Publishing Multidimensional Statistical Linked Data

Airton Zancanaro, Leandro Dal Pizzol, Rafael de Moura Speroni, José Leomar Todesco, Fernando O. Gauthier

Department of knowledge engineering – Universidade Federal de Santa Catarina (UFSC)

Florianópolis - Brazil

{airtonz; leandro; speroni; tite; gauthier}@egc.ufsc.br

Abstract—The access information law approved by the Brazilian government regulates the provision of open government data in the Web. However, they are heterogeneous, unstructured and derived from independent sources, making it difficult to interconnect. This paper presents a process of identifying sources, ontology generation, mapping and publishing statistical linked data in the form of multidimensional cubes, represented by the RDF Data Cube Vocabulary. In this process, data are transformed and assigned semantic meaning through its connection with domain ontologies. Through a web application, the publication of these data is automated, allowing for future analysis operations with the use of Online Analytical Processing (OLAP). As a result, the approach is expected to increase the scale in the publication of statistical linked data, and therefore, increasing the potential for analysis.

Keywords–Linked Data; OLAP; Open Data; RDF Data Cube; Statistical Data.

I. INTRODUCTION

In November, 18th, 2011, Brazil approved the Information Access Law ("Lei de Acesso à Informação", law number 12.527), regulating and granting the right to access public information of the Brazilian government, which is assured by its Federal Constitution. Taking effect in May, 16th, 2012, such law represents a big leap towards transparency and citizenship, forcing public agencies to consider openness a rule and confidentiality an exception, broadening citizen participation in governmental actions.

Such initiative gave rise to a higher availability of data on the *Web,* being originated from several sectors of public administration. Even so, such data is typically heterogeneous and has no integrated statistic treatment [1]. Moreover, there are no available means to expose, share or link the data so that they can add more information [2].

Statistical data are important sources of information for: a) the Government, as a way of verifying the Strong and weak points of their administration, therefore contributing for better policymaking decisions; b) science, as an important tool for accepting or rejecting a theory; and c) for business, as a way of supporting strategic decisions of the administration. For that matter, it is paramount that such data are semantically linked to ontologies or knowledge databases [3].

Two of the main challenges mentioned by Kämpgen, O'Rain and Harth [4] towards the use of OLAP in statistical linked data are: a) OLAP requires a data cube model, with dimensions and measurements; b) OLAP queries are complex and require specialized data models, such as the star model in relational databases, to run efficiently.

On that basis, the Federal University of Santa Catarina's (UFSC) Knowledge Engineering and Management Graduation Program (EGC) researchers, through the Knowledge Engineering Laboratory (LEC), developed a supporting tool for the publishing of statistical data series in an open multidimensional model pattern, using OLAP data cubes. The main contribution of this paper is to automate the process of cube construction that is extremely complex when running on a non-automated way. Furthermore, enables semantic search while using SPARQL [5] language query over these datasets. Furthermore, it enables semantic search while using SPARQL language query over these datasets.

Therefore the following technologies were combined: OLAP Data Cube, ontologies and linked open data, resulting in a functional tool for generating statistical data. This method uses an ontology named cube and makes the Extraction, Transformation and Loading (ETL) in an OLAP structure that is mounted according to this ontology.

The research related to this work is further presented, in Section II. In Section III, RDF language is described. In Section IV, we explain the pattern for exchanging and sharing SDMX data. In Section V, the RDF Data Cube vocabulary is described. Section VI addresses the obtainment, processing and publication of data. The final considerations are in Section VII.

II. RELATED WORKS

Researches, such as Hull [6], point out the integration of different databases for the purpose of adding more content to what is being searched on the Web. Thus, the primary intention of linked open data [7] is to publish, share and connect different databases openly available on the Web. In order to make this possible, Berners-Lee [8] identified four principles that standardize the publication of the data that form the so-called Web of data: use Uniform Resource Identifiers (URIs) to identify things, use HTTP to find these names on the web, provide useful information in the form of Resource Description Framework (RDF) [9], and include links to other URIs so that you can discover more things. These principles, associated with the use of ontologies [10] and the SPARQL query language [11] form a set of

technologies that have been established for the publication and the integration of data [12], [13] and [14].

