

6 Conclusion

6.1 Contributions

This dissertation proposed a mediator to consume linked data cubes stored in relational databases and described in a catalogue. The mediator is part of a broader architecture built to help describing and consuming statistical data, exposed as RDF triples, but stored in relational databases. The mediator exposes data cubes to the applications through a RESTful Web API.

To demonstrate the consumption of the data cubes in the mediator, this dissertation also presented two data flows: the request of data cube metadata and the request of data cube observations, using a small Brazilian dataset containing demographic data.

In general, the work contributes to increase the adoption of REST Principles in Linked Data applications, since both approaches have the notion of resource as a key concept, but focused on different aspects of the systems. The REST Principles help to easily access and manipulate resources.

This dissertation also aims at facilitating the understanding of how to represent data cubes and how to consume them by following the Linked Data Principles and using its recommended vocabularies. Furthermore, it also aims to promote the adoption of the R2RML mapping language, recently accepted as a W3C recommendation, as a standard to express the mappings from relational databases into RDF triples.

6.2 Limitations and Future work

One of the main drawbacks noticed in the preparation of the datasets (see Section 5.4.1) was the conversion of CSV files into star schemas. The manual process of normalizing the data was costly. In order to automate this process, the

Data Cube Discovery is currently being implemented to cover the identification of statistical databases in the database of origin (Ortiga 2013).

Another limitation of this work is related to the manual creation of the R2RML mapping files used to triplify the data cubes. To solve this problem, we need a tool that automatically generates customized R2RML mappings based on a set of semantic mappings that model the relationship between the relational database schema and the target ontology in RDF (Neto 2013). Furthermore, in the process of identifying the target ontology, the component *Data Source Recommendation* is currently being developed to identify RDF sources using *MapReduce* computing on top of the cloud computing paradigm (Talavera 2012).

We also intend to extend the architecture in such a way as to enable several catalogues to operate as a federation, supporting distributed search over the catalogues. Finally, another future improvement would be to allow operations over data cubes, as well as to allow keyword search over data cube metadata.