Novel Developments in Ontology-Based Data Access and Integration: Part 2. Systems and Use Cases

Diego Calvanese, **Guohui Xiao**

KRDB Research Centre for Knowledge and Data
Free University of Bozen-Bolzano, Italy

17th International Conference of the Italian Association for Artificial Intelligence (AI*IA 2018)
Trento, Italy, 20 November 2018

OBDA is by now a mature technology

Ontology-based querying of relational data sources is supported by several systems, both open-source and commercial:

- Mastro [C., De Giacomo, et al. 2011] [1]
  Sapienza Università di Roma & OBDA systems SRL, Italy
- Morph [Priyatna, Corcho, and Sequeda 2014] [2]
  Technical University of Madrid, Spain
- Ontop [C., Cogrel, et al. 2017] [3]
  Free University of Bozen-Bolzano, Italy
- Stardog [4], Stardog Union, US
- Ultrawrap [Sequeda and Miranker 2013] [5], Capsenta, US
- Oracle Spatial and Graph [6]

[1] http://www.obdasystems.com/it/mastro
[2] https://github.com/oeg-upm/morph-rdb
[3] http://ontop.inf.unibz.it
[4] http://www.stardog.com
[5] https://capsenta.com/ultrawrap
[6] http://www.oracle.com/technetwork/database/options/spatialandgraph

unibz

http://ontop.inf.unibz.it/

- State-of-the-art OBDA system developed at the Free University of Bozen-Bolzano.
- Compliant with all relevant Semantic Web standards:
  RDF, RDFS, OWL 2 QL, R2RML, and SPARQL
- Supports all major relational DBs:
  Oracle, DB2, MS SQL Server, Postgres, MySQL, Denodo, Teiid, Exareme, etc.
- Open-source and released under Apache 2 license.
- Development of *Ontop*:
  - development started in 2009
  - already well established and widely adopted:
    +200 members in the mailing list
    +9000 downloads in last 2 years
  - main development was carried out in the context of the EU project Optique

unibz

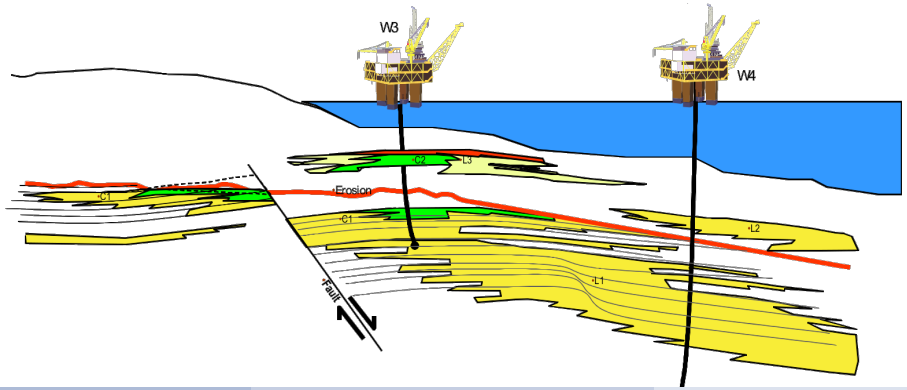
Selected Use Cases of OBDA

- Oil & Gas: Statoil [Kharlamov et al. 2017a]
- Turbine Diagnoses: Siemens [Kharlamov et al. 2017b]
- Development of data integration solutions: SIRIS Academic SL Barcellona [Mosca, Rondelli, and Rull 2017]
- Log Extraction in Process Mining – Euregio financed project KAOS [Calvanese, Kalayci, et al. 2017]
- Cultural heritage: EPNet project [Calvanese, Liuzzo, et al. 2016]
- Maritime security: EMSec project [Brüggemann et al. 2016]
- Manufacturing [Petersen et al. 2017]
- Health care: electronic health records [Rahimi et al. 2014]
- Public debt: the Italian Ministry of Economy and Finance [Antonioli et al. 2014]
- Smart cities: IBM Ireland [López et al. 2015]
- ...