These principles can be observed in the work of Kämpgen and Harth [15], which aimed at using linked data from various databases available on the web in an OLAP system. For that, the transformation and integration of data were made into an appropriate format using OLAP operations and SPARQL queries.

The OLAP, also known as "OLAP cube" [10], was officially described in an article submitted to Arbor Software Corp. in 1993 entitled "Providing OLAP (Online Analytical Processing) to User-Analysts: An IT Mandate", written by W.H. Inmon, R. Kimball, and E.F. Codd [16], although its concept is known more earlier [17]. Date [17] defines OLAP as an interactive process of creating, managing, analyzing and reporting data. Typically its operations are roll-up (increases the level of aggregation) and drill-down (decreasing aggregation and increases the breakdown providing a smaller granularity) [18] along one or more dimensions. Slice and dice (selection and projection) are responsible for working with the information, changing positions whenever necessary and pivoting (reorientation and multidimensional view of data), with the ability to summarize and group data in various formats [19].

Works, such as Kämpgen, O'Rain and Harth [4], suggest a new way of interacting with linked statistical data in an RDF-modeled cube, a format of structured data that enables querying to multiple data sources through the use of SPARQL language. To this end Zapilko and Mathiak [1] developed an approach for the purpose of assisting researchers to statistically analyze the linked data with the aid of SPARQL.

The main advantage for the publication and consumption statistical data, according to Kämpgen [20], is the ease of integration and enrichment of data using other sources. In addition Cyganiak et al. [21] shows a number of benefits to publish statistics using RDF standards: a) the ability to access the annotations generated by third parties, b) statistical data can integrate a wider range of linked data, c) the possibility to perform operations of slice and dice in the datasets and the granularity of the information, and d) the flexibility offered by RDF in publishing statistical data.

For Salas et al. [3], statistical data are the main source of information for governments, researchers and administrators. In this sense, the author proposes two tools that use the RDF Data Cube vocabulary in order to provide the representation of statistical data in the multidimensional format. The first is the *OLAP2DataCube*, which allows the analysis of a large amount of data and its efficient transformation into RDF. The second tool, which is the *CSV2DataCube*, offers conditions to transform data available in CSV format to RDF.

On the other hand, the RDF data cube vocabulary was, according to Follenfant, Trastour and Corby [22], introduced by Cyganiak, Reynolds and Tennison [23] with

the purpose of allowing the publication of statistical data on the Web, providing a metamodel for a set of multidimensional data [2].

Finally, Casanova et al. [24] highlights the lessons learned in the conversion of government data in RDF and graphically presents a comparison between the American and Brazilian data.

III. RDF

The RDF is a language originally created to represent and identify semantic content and information on Web pages [9]. However, in a generalized concept of Web, RDF can be used to identify any information or resource existing in the Web. Normally, it is used when the information that is wanted to be retrieved will be processed by machine rather than only displayed to the user.

According to Manola and Miller [9], RDF provides a common way of expressing information enabling the exchange of information among the applications without losing the meaning.

Fundamentally, the RDF vocabulary is fully extensible and consists of identifying objects via URIs [25]. URIs are used to name things, identify resources and properties in RDF.

The RDF properties may be thought of as attributes of resources and thus corresponds to traditional attribute-value pairs [26]. They also represent relationships between resources, resembling to an entity-relationship diagram where resources correspond to objects and properties correspond to instance variables.

The RDF data model consists of three basic types of objects [27] they are:

a) Resources: all that can be described by RDF expressions and identified by a URI (whole or part of web pages, a figure, or even an object that is not directly accessible via the Web, for example: a printed book);

b) Properties: specific aspects, characteristics, attributes or relations used to describe a resource. Each property has a specific meaning, defines its permitted values, the types of resources that can describe, and its relationship with other properties [26] and;

c) Sentences: structured information composed of subject (resource), predicate (property) and object (property value). The object of a statement can be another resource or it can be a literal, that is, a resource (specified by a URI), a simple string or other primitive data type defined by XML.

