppinen, T., Keßler, C., Fritz, F.: Design and development of a linked open data-based health information representation and visualization system: Potentials and preliminary evaluation: *JMIR medical informatics*, 2(2), 2014.
127. Trinh, T.-D., Do, B.-L., Wetz, P., Anjomshoaa, A., Tjoa, A. M.: Linked widgets: An approach to exploit open government data. In *Proceedings of International Conference on Information Integration and Web-based Applications & Services*, IIWAS '13, pages 438:438–438:442, New York, NY, USA, 2013. ACM.
128. Tschinkel, G., Veas, E., Mutlu, B., Sabol, V.: Using semantics for interactive visual analysis of linked open data. In *Proceedings of the 2014 International Conference on Posters & Demonstrations Track - Volume 1272*, ISWC-PD'14, pages 133–136, Aachen, Germany, Germany, 2014. CEUR-WS.org.
129. Vaisman, A. Zimányi, E. *Data Warehouses and the Semantic Web*, pages 539–576. Springer Berlin Heidelberg, Berlin, Heidelberg, 2014.

130. van der Waal, S., Wecel, K., Ermilov, I., Janev, V., Milošević, U., Wainwright, M. *Lifting Open Data Portals to the Data Web*, pages 175–195. Springer International Publishing, Cham, 2014.
131. Van Nuffelen, B., Janev, V., Martin, M., Mijovic, V., Tramp, S. *Supporting the Linked Data Life Cycle Using an Integrated Tool Stack*, pages 108–129. Springer International Publishing, Cham, 2014.
132. Vilches-Blázquez, L. M., Villazón-Terrazas, B., Corcho, O., Gómez-Pérez, A.: Integrating geographical information in the linked digital earth: *International Journal of Digital Earth*, 7(7):554–575, 2014.
133. Vrandecic, D., Lange, C., Hausenblas, M., Bao, J., Ding, L.: Semantics of governmental statistics data: *Semantic Web*, 2010.
134. Wagner, A., Haase, P., Rettinger, A., Lamm, H.: Discovering related data sources in data-portals. In *First International Workshop on Semantic Statistics*, 2013.
135. Zancanaro, A., Pizzol, L., de Moura Speroni, R., Todesco, J. L., Gauthier, F. O.: Publishing multidimensional statistical linked data. In *Proceedings of the Fifth International Conference on Information, Process, and Knowledge Management*, pages 290–304, 2013.
136. Zapilko, B. Mathiak, B.: Performing statistical methods on linked data. In *International conference on dublin core and metadata applications*, pages 116–125, 2011.
137. Zapilko, B. Mathiak, B. *Object Property Matching Utilizing the Overlap between Imported Ontologies*, pages 737–751. Springer International Publishing, Cham, 2014.
138. Zaveri, A., Lehmann, J., Auer, S., Hassan, M. M., Sherif, M. A., Martin, M.: Publishing and interlinking the global health observatory dataset: *Semantic Web*, 4(3):315–322, 2013.
139. Zaveri, A., Pietrobon, R., Auer, S., Lehmann, J., Martin, M., Ermilov, T.: Redd-observatory: Using the web of data for evaluating the research-disease disparity. In *Proceedings of the 2011 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology - Volume 01*, WI-IAT '11, pages 178–185, Washington, DC, USA, 2011. IEEE Computer Society.