UNDERSTANDING ORGANISATIONS,
HELPING PEOPLE

UNiCS: An ontology-mediated open data platform for Research and Innovation

Alessandro Mosca
SIRIS LAB, Research Division of SIRIS Academic

UNICS
UNIVERSITY • ANALYTICS

JIRC @Université de François Rabelais, November 29th 2018

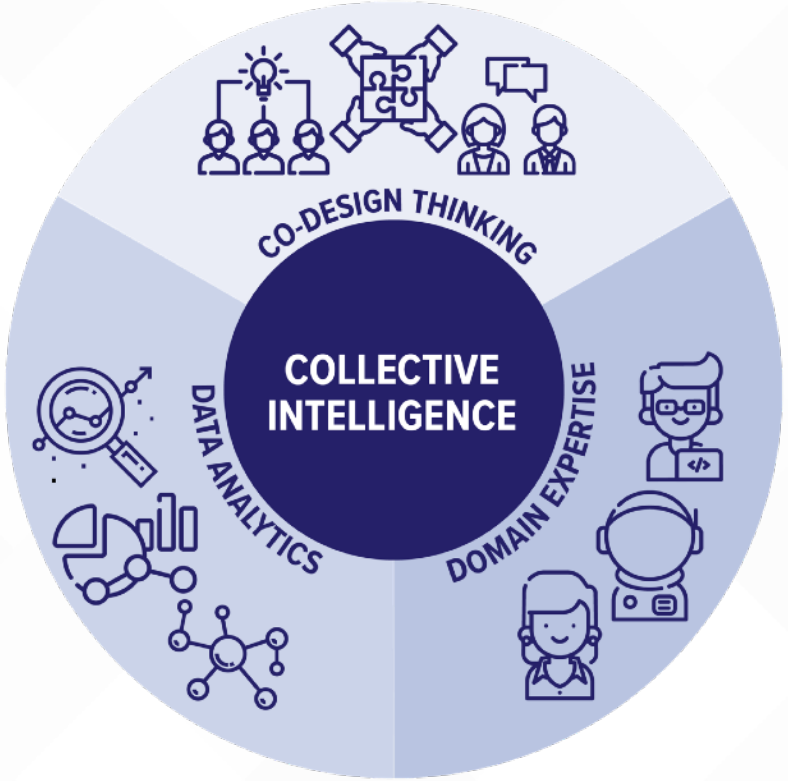
SIRIS
A·CA·DE·MI·C

Today, the SIRIS team

international and interdisciplinary, with seven different nationalities and PhDs in different fields, from archeology to astrophysics, philosophy and computer science.

We combine **domain expertise**, **data analytics** and **co-design thinking**.

We stimulate **collective intelligence** for exploring scenarios, designing solutions and launching transformations.



The logo-slide… a must

Regional Governments

"I want to understand how the regional S3-related initiatives fit within the general R&I panorama and other funded projects. I want to analyse the collaboration networks which exist already within a given thematic area, and understand which kinds of actors are mostly involved."

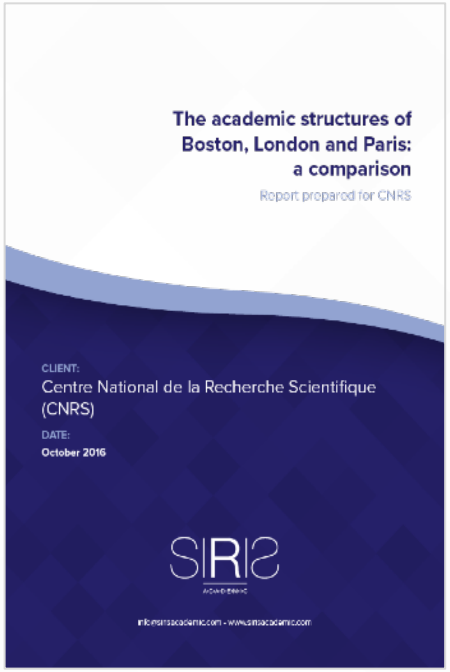
HE&R Institutions

"I'm interested in enhancing the attractiveness of my higher education organisation and I want to get a better understanding of the student mobility patterns and their evolution over time. Is there anyone else similar to us on this respect?"

R&I Funding Bodies

"Research and innovation ecosystems are fast changing and fast growing. I need to define our research and innovation priorities, but also inform the shaping our future investments and then monitor and learn to eventually reshape our existing policies."

Strategic and analytical reports to inform decisions



Goals, Strategies, Instruments… Data!

Regional Governments

"I want to understand how the regional **S3-related initiatives** fit within the general R&I panorama and other **funded projects**. I want to analyse the collaboration networks which exist already within a given **thematic area**, and understand which kinds of **actors** are mostly involved."

HE&R Institutions

"I'm interested in enhancing the attractiveness of my higher education organisation and I want to get a better understanding of the **student mobility patterns** and their evolution over time. Is there **anyone else similar to us** on this respect?"

R&I Funding Bodies

"**Research and innovation ecosystems** are fast changing and fast growing. I need to define our research and innovation **priorities**, but also inform the shaping our future **investments** and then monitor and learn (via specific **indicators**) to eventually reshape our existing policies."

Our technological requirements

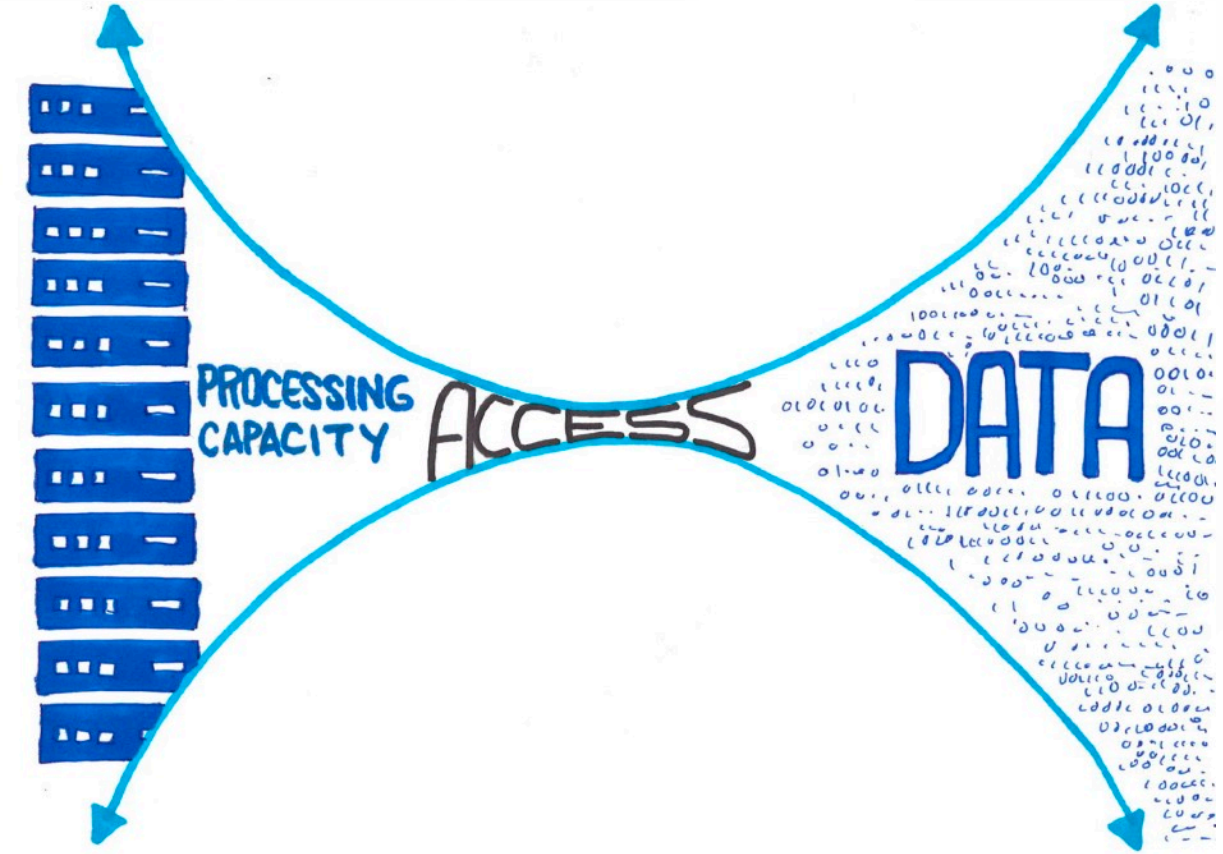
Our solutions provide end-users with:

- a running technology for **integrating and accessing data** in a way that is conceptually sound with the **experts domain knowledge**

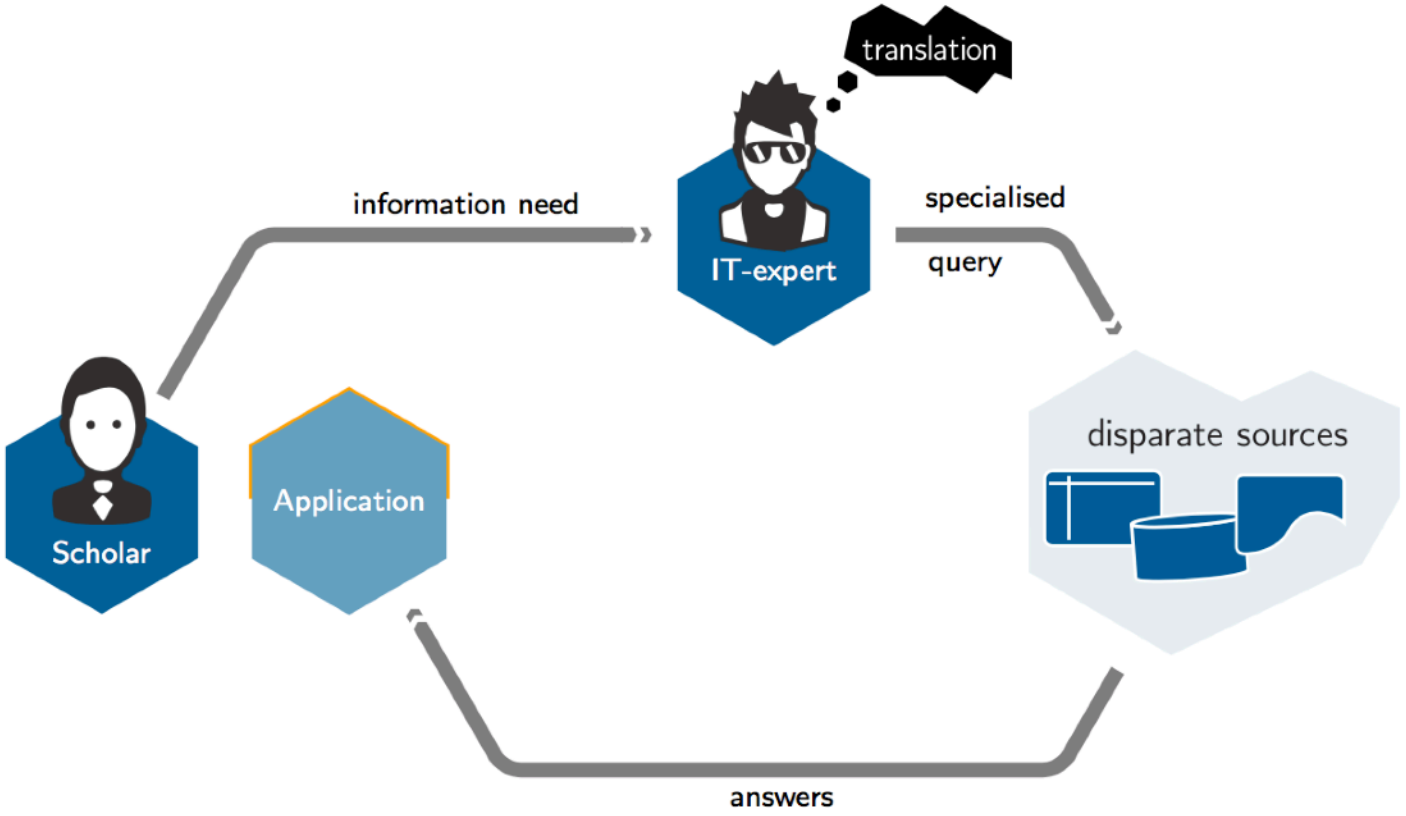
- a **semantically-transparent platform**, ready to acquire and be complemented with new **data from different sources**

- a theoretically grounded mechanism to **homogenise information** originally stored in **different formats** and according to **different conceptualisations**

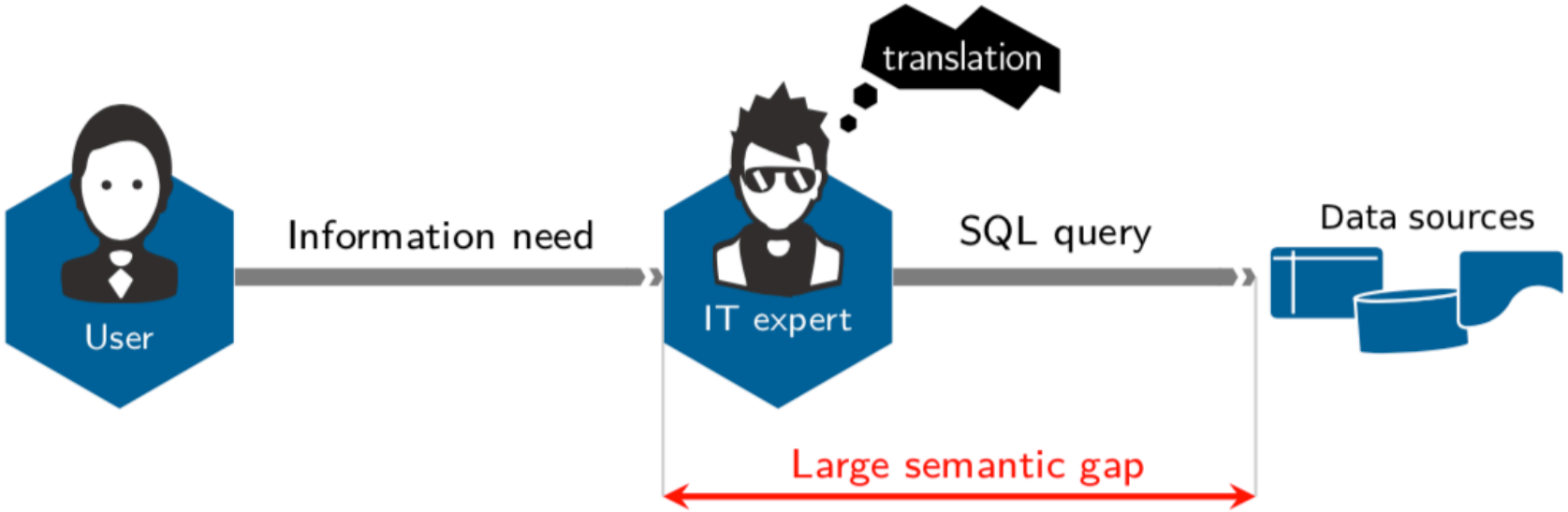
Data access bottleneck



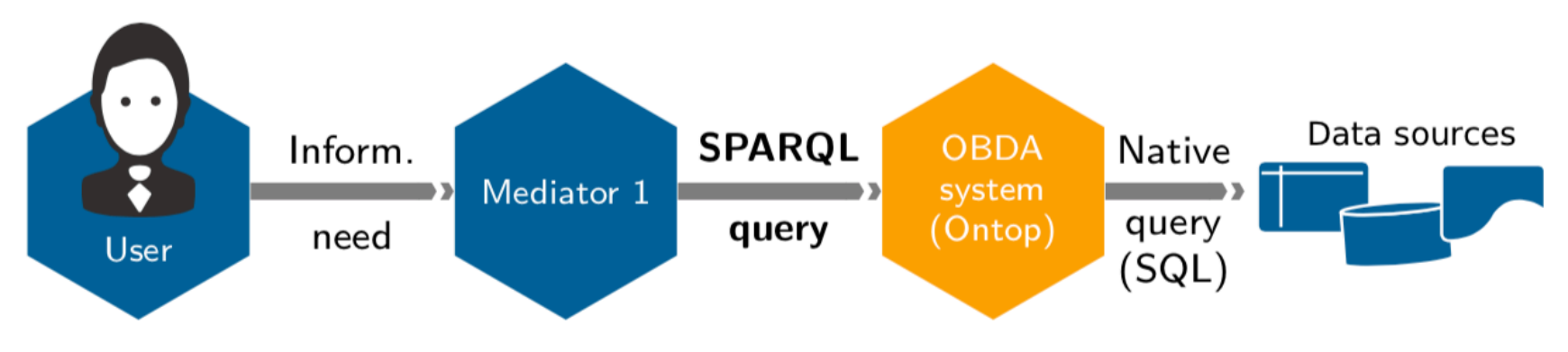
How is this problem usually addressed?