Figure 1 illustrates the sentence "the archive 'exemplo.rdf' was created by the Knowledge Engineering Laboratory LEC" by using the syntax and RDF.

```
1  <rdf:RDF
2      xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
3      xmlns:dc="purl.org/dc/elements/1.1/">
4      <rdf:Description rdf:about="http://www.lec.ufsc.br/LEC/exemplo.rdf">
5          <dc:creator>LEC</dc:creator>
6      </rdf:Description>
7  </rdf:RDF>
```

Figure 1. Sentence "the archive exemplo.rdf was created by LEC",
expressed in RDF syntax.

The same sentence presented by Figure 1 can be expressed graphically by using arcs (properties) and nodes (resources or objects) as illustrated in Figure 2.
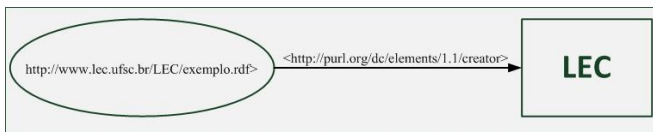
Figure 2. Sentence of Figure 1 graphically expressed.

Another way to express the sentences RDF is the Notation 3 (N3) [28], in which the three elements are listed in order: subject, predicate and object. Its objectives are to optimize the expression of logic and data in the same language and to allow the RDF can be expressed and rules to be integrated seamlessly to the RDF to be the most readable, natural and symmetrical as possible.

In order to promote interoperability and comparability between datasets using RDF, Milošević et al. [2] describes a syntax for the exchange and sharing of statistical data and metadata known as SDMX-RDF which will be presented in sequence.

IV. SDMX

Proposed in 2001 by the International Organization for Standardization (ISO), Statistical Data and Metadata eXchange (SDMX) is a standard for exchanging and sharing of statistical data and metadata between organizations. The SDMX standard for exchange of such data is a joint proposition of seven organizations worldwide, among which the U.S. Federal Reserve Federal Reserve, European Central Bank, the World Health Organization (WHO), the International Monetary Fund (IMF) and the United Nations (UN).

The SDMX has the SDMX-RDF syntax which, according to Cyganiak, Reynolds and Tennison [23], consists of the same model information specified in SDMX, but with information expressed in RDF, allowing the simple discovery and publication of linked data to the Web. SDMX-RDF defines classes and predicates to represent RDF statistical data compatible with the SDMX information model.

The key component of the SDMX standards, according to Cyganiak, Reynolds and Tennison [23], is the Content-Oriented Guidelines (COGs), a set of domain concepts, code lists and categories that support interoperability and comparability between data sets, providing a common language SDMX between applications. The RDF versions of these components are available as part of SDMX-RDF, and should be reused where possible.

V. RDF DATA CUBE VOCABULARY

The concept of multidimensional modeling, as it is known today, was proposed by Kimball [29] and subsequently deepened and enhanced in Kimball [30]. According to the Kimball [30], the great advantage of the multidimensional model is its simplicity, which is essential to enable users to understand databases, and allow recovery in an efficient way.

Multidimensional models are designed to store statistical data sets that, according to Cyganiak, Reynolds and Tennison [23], comprise a collection of observations made at some points across a logical space. This collection is characterized by a set of dimensions which define the scope of each observation along with metadata describing what was measured, as was measured and how the observations are expressed.

Statistical data can be set in a multidimensional way in space, that is, as a hypercube. A cube is arranged according to a set of dimensions, attributes and measures.

The dimensions are used to identify the observations, that is, the set of values for each dimension represents a single observation. Examples of dimensions include the time that the observation applies, or the geographic region that the observation covers.

The attributes, for its part, qualify and interpret the observed values. They allow the specification of measurement units and scale factors.

Lastly, the measures represent the phenomenon to be observed, for example, the population growth of a municipality.

The formalization of the understanding of data cubes based on Data Cube vocabulary (QB) is presented in the following and illustrated in Figure 3.

