The OBDA-based "Observatory of Research and Innovation" of the Tuscany Region

Alessandro Mosca, Bernardo Rondelli and Guillem Rull

SIRIS Lab, Research division of SIRIS Academic, Barcelona, Spain
{name.surname}@sirisacademic.com

Abstract. The Tuscany's Observatory of Research and Innovation portal is an instrument to promote more transparent and inclusive governance in the region. We show its interactive dashboard and underlying SPARQL endpoint, powered by SIRIS Academic's UNiCS platform, which integrates Open Data on the Higher Education & Research field, following the Ontology-Based Data Access approach.

Keywords. OBDA, Higher Education & Research, Data-driven policies, Interactive Data Visualisation

1. Introduction

In line with the orientation adopted by the EU in its Europe 2020 strategy, the Italian region of Tuscany has defined a set of policies aimed at supporting the Higher Education and Research (HE&R) system, and promoting innovation in the Tuscan territory. Tuscany has provided the Regional Research and Innovation Observatory as a tool to support the implementation of the Regional Development Program (PRS 20016/20[1]). The PRS is the cornerstone of regional policies: "[A] tool that expresses a vision for the future of Tuscany and proposes constructive dialogue with the actors of the territory" (E. Rossi, Region's President, PRS introductory speech). This tool has the ambition to communicate and enhance the strengths of the research system, and to host information on research, innovation and higher education.

Within the Observatory, and in support of the PRS, the Tuscany Region has decided to have an information dashboard capable of integrating HE&R data, keeping them up to date, and supporting policy makers in designing their policies. This paper shows the Observatory's interactive dashboard, currently located at `toscanaopenresearch.it`[2], and the underlying SPARQL endpoint, which is powered by SIRIS Academic's UNiCS platform [5], a system that integrates Open Data on the HE&R field and makes them accessible to users through a unified domain ontology, following the so-called Ontology-Based Data Access (OBDA) approach [6,3].

---

[1] `https://goo.gl/6VY4Co`
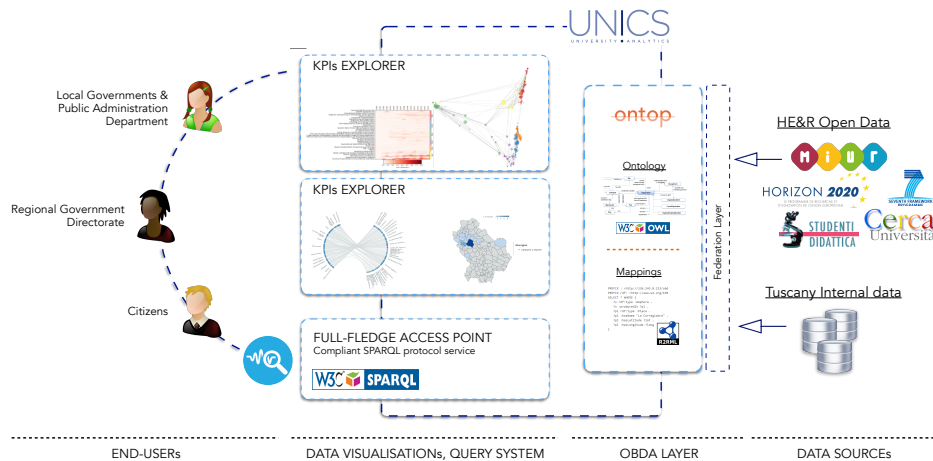[2] Guest users can login with username/password: dao2017/dao2017

Figure 1. UNiCS platform architecture tailored for the Tuscany's Observatory of Research and Innovation

The paper is structured as follows. First, Section 2 introduces the underlying platform that provides data to the Observatory, then Section 3 describes the Observatory's interactive dashboard.

2. The UNiCS Platform

University Analytics (UNiCS) integrates open data repositories about HE&R in Europe and makes them available via a dedicated SPARQL endpoint [9,8]. Queries are posed in terms of a domain ontology that provides a homogeneous view of the otherwise disparate integrated datasets. While data are originally stored in relational databases (DBs), UNiCS users see them as RDF data, the standard data model in Linked Data[3]. This is made possible by -ontop- [1], an OBDA system that allows querying relational DBs as virtual RDF graphs using SPARQL. Given the domain ontology, and an R2RML [7] mapping description that connects the ontology with the underlying DBs, -ontop- translates the users' SPARQL queries into SQL ones that are then run on the federated DB.

In the context of the Tuscany's Observatory, the core architecture of UNiCS includes a relational DB into which the different Italian and European open data repositories have been integrated, as per the Data Exchange [2] approach. The reason why these datasets have been copied into a single relational DB is that they are not available as proper, queryable DBs, but only as downloadable CSVs, so moving them into a more suitable storage is required. The datasets include official Italian student and researcher data[4] coming from the MIUR (Ministero dell'Istruzione, dell'Università e della Ricerca), and European data on FP7 and H2020 research projects [4]. An extension to this architecture is planned to incorporate internal data managed by the Tuscany Region, which will be federated with the UNiCS DB, and mapped into the UNiCS domain ontology. The architecture is depicted in Figure 1. (Being able to integrate not just relational sources but