Issue: the “Semantic Gap”



Ontology-based Data Access (OBDA)



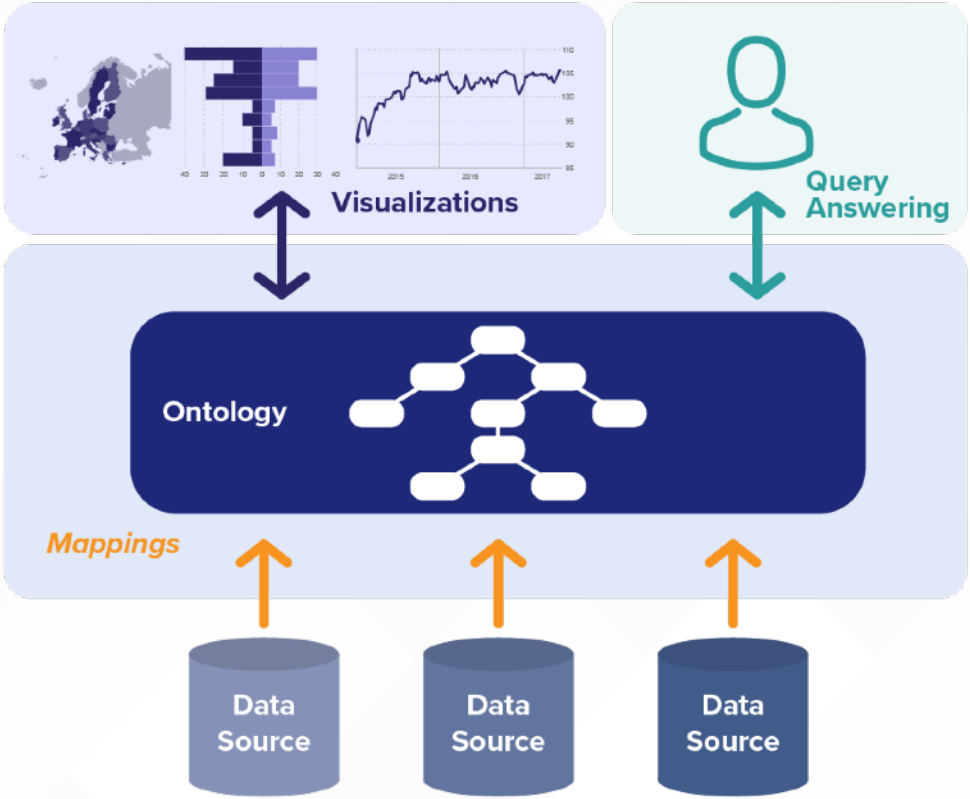
The Ontology-based Data Access (OBDA) architecture

Exploits the domain knowledge to support the high-level translation of the information needs, and to enrich the extracted answers.

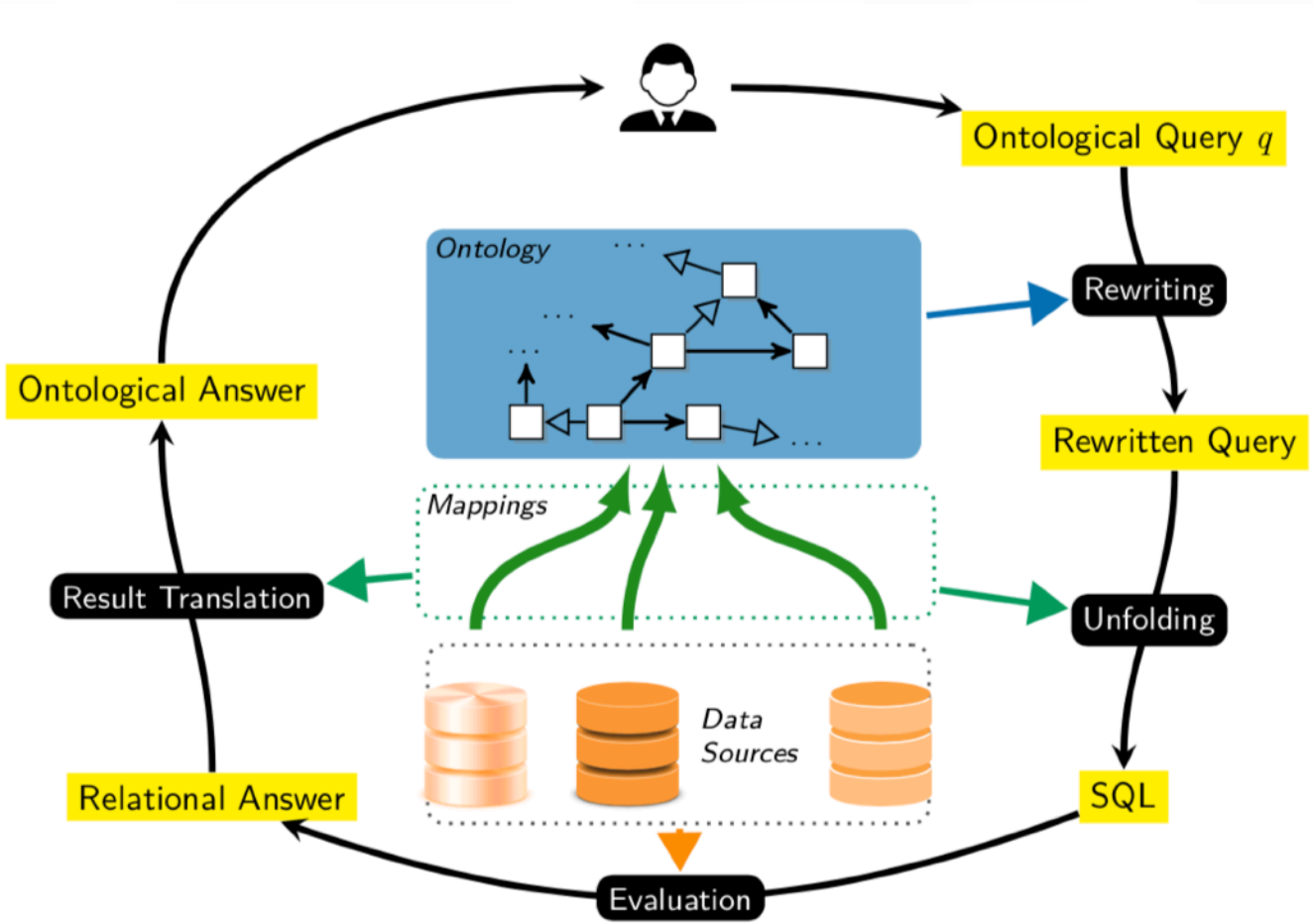
Accepts as input intuitive queries posed by the users on the basis of domain-centred vocabulary.

Combines and filters the extracted data to provide to the users the answers to their requests.