With QB vocabulary as points Kämpgen, O'Rain and Harth [4], there is greater ease in handling and ability to capture the statistical semantics of the linked data. As examples of projects that use the same vocabulary we have the UK government program *Combined Online Information System* (COINS) and the North American program of the U.S. Security and Exchange Commission.
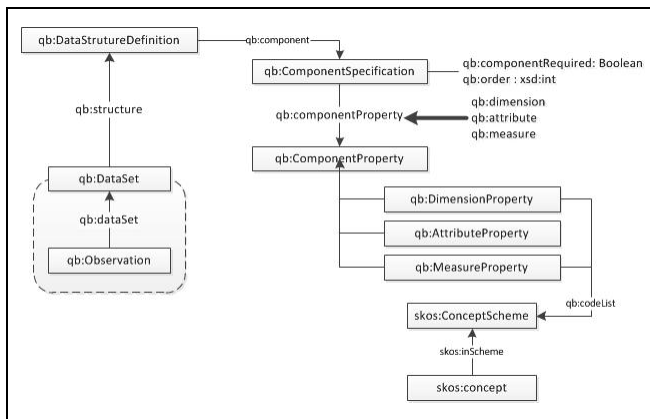
Figure 3. The data cube core [23].

Figure 3 illustrates, in a general way, the division of the *QB* structure in classes and properties which include the datasets used - its dimensions, attributes and measures - besides the observations and operations that will be applied on the data.

The main classes and properties of the structure of the data cube are:

a) DataSets: Class *qb:DataSet* - represents a collection of observations, possibly organized into multiple slices, as the common dimensional structure is determined;

b) Observations: Class *qb:Observation* - represents a single observation of the cube, which can have one or more measurement values associated. Related to this class we found properties *qb:dataset*, which indicates what set of data that observation belongs, and *qb:observation*, that indicates an observation contained within a slice of the data set;

c) Data Structure Definitions (DSD): Class: *qb:DataStructureDefinition* - defines the structure of data set (*Dataset*) or a slice. Associated to this class are the properties: *qb:structure*, which indicates the structure to which this data set belongs, and *qb:component*, responsible for the specification of the component that is included in the structure of the dataset;

d) Dimensions, Attributes and Measures: Class: *qb:ComponentProperty* - subclass of *rdf:Property*, a super abstract property of all properties that represent dimensions, attributes and measures. Class *qb:DimensionProperty* - represents components that form the cube dimensions. Class *qb:AttributeProperty* - formed by components which represent the attributes of the observations in the cube, for example, units of measure. Class *qb:MeasureProperty* - represents the measured values for the observed phenomenon. Associated to this class have the property *qb:measureType*, a generic measure of size. The value of this dimension indicates how far (within the set of measures of DSD) is provided by the value of observation, or other primary measure;

e) Slices: Class *qb:Slice* –denotes a subset of a dataset that is set by setting a subset of dimensional values. Its

property *qb:slice*, indicates the subset defined by setting a subset of the values of the dimension.

A. Multicubes

Cubes which share dimensions constitute a a dimensional model of multiple cubes. In *QB*, a cube corresponds to multiple cubes that use instances of *qb:ComponentProperty* with equivalence, and thus can be connected using the property *owl:sameAs*.

Similarly, members of a cube can be equivalent as in the case of Figure 4 which shows the relationship afforded by binding property *owl:sameAs*.
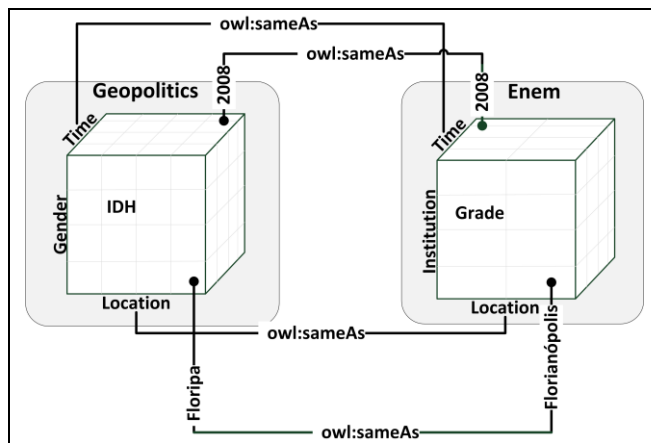

Figure 4. Example of a multicube.