unibz

Statoil

- Optique use case partner
- Exploration domain: analyse existing relevant data in order to find exploitable accumulations of oil or gas
- Improve the efficiency of the information gathering routine for geologists
- Efficient, creative data collection from multiple large volume data sources

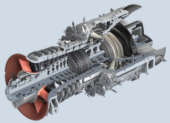


unibz

Siemens use case

- Optique use case partner
- Siemens energy department
- streaming and temporal data

- Siemens
  - produces huge appliances, e.g., turbines
  - installs them in plants

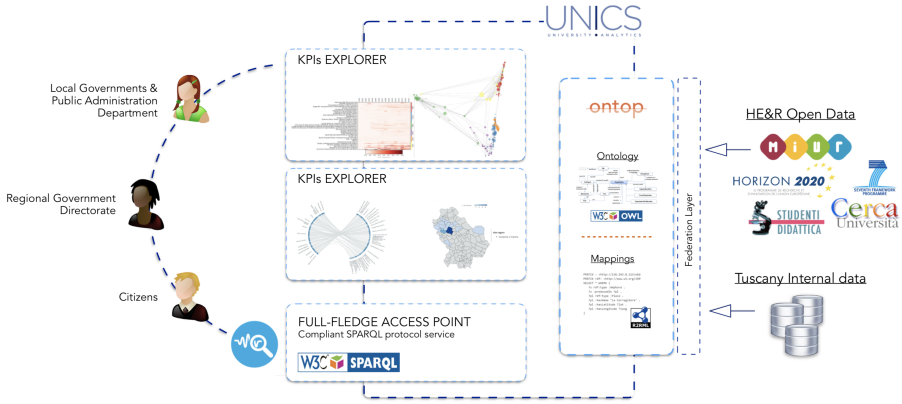


- Siemens service centers
  - offers constant monitoring and diagnostics service
  - over 50 service centers world wide
  - each SC is responsible for several thousands appliances
  - their job: monitoring and diagnostic of turbine
- Monitoring and diagnostic tasks
  - reactive and preventive diagnostics
    - offline, after an issue is detected
  - predictive analyses
    - real-time, to avoid issues

unibz

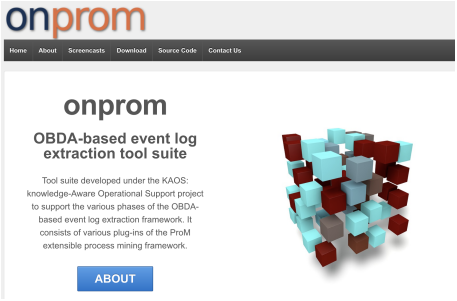
Tuscany's Observatory of Research and Innovation

- Carried out by SIRIS Academic SL Barcellona and the Tuscany Region
- Support the Higher Education and Research system, and promoting innovation in the Tuscan territory [Mosca, Rondelli, and Rull 2017]



Log Extraction in Processing Mining

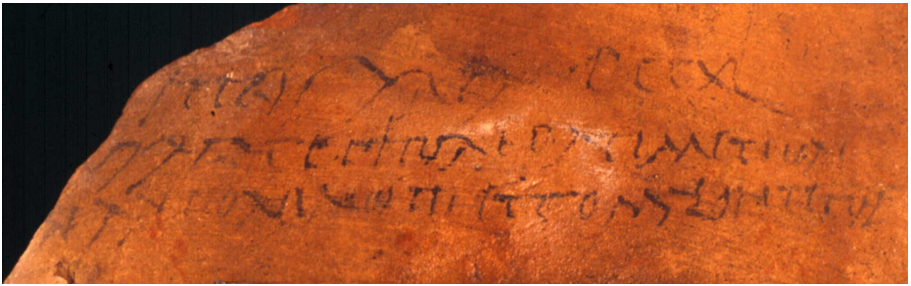
- Process Mining as a way to reconcile model-driven management and the real behaviours
- Processing mining tools (e.g. Disco [fluxicon.com]) require data in a particular format (XES)
- Data preparation is an issue in presence of legacy data
- onprom[7] is an effective tool chain for extracting event logs from legacy databases