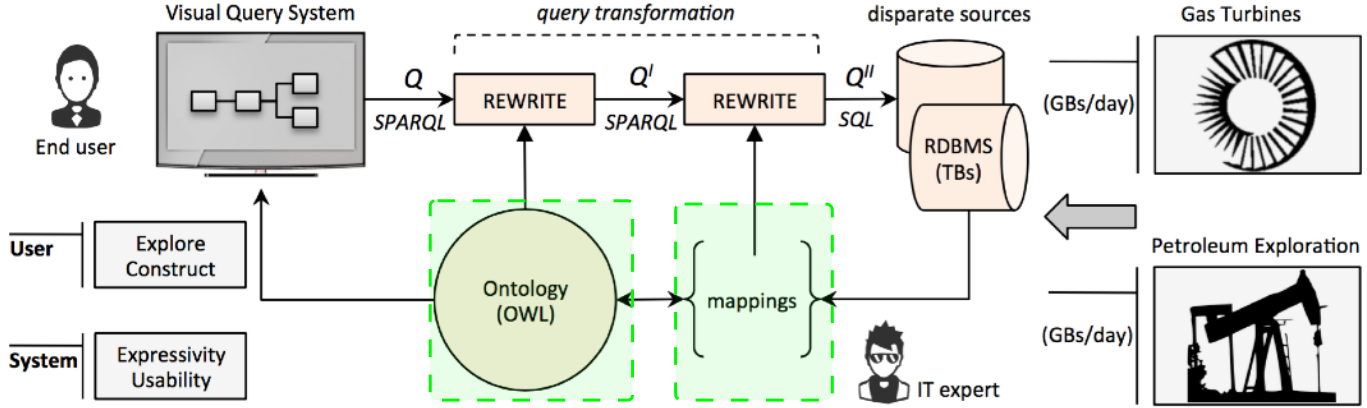
Deals with the variety and complexity of the data sources, extracting from them the necessary data.



An OBDA cycle - I



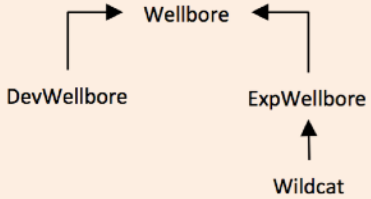
An OBDA cycle - II



Database

wlb_dev(name, . . .)
wlb_exp(name, purpose, . . .)

Ontology



Mappings

DevWellbore(name) → SELECT name FROM wlb_dev

ExpWellbore(name) → SELECT name FROM wlb_exp

Wildcat(name) → SELECT name FROM wlb_exp
WHERE purpose = 'WILDCAT'

Q *original user query*	***Q^I*** *rewrite w.r.t. ontology*	***Q^II*** *rewrite to SQL*	***Q^III*** *optimise*
Wellbore(x) →	Wellbore(x) U DevWellbore(x) U ExpWellbore(x) U Wildcat(x) →	SELECT name FROM wlb_dev UNION SELECT name FROM wlb_exp UNION SELECT name FROM wlb_exp WHERE purpose ='WILDCAT' →	SELECT name FROM wlb_dev UNION SELECT name FROM wlb_exp

Optique

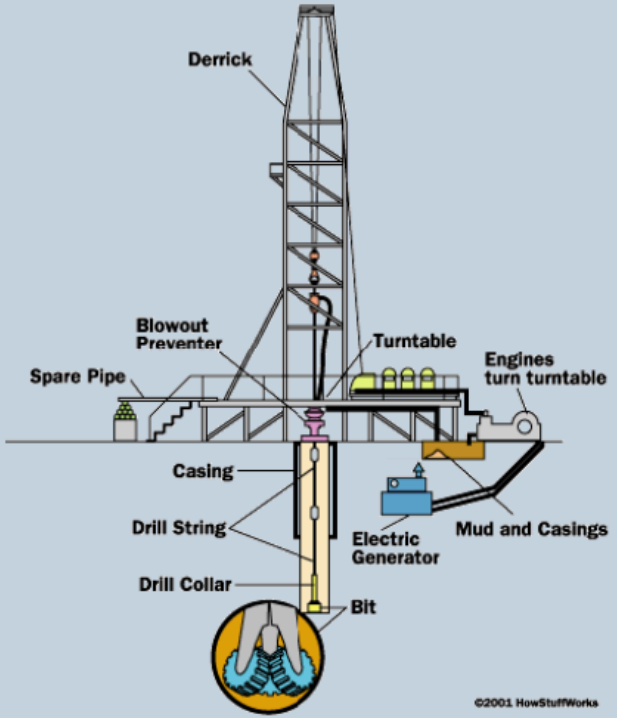
A conceptual model of (some aspect of) the world

- Introduces **vocabulary** relevant to domain
- Specifies **meaning** (semantics) of terms

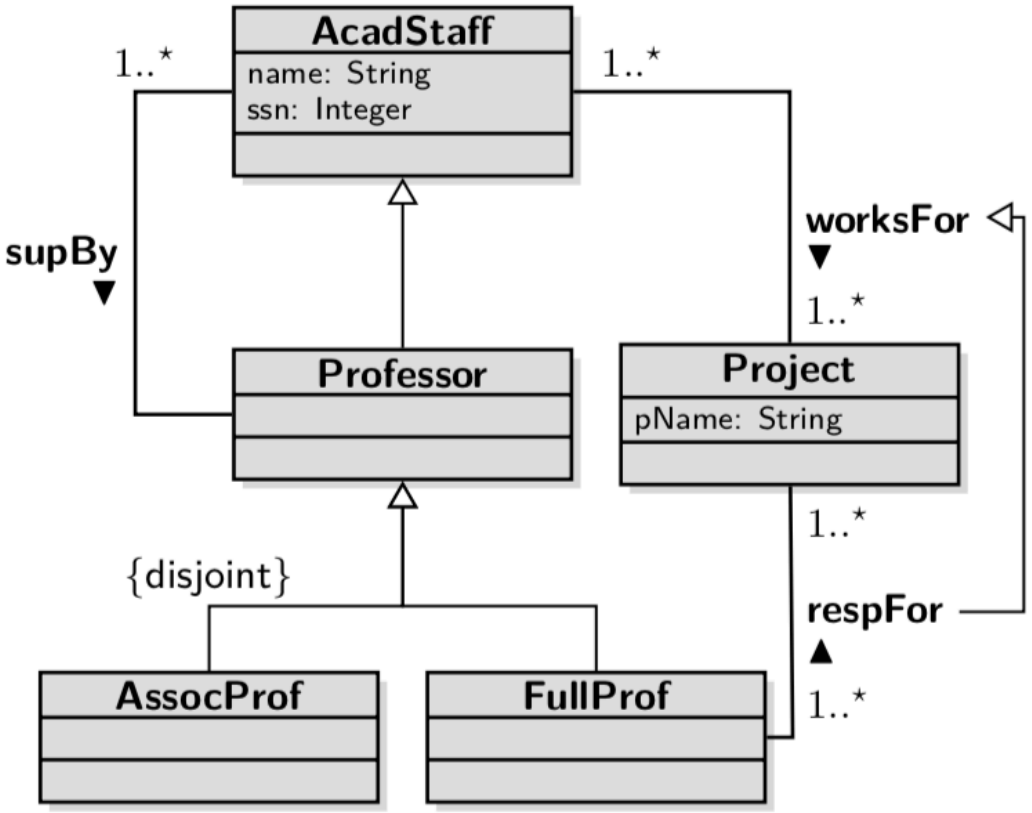
  Oil pipeline is a pipeline from a facility that is an oil facility

- **Formalised** using suitable logic

$$\mathsf{OilPipeline} \sqsubseteq \mathsf{Pipeline} \sqcap \exists \mathsf{fromFacility}.\mathsf{OilFacility}$$



Ontology



$$
\begin{aligned}
\text{Professor} &\sqsubseteq \text{AcadStaff} \\
\text{AssocProf} &\sqsubseteq \text{Professor} \\
\text{FullProf} &\sqsubseteq \text{Professor} \\
\text{AssocProf} &\sqsubseteq \neg\text{FullProf} \\
\text{AcadStaff} &\sqsubseteq \exists \text{ssn} \\
\exists \text{ssn} &\sqsubseteq \text{AcadStaff} \\
\exists \text{ssn}^{-} &\sqsubseteq \texttt{Integer} \\
\exists \text{worksFor} &\sqsubseteq \text{AcadStaff} \\
\exists \text{worksFor}^{-} &\sqsubseteq \text{Project} \\
\text{AcadStaff} &\sqsubseteq \exists \text{worksFor} \\
\text{Project} &\sqsubseteq \exists \text{worksFor}^{-} \\
\text{respFor} &\sqsubseteq \text{worksFor} \\
&\vdots
\end{aligned}
$$

The OWL 2 QL profile

- **OWL 2 QL** is one of the three standard profiles of OWL 2.