In Figure 4, there are two datasets, one representing Brazilian Geopolitics and other representing the scholar data in Brazil (Enem [31]). Both have a dimension which denotes a geographical entity and has a member which denotes a city, such as Florianópolis. Also, both may use the same time dimension with literal values. If there is a statement *owl:sameAs* between the geographical dimensions, the two cubes can be represented as a multicube.

B. Relating OLAP operations to SPARQL in QB

The set of terms in an RDF Triple Store consists of URIs, blank nodes and literals. Triple store management system is a database for RDF, in which a triple (s, p o) is called RDF triple, where "s" is the subject, "p" represents the predicate and "o" is the object. These systems provide management and access to data through APIs and query languages for RDF data. Many Triple Quad Stores Stores are indeed due to the need to maintain the provenance of RDF data within the system. Any Triple Store that supports graphs will probably be a Quad Store [32] and [33].

Given a Triple Store with statistical linked data, we use basic queries in SPARQL about this dataset to return a specific set of multidimensional elements comprising their respective URIs or a blank node. In this section we present common OLAP queries on a multidimensional model using SPARQL in QB. The similar operations (projection, slice,

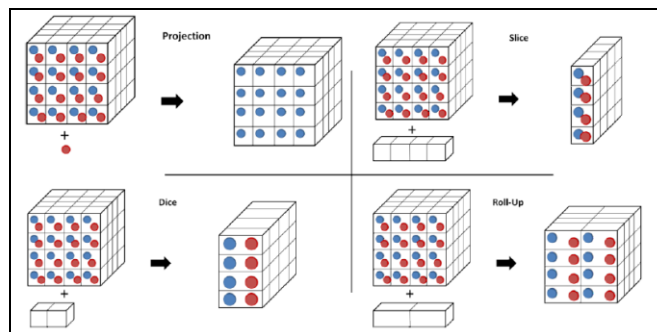dice and roll-up) can be found in [4], [34] and [35] and are illustrated in Figure 5.

Figure 5. Representation of OLAP operations with inputs e outputs. Source: [4].

a) Projection: DataCube x Measure → DataCube - removes a measure of the entry cube and allows you to see only a specific measure. In the example above, all triple referring to a measurement are removed, resulting in a query subcube;

b) Slice: DataCube x Dimension → DataCube - removes a dimension from the entry cube and all its contents added over the members of a dimension;

c) Dice: DataCube x Dimension x Value → DataCube - allows to filter and aggregate on certain dimension members. Note that Dice is not an selection operation, but a filter combined with the Slice operation;

d) Roll-up: DataCube x Dimension x Level → DataCube - allows to create a cube that contains instance data at a high level of aggregation. Remark: the Drill-Down operation has not been set yet, since it can be viewed as an inverse operation to roll up.

VI. COLLECTION, PROCESSING AND PUBLICATION OF DATA

The proposed for publication of data consists of four main steps: identification of sources identifying, ontology generation, mapping and publication of data; shown in Figure 6. Using the data from outside sources, it is intended to add semantic meaning by connecting them to other sources of data, thus publishing them in the form of a multidimensional cube.

Figure 6. Steps in the process of obtaining, processing and publication of data.

Sources Identifying: the composition of logical cubes composed of dimensions, measures, attributes, hierarchies and levels, allows to represent in a simple way real-world entities. Furthermore, it can facilitate the analysis of measures, to define which dimensions and attributes can represent significant data, and organize the dimensions of a given scope in levels and hierarchies.

In the choice of sources, some criteria must be observed such as the accuracy that indicates whether the values are stored in accordance with the actual values; the temporality, which indicates whether the recorded values are updated; the completeness, which indicates whether the needed values are stored and that these have an appropriate depth and width; and the consistency of the data [36]. Last but not least, the quality of the data must meet the requirements of its use.

To illustrate this, we used Brazilian governmental open data sources [37]. This choice is justified for these data sets have features such as: historical time series and the division by municipalities, considering also their spatial location, essential in the construction of cube.