[7] http://onprom.inf.unibz.it/

EPNet project

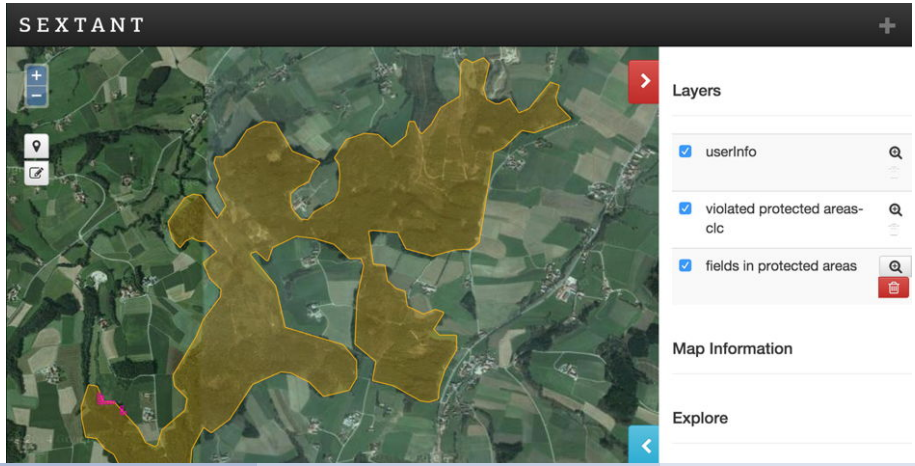
- Ontology-based data integration for humanities and archaeologists
- ERC advanced grant EPNet "Production and distribution of food during the Roman Empire: Economics and Political Dynamics".
- Linking three datasets:
  1. the EPNet relational repository
  2. the Epigraphic Database Heidelberg
  3. the Pleiades dataset



unibz

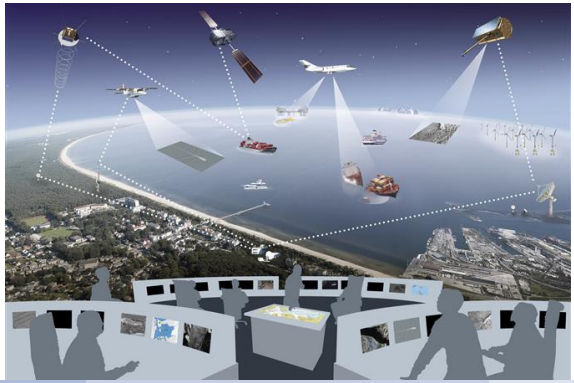
Melodies project

- EU FP7 Melodies project: working with Open Data, 16 partners.
- Geospatial extension Ontop-spatial used for accessing geospatial data
- Use cases: urban development, land management, disaster management



EMSec project

- German BMBF project EMSec, collaborated with Airbus:
- real-time services for maritime security
- Geo-spatial support by *Ontop*-spatial (developed as a fork of *Ontop*)
- SPARQL federation to access different kinds of data sources:
  - SPARQL endpoints of *Ontop* over *in situ* data
  - open SPARQL endpoints: Geonames, DBPedia



unibz

Hands-on with Ontop

- Scenario: Integration of the information systems of two universities
- Requirements
  - Java 8
  - H2 with preloaded datasets
  - Protégé bundle with Ontop for editing ontologies and mappings
  - Tomcat bundle with Ontop for deploying a SPARQL endpoint
- Programme
  - Mapping the first data source
  - Mapping the second data source
  - Deploying a SPARQL endpoint
- Instruction: Session 1 of the Ontop Tutorial:
  `https://github.com/ontop/ontop-examples/edit/master/ontop-v3-tutorial/`

unibz

Tips

- Protégé does not work with newer versions (>8) of Java
- Reasoner needs to be restarted whenever ontology or mapping are changed. Reasoner "stop" and "start" is more reliable than "synchronize".
- The SPARQL endpoint part can be done independently

unibz

References

References I

[1] Diego C., Giuseppe De Giacomo, Domenico Lembo, Maurizio Lenzerini, Antonella Poggi, Mariano Rodriguez-Muro, Riccardo Rosati, Marco Ruzzi, and Domenico Fabio Savo. "The Mastro System for Ontology-Based Data Access". In: *Semantic Web J.* 2.1 (2011), pp. 43–53.