---

[3]`https://www.w3.org/DesignIssues/LinkedData.html`

[4]Ministero dell'Istruzione, dell'Università e della Ricerca: (i) Anagrafe nazionale studenti: `anagrafe.miur.it`; (ii) Cerca università: `cercauniversita.cineca.it`
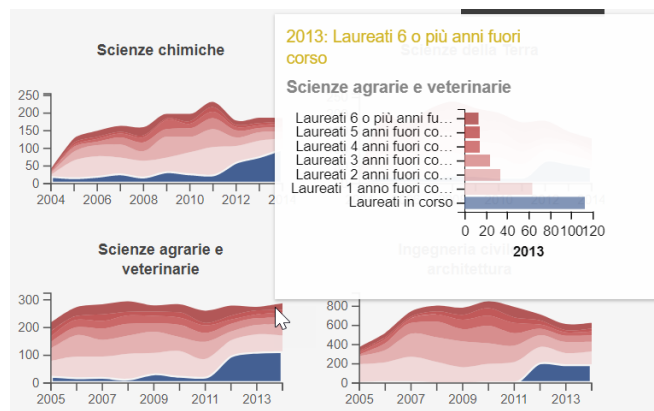
Figure 2. Interactive visualisation that shows a pop-up windows with additional information as the user mouses over.

also datasets that are behind given SPARQL endpoints is an -ontop- extension we are currently working on in collaboration with the KRDB Research Centre[5]).

3. Data visualisation

Currently, the Observatory consists mainly of an interactive dashboard, hosting data visualisations (co-designed with the relevant stakeholders) fed by the underlying UNiCS SPARQL endpoint. The visualisations are generated in real time by Javascript code running on the user's browser, which interrogates the endpoint for the necessary data. Visualisations are interactive: the user can click on different components to drill down on the results being displayed, as well as apply filters to focus on a particular data subset. Pop-up windows are also displayed with additional information that is not originally provided by the visual representation of the data, once the user selects a specific item in it. As an example, Figure 2 shows the distribution of graduated students per Italian bachelor faculty, and the different colours are meant to represent the overall number of years they spent at the university. Once the user mouses over the graphic of a given faculty, a window appears with a summarized view of the data behind, showing numbers that are not visible in the original visualisation. Users can either download the data behind each visualisation or copy and paste the queries which generate those data (see Figure 3), and execute them, possibly modified according to new specific needs.

The dashboard is now conceptually divided into four main sections: *Teachers & researchers* (showing distributions per gender, age, and disciplinary sectors), *Teaching* (on student data, with provenance, success rates per bachelor and master degrees), *Research at universities* (on EU funded projects, relative/absolute budgets, targeted to public organisations), and *Research at private companies* (on EU/regional funded projects, targeted to private organisations).

The Observatory's portal also includes a dedicated SPARQL endpoint and the LODE-powered documentation of the relative domain ontology[6]. The endpoint includes

---

[5]http://www.inf.unibz.it/krdb/, Free University of Bozen-Bolzano, Italy.
[6]http://34.250.237.252/toscana/sparql/docs/index.html

Figure 3. The dashboard allows users to see the SPARQL queries behind each visualisation, and also gives them the option to download the combined result of the queries in CSV format.

a library of pre-defined queries that either refer to the dashboard visualisations or have been collaboratively specified with the managers of the regions to satisfy specific needs and strategic demands. Users that are not familiar with SPARQL can then profit of the library, modify existing queries and execute them. The portal's visualisations and the SPARQL endpoint can be used by policy makers to *get a better understanding of the current situation, and both monitor the effectiveness of recent policies and also be able to design new policies based on evidence rather than intuition.*

References

[1] Calvanese, D., Cogrel, B., Komla-Ebri, S., Kontchakov, R., Lanti, D., Rezk, M., Rodriguez-Muro, M., Xiao, G.: Ontop: Answering SPARQL queries over relational databases. Semantic Web 8(3), 471–487 (2017)

[2] Fagin, R., Kolaitis, P.G., Miller, R.J., Popa, L.: Data exchange: semantics and query answering. Theor. Comput. Sci. 336(1), 89–124 (2005)

[3] Kontchakov, R., Lutz, C., Toman, D., Wolter, F., Zakharyaschev, M.: The combined approach to query answering in DL-Lite. pp. 247–257. KR'10, AAAI Press (2010)

[4] EU Publications Office: CORDIS. http://cordis.europa.eu/ Accessed 12 Sept 2017

[5] SIRIS Academic: UNiCS. http://university-analytics.com/ Accessed 12 Sept 2017

[6] Poggi, A., Lembo, D., Calvanese, D., De Giacomo, G., Lenzerini, M., Rosati, R.: Journal on data semantics. chap. Linking Data to Ontologies, pp. 133–173. Springer-Verlag, Berlin, Heidelberg (2008)

[7] W3C: R2RML: RDB to RDF Mapping Language. https://www.w3.org/TR/r2rml/ Accessed 12 Sept 2017

[8] W3C: SPARQL 1.1 Protocol. https://www.w3.org/TR/sparql11-protocol/ Accessed 12 Sept 2017

[9] W3C: SPARQL 1.1 Query Language. https://www.w3.org/TR/sparql11-query/ Accessed 12 Sept 2017