5
Conclusions and Future Work

This dissertation presented an approach for describing, extracting and publishing statistical data on the Web of Data. For this purpose, it described a mediation architecture comprising a Catalogue of Linked Data Cube Descriptions that stores RDF data cubes descriptions in a standardized way to help the data consumption. This approach provides a dynamic generation of RDF triples describing the data cube observations, which can be consumed by the user on demand. The dynamic generation of triples avoids storing the complete triplification of data cubes redundantly, which otherwise lead to problems such as storage space, velocity of consumption and mismatch of data.

The catalogue stores the descriptions of the data cubes (excluding their observations) and their mappings to star schema views of the relational databases. The descriptions of the data cubes use a vocabulary created for representing multi-dimensional spaces as the cube model, known as the RDF Data Cube Vocabulary. The mappings use R2RML, a W3C recommendation.

The catalogue also supports connections to other external databases, through the dimensions of the data cubes. The linkage between the dimensions and external data is essential to insert the data cubes in other contexts – following the Linked Data principles – thereby helping making the semantics of the data cubes explicit and enabling the discovery of related information by cross-checking between different data sources. Also, the catalogue enables the reuse of the dimensions that are common to the different data cubes.

The catalogue offers SPARQL queries as well as keyword queries to search the data cube descriptions it stores. It also features a module to populate the catalogue with new data cubes.

As for future work, it is intended to: create new functionalities that enable several catalogues to operate as a federation, supporting distributed search over several catalogues; create a tool that updates the linked data cube descriptions in the catalogue every time the RDBs suffer changes; and create a solution for the existence of different temporal versions for the same data cube.