[2] Freddy Priyatna, Oscar Corcho, and Juan F. Sequeda. "Formalisation and Experiences of R2RML-based SPARQL to SQL Query Translation Using morph". In: *Proc. of the 23rd Int. World Wide Web Conf. (WWW)*. 2014, pp. 479–490. DOI: 10.1145/2566486.2567981.

[3] Diego C., Benjamin Cogrel, Sarah Komla-Ebri, Roman Kontchakov, Davide Lanti, Martin Rezk, Mariano Rodriguez-Muro, and Guohui Xiao. "Ontop: Answering SPARQL Queries over Relational Databases". In: *Semantic Web J.* 8.3 (2017), pp. 471–487. DOI: 10.3233/SW-160217.

[4] Juan F. Sequeda and Daniel P. Miranker. "Ultrawrap: SPARQL Execution on Relational Data". In: *J. of Web Semantics* 22 (2013), pp. 19–39.

[5] E. Kharlamov et al. "Ontology Based Data Access in Statoil". In: *J. of Web Semantics* 44 (2017), pp. 3–36. DOI: 10.1016/j.websem.2017.05.005.

unibz

References

References II

[6] E. Kharlamov et al. "Semantic Access to Streaming and Static Data at Siemens". In: *J. of Web Semantics* 44 (2017), pp. 54–74.

[7] Alessandro Mosca, Bernardo Rondelli, and Guillem Rull. "The OBDA-Based "Observatory of Research and Innovation" of the Tuscany Region". In: *JOWO*. Vol. 2050. CEUR Workshop Proceedings. CEUR-WS.org, 2017.

[8] Diego Calvanese, Tahir Emre Kalayci, Marco Montali, and Stefano Tinella. "Ontology-Based Data Access for Extracting Event Logs from Legacy Data: The onprom Tool and Methodology". In: *BIS*. Vol. 288. Lecture Notes in Business Information Processing. Springer, 2017, pp. 220–236.

[9] Diego Calvanese, Pietro Liuzzo, Alessandro Mosca, José Remesal, Martin Rezk, and Guillem Rull. "Ontology-based data integration in EPNet: Production and distribution of food during the Roman Empire". In: *Eng. Appl. of AI* 51 (2016), pp. 212–229.

[10] Stefan Brüggemann, Konstantina Bereta, Guohui Xiao, and Manolis Koubarakis. "Ontology-based data access for Maritime Security". In: *Proc. of ESWC*. 2016.

unibz

References III

[11] Niklas Petersen, Lavdim Halilaj, Irlán Grangel-González, Steffen Lohmann, Christoph Lange, and Sören Auer. "Realizing an RDF-Based Information Model for a Manufacturing Company - A Case Study". In: *International Semantic Web Conference (2)*. Vol. 10588. Lecture Notes in Computer Science. Springer, 2017, pp. 350–366.

[12] Alireza Rahimi, Siaw-Teng Liaw, Jane Taggart, Pradeep Ray, and Hairong Yu. "Validating an ontology-based algorithm to identify patients with Type 2 Diabetes Mellitus in Electronic Health Records". In: *Int. J. of Medical Informatics* 83.10 (2014), pp. 768–778.

[13] N. Antonioli et al. "Ontology-based Data Management for the Italian Public Debt". In: *Proc. of FOIS*. Vol. 267. FAIA. 2014.

[14] V. López, M. Stephenson, S. Kotoulas, and P. Tommasi. "Data Access Linking and Integration with DALI: Building a Safety Net for an Ocean of City Data". In: *Proc. of the 14th Int. Semantic Web Conf. (ISWC)*. Vol. 9367. Lecture Notes in Computer Science. Springer, 2015, pp. 186–202.

unibz