- Derived from the ***DL-Lite*$_R$** Description Logic of the *DL-Lite-family*:
  - Groups the domain into **classes** of objects with common properties.
  - Binary relations between objects (**object properties**).
  - Binary relations from objects to values (**data properties**).

- Is considered a lightweight ontology language:
  - controlled expressive power
  - efficient inference

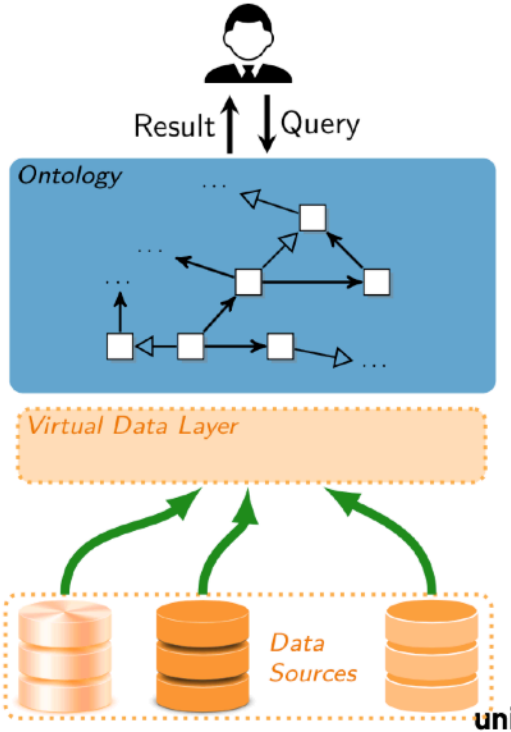
- Optimised for accessing large amounts of data (i.e., for data complexity):
  - **First-order rewritability of query answering**: queries over the ontology can be rewritten into SQL queries over the underlying relational database.
  - Consistency checking is also first-order rewritable.

In an OBDA instance $\mathcal{O} = \langle\langle \mathcal{T}, \mathcal{M}, \mathcal{S}\rangle, \mathcal{D}\rangle$, the **mapping** $\mathcal{M}$ encodes how the data $\mathcal{D}$ in the source(s) $\mathcal{S}$ should be used to populate the elements of $\mathcal{T}$.

Virtual data layer

The data $\mathcal{D}$ and the mapping $\mathcal{M}$ define a virtual data layer $\mathcal{V} = \mathcal{M}(\mathcal{D})$

- Queries are answered w.r.t. $\mathcal{T}$ and $\mathcal{V}$.
- We do not really materialize the data of $\mathcal{V}$ (it is virtual!).
- Instead, the intensional information in $\mathcal{T}$ and $\mathcal{M}$ is used to translate queries over $\mathcal{T}$ into queries formulated over $\mathcal{S}$.



Mapping assertions – RDB-RDF

Class instance (`:Student`)

Source	$q(s) \leftarrow \texttt{uni1-student}(s, f, l)$
	
	`SELECT s_id`
	`FROM uni1.student`

Target	$\text{Student}(\text{URI}_1(s))$ `ex:uni1/student/{s_id} a :Student .`

Data property (`foaf:firstName`)

Source (SQL)	`SELECT s_id, firstName, lastName` `FROM uni1.student`

Target (RDF)	`ex:uni1/student/{s_id}`
	`foaf:firstName "{firstName}"^^xsd:string ;`
	`foaf:lastName "{lastName}"^^xsd:string .`

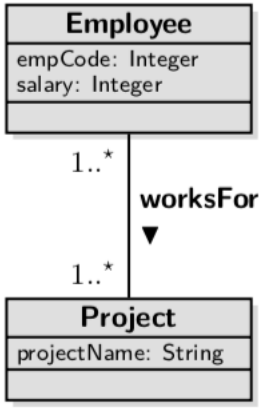
Object property (`:teaches`)

Source	`SELECT *` `FROM "uni1"."teaching"`
Target (RDF)	`ex:uni1/academic/{a_id} :teaches` `ex:uni1/course/{c_id} .`

Magic number

Source	`SELECT * FROM "uni1"."academic"` `WHERE "position" = 1`
Target (RDF)	`ex:uni1/academic/{a_id} a :FullProf .`

Mapping assertions – RDB-RDF



D_1:

SSN	*PrName*
23AB	optique
...	...

D_2:

Code	*Salary*
e23	1500
...	...

D_3:

Code	*SSN*
e23	23AB
...	...

m_1:
```
SELECT SSN, PrName
FROM D1
```
$\rightsquigarrow$ Employee(**pers**(*SSN*)),
Project(**proj**(*PrName*)),
projectName(**proj**(*PrName*), *PrName*),
worksFor(**pers**(*SSN*), **proj**(*PrName*))

m_2:
```
SELECT SSN, Salary
FROM D2, D3
WHERE D2.Code = D3.Code
```
$\rightsquigarrow$ Employee(**pers**(*SSN*)),
salary(**pers**(*SSN*), *Salary*)

Resource Description Framework (RDF), data

Class membership:

Fact	Prof(*uni2/person/1*)
RDF triple	`<uni2/person/1> a :Prof`

Note: This is an abbreviation for

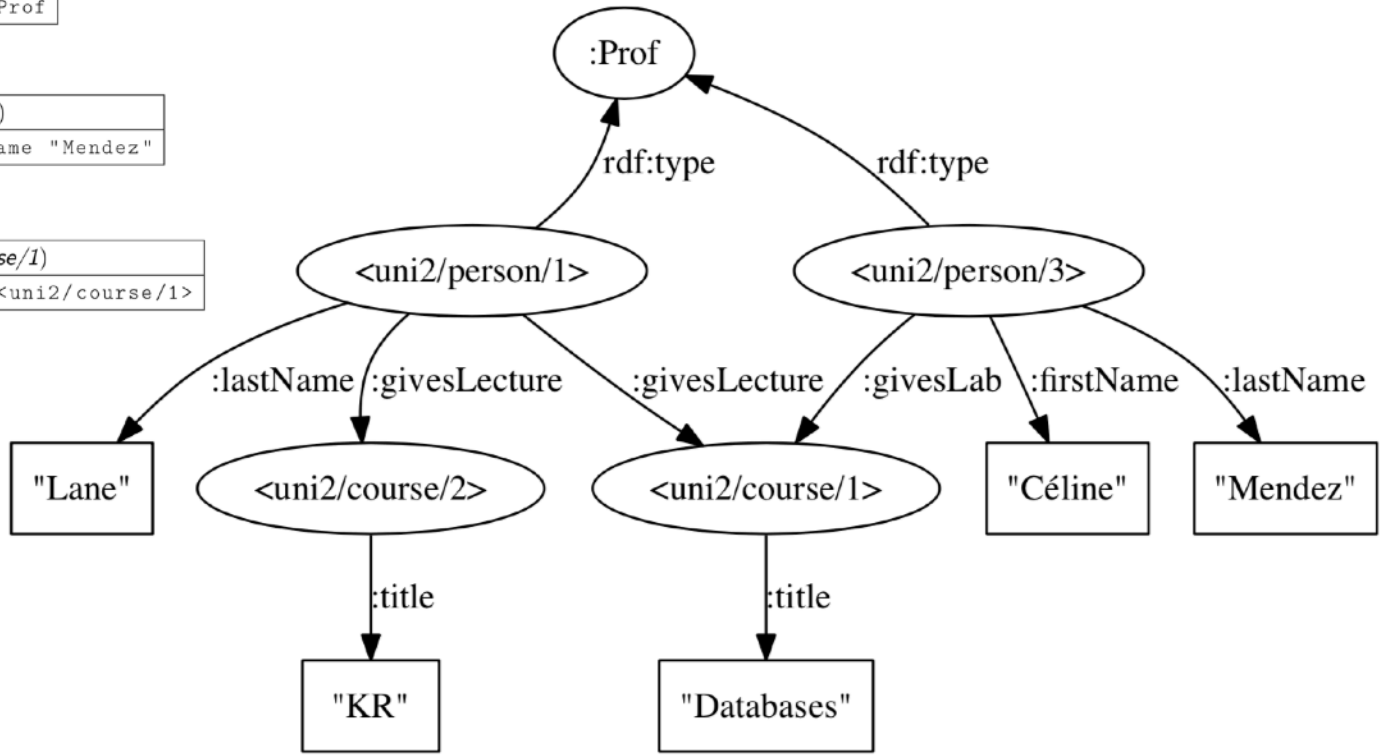
RDF triple	`<uni2/person/1> rdf:type :Prof`