Ontology Generation: the governmental data sources available have no semantic meaning and for this it is necessary that these sources are represented by a domain ontology. An ontology provides an explicit specification of a conceptualization [38] and its objective is to define which primitives are necessary for the representation of knowledge in a given context. In this process, the ontologies are used to provide a semantic representation of the dimensions and observations that describe the data to be published. In this stage, it is necessary to generate a proper ontology or to use an existing one to represent the data set.

Data Mapping: After the identification of the different data sources the ETL process is performed, through which data are integrated to form a single assembly. The data from this set are analyzed and linked to concepts represented in the ontologies. Thus, we want to clarify this relationship through a mapping that indicates, for each column of data, the URI of the corresponding concept. Also in this step the values contained in the dataset to the corresponding URIs are mapped, based on the representation of domain ontologies. Thus, each data value will be represented by the URI that identifies it.

Data Publishing: The publication step is the transformation of the data set already properly mapped to a multidimensional OLAP cube model. The adopted vocabulary for the publication is the RDF Data Cube, which represents, in the form of RDF, structures and standardized data appropriately for subsequent processing through OLAP operations. Furthermore, the fact of using RDF and URIs ensures integration of the dataset with the ontology and other datasets.

Figure 7 shows the interface through which the user makes the choice of the file containing the dataset as well as the definition of what are the dimensions that will form the multidimensional cube, and the measures that will make the facts stored in it.

Figure 7. Web application interface for publishing data.

Through the Web application developed, the dataset obtained is mapped, linked to other data sources and then converted into the format cube RDF.

In the processing each data set corresponds to an instance of the class *qb:Dataset*, which has its structure defined by an instance of *qb:DataStrucureDefinition*. This, in turn, is described in dimensions (represented by the property *qb:dimension*) and measures (represented by the property *qb:measure*).

Each of the observations that make up the cube are represented by an instance of the class *qb:Observation* (Figure 8), and its properties are associated with references to the *qb:Dataset* and for each of the dimensions and measures.

Figure 8. Example of an observation.

Figure 8 presents an example of representation of a observation of the cube, in RDF form. Lines 1 and 7 delimit the observation indicating that it is a by stating that it is an instance of the class *qb:Observation*. The line 2 indicates the observation this cube belongs to, represented by the property *qb:dataSet*. Lines 3 and 4 indicate the references for the dimension values, while lines 5 and 6 present the measurements relating to the observation. The result of the transformation is a file containing an RDF graph as shown in Figure 9.

Figure 9. Excerpt of the generated RDF file.

The RDF graph generated can then be loaded into a Triple Store, so that data are made available on a linked data infrastructure and to allow SPARQL queries.

VII. CONCLUSION AND FUTURE WORK

The provision of open government data has been gaining momentum in many countries. The standardization and structure, however, are not yet a concern of the agencies responsible for disseminating them, complicating the analyzes on them, especially by machines.

In this work it was presented a process for publication of statistical data related to a multidimensional cube format. The use of a standard vocabulary representing the cube structure allows publication of a large amount of statistical data from different sources.

The choice of RDF Data Cube vocabulary as standard allows not only the publication scale as well as increases the potential for analysis, since the operations are also standardized. It also permits that the published data may be used by others applications.

The concern with the mapping of data ensures semantics and connect them with external sources and with other cubes that are already stored on Triple Store. In the case of cubes that share semantic concepts and have dimensions in common, multiple cubes are materialized.

Upcoming efforts that follow this work are the development of tools for visualizing multidimensional statistical data, using standard OLAP operations. These tools will provide a more powerful analysis of the data, besides allowing identifying links between different datasets.

Among the expected benefits of this proposal is the publishing of large-scale statistical series of linked data, which would serve as a base for open data portals, whether governmental or not.

The publication of the data in the form of OLAP cube allows the use of standard operations (e.g., slice, dice, roll-up, drill-down), which facilitates the analysis.

The published data, along with analysis tools enable a more agile application for data presentation. New mashups

can be created according to the needs, only requiring the specification of which cubes and which operations to be applied.

