

UNiCS: The open data platform for Research and Innovation ^{*}

Xavi Gimenez, Alessandro Mosca, Fernando Roda,
Bernardo Rondelli, and Guillem Rull

SIRIS Lab, Research Division of SIRIS Academic, Barcelona, Spain
<http://www.sirisacademic.com/wb/siris-lab/>

Abstract. The paper introduces UNiCS, the open platform for Research and Innovation data management. The UNiCS platform follows the Ontology-Based Data Access (OBDA) approach, which eases the access to a vast amount of heterogeneous data and offers the final users the possibility to formulate queries using terms from the knowledge domain they are experts in. In UNiCS, each query gets transformed in a set of optimised queries to different data sources. Moreover, the OBDA approach makes the semantics of the data explicit, thus offering an intuitive way to access, explore, visualise, analyse, and post-process them.

Keywords: Research and Innovation data · Ontology-mediated data management · OBDA · Data Access · Data integration · Query answering.

1 Introduction

Research and Innovation (R&I) ecosystems involve data and knowledge flows across enterprises, academia, funding institutions, public authorities and citizens. The main problem here is that key R&I data elements are currently dispersed across a multitude of distinct, heterogeneous datasets, which are often neither in structured format nor systematically shared. This poses serious challenges to R&I decision and policy makers who are engaged in devising proper financial and political instruments to tune R&I dynamics, and in monitoring their impact in time. For them, in order to be driven by factual evidence, it becomes mandatory to overcome the limitations imposed by the usage of separated data silos, and to provide meaningful, integrated access to data with the appropriate granularity. In such a context, ontology-mediated data management (based on Semantic Web technologies) can help bringing together inputs and outcomes data from a variety of sources, in an (linked) open and interoperable fashion. The paper introduces the main characteristics of the UNiCS platform, an ontology-mediated data access and data integration platform for R&I policy and decision makers. UNiCS provides end-users with: (i) a running technology for accessing data in a way that is conceptually sound with their own domain knowledge; (ii) a semantically-transparent platform, ready to acquire and be complemented with new data from different sources; and (iii) a theoretically grounded mechanism to homogenise information stored in different formats and according to different conceptualisations.

^{*} Supported by SIRIS Academic SL.

2 The UNiCS platform

Since the mid 2000s, *ontology-mediated data management* has become a popular approach for providing integrated, uniform access to heterogeneous data sources. The UNiCS platform builds on top of this approach, its principles and methods: it hosts a core ‘ontology-based data access (OBDA)’ engine [1], which is fed with a R&I ontology and mappings for the sake of the data integration and the data access functionalities the platform offers. But, UNiCS is not only that: a number of dedicated data visualisations are implemented in its the front-end and directly connected to the OBDA engine, which is the place where the data are retrieved via suitable SPARQL queries. As one would expect, *the ontology, the mappings and the visualisations are fully customisable on a project-by-project basis, according to the specific information needs, as well as reporting, analytical and communication goals of the end-users.*

The overall architecture of the UNiCS system is shown in Fig.1, where as data sources we have considered the prototypical ones taken into account by R&I decision and policy makers: multidimensional data produced by governmental authorities, data on the HE&R sector, unstructured data coming from internal repositories, and proprietary data that, most of the time, are the result of ad-hoc analyses performed in house). In UNiCS, a conceptual layer is given in the form of an *ontology*, which captures knowledge about the R&I domain, and provides a high-level conceptual view of the underlying data sources in terms of a shared terminology. The UNiCS ontology is connected to the data sources through a declarative specification, given in terms of *mappings* that relate entities in the ontology (classes and properties) to (SQL) views over data: users can then query the data sources using the shared ontology terminology, without the need of understanding the precise structure of the sources, the relations between them, or the encoding of the data. Making use of the mappings, UNiCS translates the user queries into SQL queries formulated over the sources, while at the same time exploiting the domain knowledge encoded in the ontology to overcome incompleteness in the data and enrich query answers.

Ontology and the mappings, that are domain-dependent components of the platform, but the way UNiCS implements the OBDA principles is through -ontop- [6,5,4,3], a mature open-source system to query relational databases as Virtual RDF Graphs via *query-rewriting* [2]. -ontop- is responsible for the translation of the SPARQL user queries to SQL queries over the data sources, making an optimised use of the axioms in the ontology, the mappings and the statistics about the sources themselves. Its virtual approach avoids the cost of materialisation, and it allows one to profit from the maturity of relational database systems¹.

In its front-end, UNiCS comes equipped with interactive data visualisations and a full-fledge data access point based on SPARQL query answering. The UNiCS interactive visualisations are always designed together with the end-users of the platform by means of participatory and co-design methods. They offer an overview of the data plus tools for drilling down into the details or filtering them according to a selected set of

¹ -ontop- supports the virtual approach with all major RDBMSs (e.g., Oracle, IBM DB2, Microsoft SQL Server, PostgreSQL, and MySQL).

specific dimensions². Pop-up windows are also displayed with additional information that is not originally provided by the visual representation of the data. The users of UNiCS can always download the data behind each visualisation or copy and paste the SPARQL queries which generate each visualisation and execute them, possibly modified according to new specific needs.