Attribute of an individual:

Fact	lastName(*uni2/person/3*, 'Mendez')
RDF triple	`<uni2/person/3> foaf:lastName "Mendez"`

Property of an individual:

Fact	givesLab(*uni2/person/3*, *uni2/course/1*)
RDF triple	`<uni2/person/3> :givesLab <uni2/course/1>`

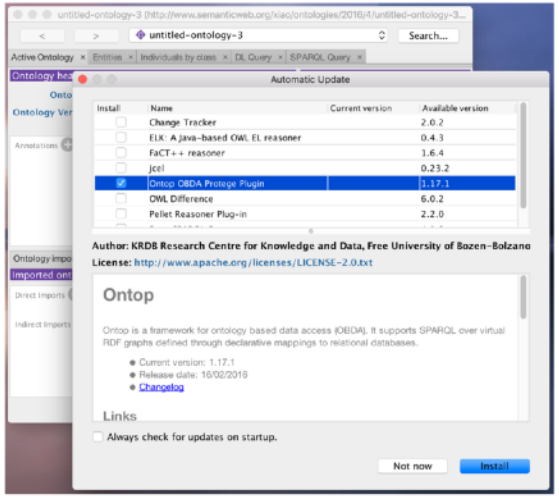


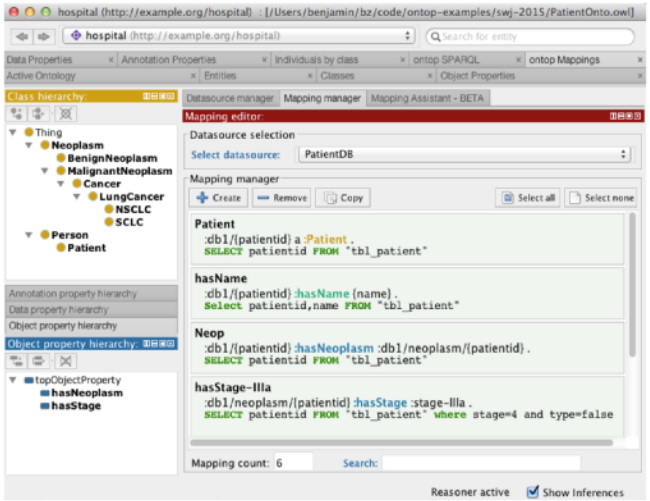
- State-of-the-art OBDA system
- Compliant with the **RDFS**, **OWL 2 QL**, **R2RML**, and **SPARQL standards**
- Supports all major relational DBs (and…)
- **Open-source** and released under *Apache 2* license

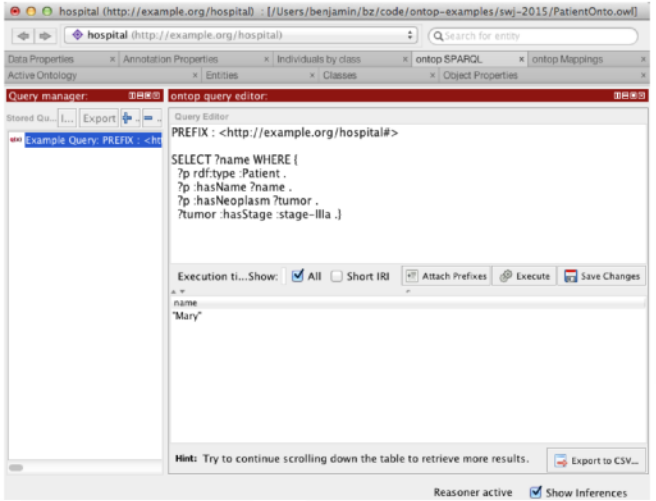
- Development of Ontop:
  - development started 6 years ago already well established:
  - +200 members in the mailing list +7000 downloads in last 18 months
  - main development carried out in the context of the EU project Optique

OBDA system

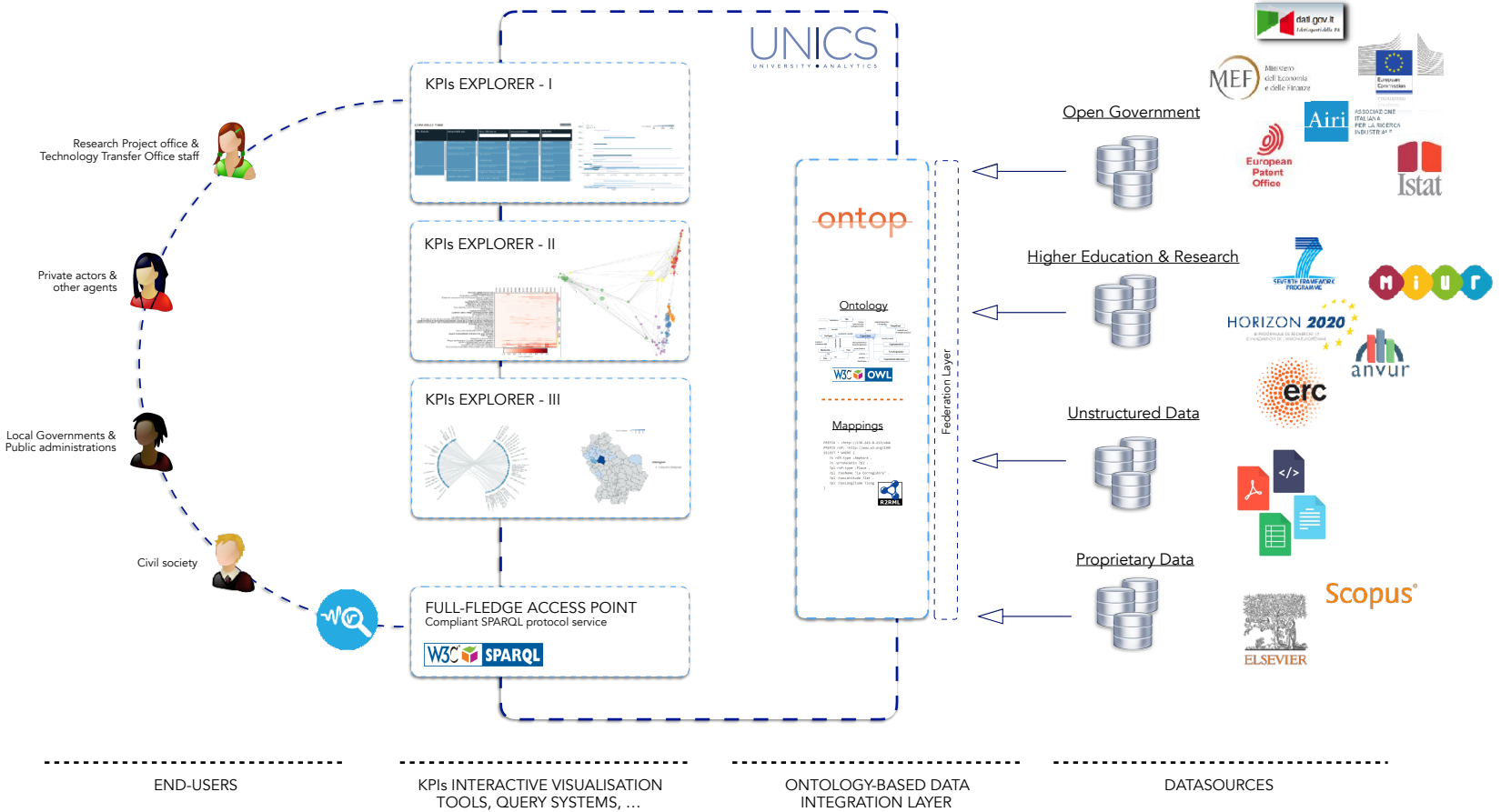
- Ontop website: http://ontop.inf.unibz.it/
- Github: http://github.com/ontop/ontop/
- Facebook: https://www.facebook.com/obdaontop/
- Twitter: @ontop4obda