In further research, integration tests will be performed with tools that enable graphical visualization of data from the data cubes developed from the process presented in this work, culminating in the publication of results in a portal public domain.

REFERENCES

[1] B. Zapilko and B. Mathiak, "Performing Statistical Methods on Linked Data," in *Proc. Int'l Conf. on Dublin Core and Metadata Applications 2011*, 2011, pp. 116–125.

[2] U. Milošević, V. Janev, M. Spasić, J. Milojković, and S. Vranes, "Publishing Statistical Data As Linked Open Data," in *Proceedings of the 2nd International Conference on Information Society TechnologyInformation Society of the Republic of Serbia*, 2012, no. September, pp. 182–187.

[3] P. E. R. Salas, M. Martin, F. M. Da Mota, S. Auer, K. Breitman, and M. A. Casanova, "Publishing Statistical Data on the Web," in *International Conference on Semantic Computing*, 2012, 6th ed., pp. 285–292.

[4] B. Kämpgen, S. O'Rain, and A. Harth, "Interacting with Statistical Linked Data via OLAP Operations," in *Proceedings of the International Workshop on Interacting with Linked Data*, 2012, pp. 36–49.

[5] E. Prud'Hommeaux and A. Seaborne, "SPARQL Query Language for RDF," *SPARQL Query Language for RDF, Technical Report*, 2004.

[6] R. Hull, "Managing Semantic Heterogeneity in Databases : A Theoretical Perspective," in *Proc. ACMSymposium on Principles of Databases*, 1997, pp. 51–61.

[7] J. Umbrich, K. Hose, M. Karnstedt, A. Harth, and A. Polleres, "Linked data-the story so far," *World Wide Web*, vol. 14, no. 5–6, pp. 495–544, Jan. 2011.

[8] T. Berners-Lee, "Linked Data: Design Issues," 2006. [Online]. Available: http://www.w3.org/DesignIssues/LinkedData.html. [Accessed: Jul. 2012].

[9] F. Manola and E. Miller, "RDF Primer W3C Recommendation," 2004. [Online]. Available: http://www.w3.org/TR/rdf-primer/ . [Accessed: Aug. 2012].

[10] M. Niinimaki and T. Niemi, "An ETL Process for OLAP Using RDF / OWL Ontologies," in *Journal on Data Semantics XIII*, 2009, pp. 97–119.

[11] E. Prud'hommeaux and A. Seaborne, "SPARQL Query Language for RDF," 2008. [Online]. Available: http://www.w3.org/TR/rdf-sparql-query/. [Accessed: Jan. 2013].

[12] T. Bray, "RDF and Metadata," 1998. [Online]. Available: http://www.xml.com/pub/a/98/06/rdf.html. [Accessed: Jun. 2012].

[13] A. Kerzazi, O. Chniber, I. Navas-Delgado, and J. F. Aldana-Montes, "A Semantic Mediation Architecture for RDF Data Integration," in *SWAP 2008*, 2008.

[14] I. F. Cruz and H. Xiao, "The Role of Ontologies in Data Integration," *Journal of Engineering Intelligent Systems*, vol. 13, pp. 245–252, 2005.

[15] B. Kämpgen and A. Harth, "Transforming statistical linked data for use in OLAP systems," in *Proceedings of the 7th International Conference on Semantic Systems - I-Semantics '11*, 2011, pp. 33–40.

[16] J. M. Pérez, R. Berlanga, M. J. Aramburu, and T. B. Pedersen, "Integrating Data Warehouses with Web Data : A Survey," *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, no. 7, pp. 940–955, 2008.

[17] C. J. Date, *Introdução a sistemas de bancos de dados*. Rio de Janeiro: Campus, 2004, p. 865.

[18] R. Elmasri and S. B. Navathe, *Sistemas de banco de dados*, 4 ed. São Paulo: Addison Wesley, 2005, p. 724.

[19] S. Chaudhuri and U. Dayal, "An overview of data warehousing and OLAP technology," *ACM SIGMOD Record*, vol. 26, no. 1, pp. 65–74, Mar. 1997.