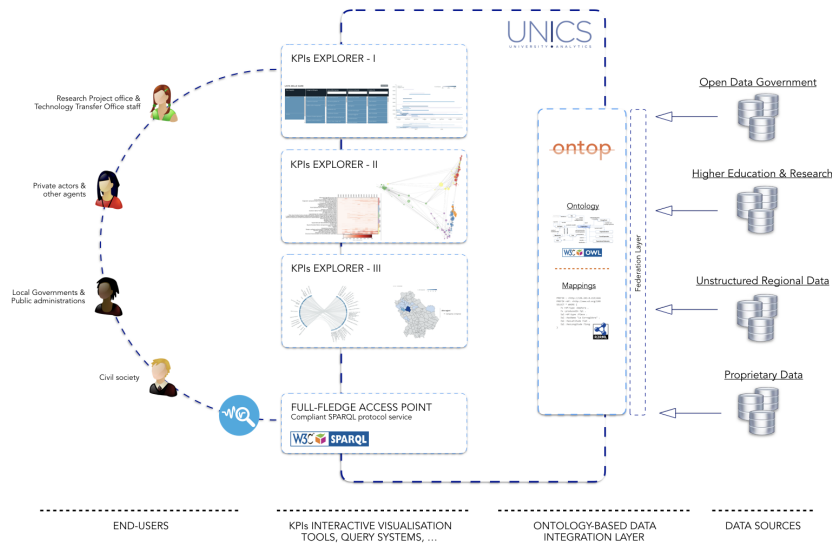


Fig. 1. UNiCS platform architecture

To conclude, the following is a list of selected applications that have been developed on top of the UNiCS platform during the last two years³: **UNiCS**⁴, the Open Data platform that integrates open data about Higher Education, Research and Innovation (HERI) in Europe. The platform repository is constantly updated by members of the SIRIS Lab, and it is fully accessible via its online SPARQL endpoint. **ToscanaOpen-Research**⁵ is the Toscana regional observatory for research and innovation: a web portal based on open and proprietary data, interactive visualisations and storytelling sections, whose main aim is to promote more transparent, evidence-based and inclusive governance of the HERI ecosystems of the region. **Smart Manufacturing Web**⁶ tool to allow for discovery of academic actors carrying out research on specific themes linked to Industry 4.0 in the region of Toscana (Italy). The web exploratory tool is the result of the combination of the application of topic modelling analytical algorithms over unstruc-

² As data are often multidimensional, alternative visualisations are provided using well known techniques like *linking and brushing* and *progressive (sequential) disclosure*.

³ Temporary user: `semantics2018`/password: `semantics2018`, whenever requested.

⁴ <http://unics.cloud>

⁵ <http://toscanaopenresearch.it> (in Italian)

⁶ <http://unics.cloud/toscana-smart-manufacturing>

tured data (mostly, scientific paper, projects and patent abstracts) and the data already integrated in the UNiCS platform. **RIS3-MCAT**⁷ is an Open Government, open source platform, with interactive visualisations summarising S3 (‘Smart Specialisation’) activities in the region of Catalonia (Spain). It covers: (i) Policy instruments and funding; (ii) Specialisation priorities and topics; (iii) Geographic and temporal distribution of activities. A fully searchable and filterable exploratory tool of the collaboration networks between Catalan actors, providing automatically computed indicators, is among the supported functionality. **Research Information System (RIS)** in University Paris Science et Lettres⁸, a key institution in the Parisian HERI landscape. The system has been developed for integrating distributed and heterogeneous data sources for (i) informing top-level strategic decision-making in the context of a period of radical change in the French HERI system, as well as (ii) open up the university to other quadruple helix actors by providing a detailed research portfolio, and (iii) generally increase the availability of pertinent data to mid-level management and individual researchers, fostering a culture change towards data use⁹.

References

1. Calvanese, D., De Giacomo, G., Lembo, D., Lenzerini, M., Poggi, A., Rodríguez-Muro, M., Rosati, R.: Ontologies and databases: The *DL-Lite* approach. In: Tessaris, S., Franconi, E. (eds.) Reasoning Web. Semantic Technologies for Informations Systems – 5th Int. Summer School Tutorial Lectures (RW), Lecture Notes in Computer Science, vol. 5689, pp. 255–356. Springer (2009)
2. Calvanese, D., De Giacomo, G., Lembo, D., Lenzerini, M., Rosati, R.: Tractable reasoning and efficient query answering in description logics: The *DL-Lite* family. *J. of Automated Reasoning* **39**(3), 385–429 (2007)
3. Calvanese, D., Giese, M., Hovland, D., Rezk, M.: Ontology-based integration of cross-linked datasets. In: Proc. of the 14th Int. Semantic Web Conf. (ISWC). Lecture Notes in Computer Science, vol. 9366, pp. 199–216. Springer (2015)
4. Giese, M., Soylu, A., Vega-Gorgojo, G., Waaler, A., Haase, P., Jiménez-Ruiz, E., Lanti, D., Rezk, M., Xiao, G., Özçep, Ö.L., Rosati, R.: Optique: Zooming in on Big Data. *IEEE Computer* **48**(3), 60–67 (2015)
5. Mosca, A., Rondelli, B., Rull, G.: The OBDA-based ”observatory of research and innovation” of the tuscan region. In: Proc. of the Joint Ontology Workshops 2017 Episode 3: The Tyrolean Autumn of Ontology, Bozen-Bolzano, Italy, September 21-23, 2017. CEUR Workshop Proceedings, vol. 2050. CEUR-WS.org (2017)
6. Rodriguez-Muro, M., Kontchakov, R., Zakharyashev, M.: Ontology-based data access: On-top of databases. In: Proc. of the 12th Int. Semantic Web Conf. (ISWC). vol. 8218, pp. 558–573. Springer (2013)

⁷ <http://unics.cloud/ris3/mcat> (in Catalan).

⁸ <http://unics.cloud/psl>

⁹ All the applications above provide open access to the underlying open datasets according to the well known ‘5* level’ of the Open Data vision.