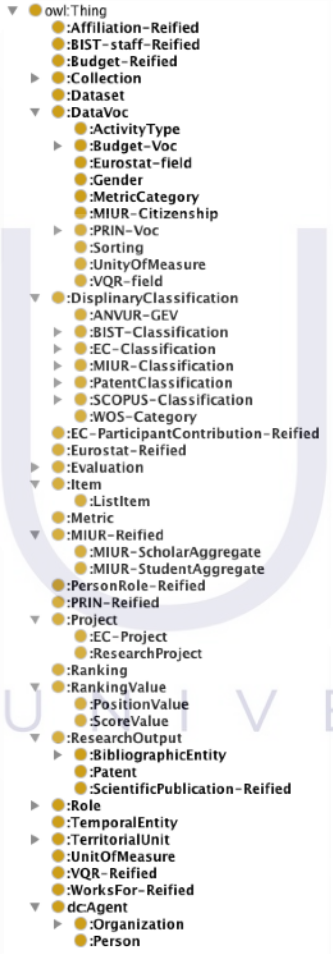


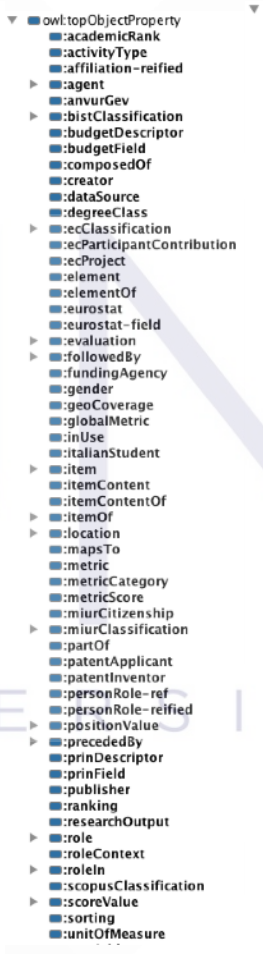


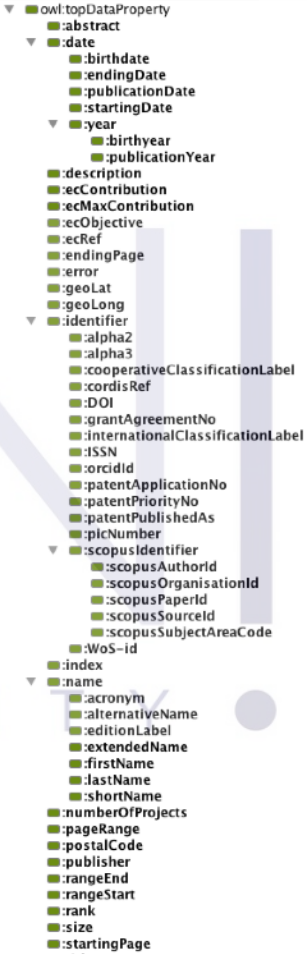
The Open Data Platform UNiCS @SIRIS



The UNiCS domain ontology

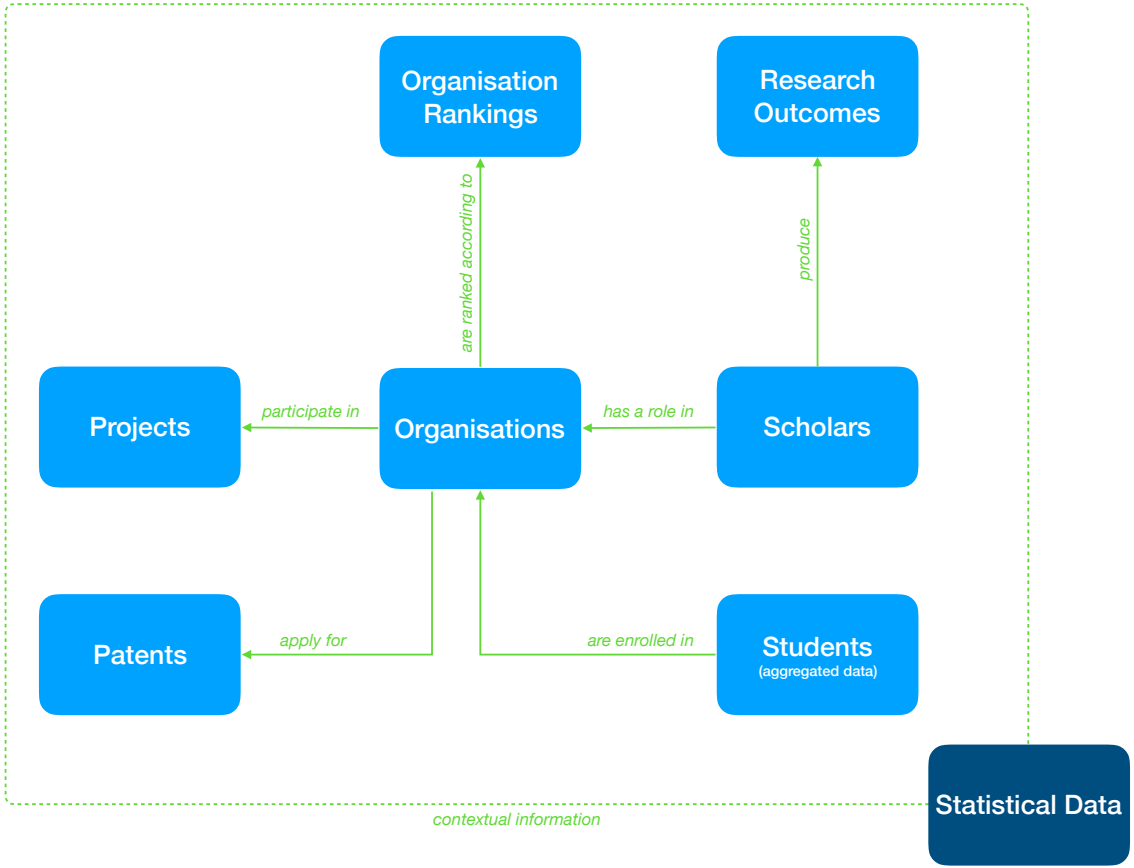




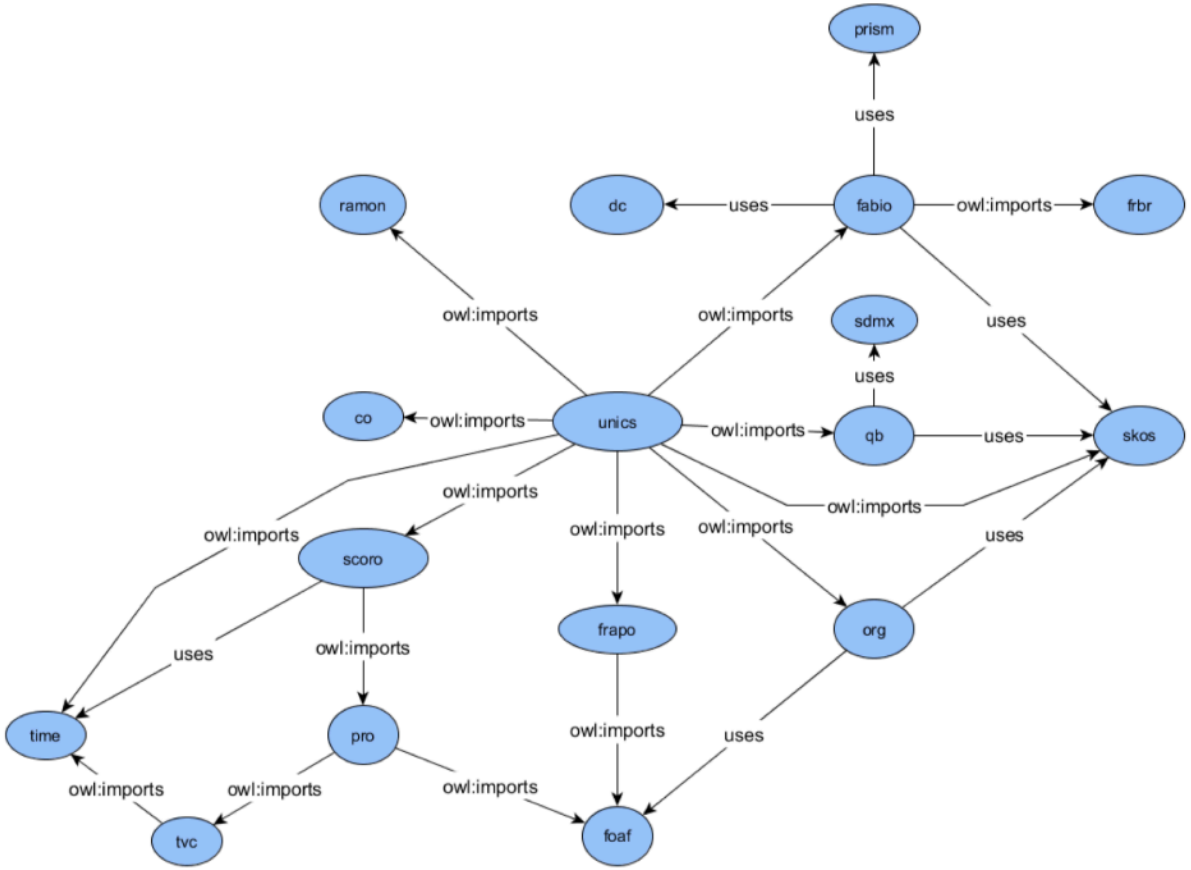


1711 owl:ObjectProperty
1174 owl:DatatypeProperty
3096 owl:Class

UNiCS: main modules



Current alignments



org	W3C - The Organization Ontology
frapo	Funding, Research Administration and Projects Ontology
dc	DUBLIN Core
skos	W3C - SKOS Simple Knowledge Organization System
ramon	RDF version of the NUTS classification system Eurostat
foaf	FOAF Essential
scoro	The Scholarly Contributions and Roles Ontology
pro	Publishing Roles Ontology
ti	Time Interval Ontology
tvc	Time-indexed Value in Context
qb	W3C - The RDF Data Cube Vocabulary
time	W3C - OWL Time
sdmx	Statistical Data and Metadata eXchange
co	Collection Ontology
fabio	The FRBR-aligned Bibliographic Ontology
prism	Publishing Requirements for Industry Standard Metadata
frbr	Functional Requirements for Bibliographic Record

The UNiCS mappings

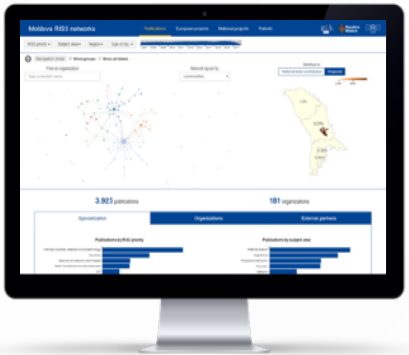
```
mappingId   M_project_class
target      :Project-{p_id} a :Project ; :identifier {p_id} .
source      select
                p.id as p_id
            from
                unics.project p

mappingId   M_project_acro
target      :Project-{p_id} :acronym {p_acro} .
source      select
                p.id as p_id,
                p.acronym as p_acro
            from
                unics.project p

mappingId   M_project_tit
target      :Project-{p_id} :extendedName {p_tit} .
source      select
                p.id as p_id,
                p.title as p_tit
            from
                unics.project p

mappingId   M_project_call
target      :Project-{p_id} :ecCall {ec_call} .
source      select
                p.id as p_id,
                p.ec_call as ec_call
            from
                unics.project p

mappingId   M_project_funsche
target      :Project-{p_id} :ecFundingScheme :ECFundingScheme-{ec_fun_scheme} .
source      select
                p.id as p_id,
                p.ec_fund_scheme as ec_fun_scheme
            from
                unics.project p
            where
                p.ec_fund_scheme is not null

mappingId   M_funsche_class
target      :ECFundingScheme-{fs_code} a :EC-FundingScheme ; :shortName {fs_code} .
source      select
                fs.code as fs_code