[20] B. Kämpgen, "DC proposal: online analytical processing of statistical linked data," *The Semantic Web–ISWC 2011*, vol. 7032, pp. 301–308, 2011.

[21] R. Cyganiak, C. Dollin, and D. Reynolds, "Expressing Statistical Data in RDF with SDMX-RDF," 2010. [Online]. Available: http://publishing-statistical-data.googlecode.com/svn/trunk/specs/src/main/html/index.html. [Accessed: Oct. 2012].

[22] C. Follenfant, D. Trastour, and O. Corby, "A Model for Assisting Business Users along Analytical Processes," *files.ifi.uzh.ch*, 2004.

[23] R. Cyganiak, D. Reynolds, and J. Tennison, "The RDF Data Cube vocabulary," 2010. [Online]. Available: http://publishing-statistical-data.googlecode.com/svn/trunk/specs/src/main/html/cube.html. [Accessed: Jul. 2012].

[24] M. A. Casanova, K. Breitman, P. Salas, D. Saraiva, V. Gama, J. Viterbo, R. Pires, E. Franzosi, and M. Chaves, "Open Government Data in Brazil," *IEEE Intelligent Systems*, no. May-June, 2012, pp. 45–49, 2012.

[25] G. Klyne and J. J. Carroll, "Resource Description Framework (RDF): Concepts and Abstract Syntax," 2004. [Online]. Available: http://www.w3.org/TR/rdf-concepts/. [Accessed: Aug. 2012].

[26] O. Lassila and R. R. Swick, "Resource Description Framework (RDF) Model and Syntax Specification," 1998. [Online]. Available: http://www.w3.org/1998/10/WD-rdf-syntax-19981008. [Accessed: Aug. 2012].

[27] D. Brickley and R. V. Guha, "RDF Vocabulary Description Language 1.0: RDF Schema," 2004. [Online]. Available: http://www.w3.org/TR/rdf-schema. [Accessed: Aug. 2012].

[28] T. Berners-Lee and D. Connolly, "Notation3 (N3): A readable RDF syntax," 2011. [Online]. Available: http://www.w3.org/TeamSubmission/n3/. [Accessed: Jan. 2013].

[29] R. Kimball, *The data warehouse toolkit: practical techniques for building dimensional data warehouses*, 1ª ed. 1996, p. 388.

[30] R. Kimball, L. Reeves, M. Ross, and W. Thornwaite, *The Data Warehouse Lifecycle Toolkit: Expert Methods for Designing*, 1ª ed. 1998.

[31] INEP, "ENEM 2012 - Passo a passo," 2013. [Online]. Available: http://www.enem.inep.gov.br. [Accessed: Jan. 2013].

[32] Virtuoso, "RDF Triple Store FAQ," 2012. [Online]. Available: http://virtuoso.openlinksw.com/dataspace/dav/wiki/Main/VOSRDFFAQ. [Accessed: Aug. 2012].

[33] J. Rusher and R. Networks, "Triple Store," 2001. [Online]. Available: http://www.w3.org/2001/sw/Europe/events/20031113-storage/positions/rusher.html. [Accessed: Aug. 2012].

[34] J. Pardillo, J.-N. Mazón, and J. Trujillo, "Bridging the semantic gap in OLAP models," in *Proceeding of the ACM 11th international workshop on Data warehousing and OLAP - DOLAP '08*, 2008, pp. 89–96.

[35] O. Romero and A. Abelló, "Automating multidimensional design from ontologies," in *Proceedings of the ACM tenth international workshop on Data warehousing and OLAP - DOLAP '07*, 2007, pp. 1–8.

[36] D. P. Ballou and H. L. Pazer, "Modeling Data and Process Quality in Multi-Input, Multi-Output Information Systems," *Management Science*, vol. 31, no. 2, pp. 150–162, Feb. 1985.

[37] Brasil, "Portal brasileiro de acesso a informação," 2013. [Online]. Available: http://data.gov.br. [Accessed: Jan. 2013].

[38] T. Gruber, "Towards principles for the design of ontologies used for knowledge sharing," *International Journal of Human-Computer Studies*, vol. 45, no. 5–6, pp. 907–928, 1995.