            from
                unics.funding_scheme fs
```

Some of our platforms

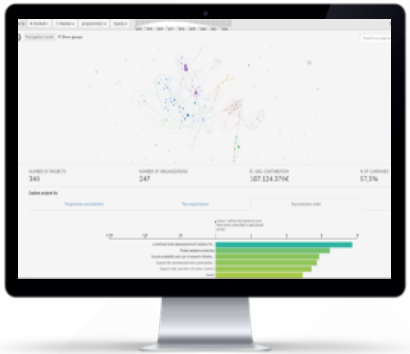
SIRIS A·C·A·D·E·M·I·C

Regional Observatory for Research and Innovation

H2020 regional explorer



Smart Specialisation (S3) Monitoring Tool



Research Information System @PSL

Open Contracting Open Basento

UNiCS The Open Platform for HERI

Toscana “Observatory for Research and Innovation”

TEACHING AND RESEARCH STAFF

The staff involved in research and teaching in universities, high schools and research centres in Tuscany form a fundamental pillar of the regional research and innovation system. The graphs, all interactive, allow you to observe the trend over time of the number of academic staff divided by gender, age, classification and subject with two levels of detail: the 14 areas of the National University Council (CUN) and more than 300 scientific disciplinary sectors (SSD). The data can be viewed at the aggregate regional level or for each university, and in some cases there is a comparison with the national level. Although in Tuscany there are national public research centres of excellence, the information relating to them has not been integrated into the portal as they are not available in an open format on a territorial basis, but only with reference to the headquarters.

4.142
Teachers
TOTAL NUMBER OF TEACHERS AND RESEARCHERS (2018)

53
Years
AVERAGE AGE OF TEACHERS AND RESEARCHERS (2014)

37
%
PERCENTAGE OF FEMALE TEACHERS AND RESEARCHERS (2017)

STUDENTS

Human capital, especially young people with university education, is today one of the main resources for the economic and social development of the territory. Tuscany guarantees a very rich and diversified range of training options and offers a wide range of opportunities both in terms of subject areas and types of courses.
The graphs show the number of students enrolled and graduating from Tuscan universities divided by year, residence, diploma of studies of origin and degree classes.

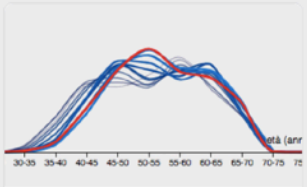
19.323
Registrations
NUMBER OF STUDENTS REGISTERED TO MASTER'S DEGREES (ACADEMIC YEAR 2017/18)

21.770
Registrations
NUMBER OF STUDENTS REGISTERED TO SINGLE CYCLE DEGREES (ACADEMIC YEAR 2017/18)

67.347
Registrations
NUMBER OF STUDENTS REGISTERED TO BACHELOR'S DEGREES (ACADEMIC YEAR 2017/18)

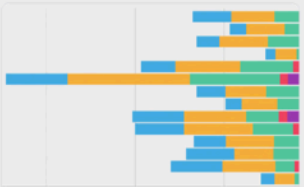
2.003
Registrations
NUMBER OF INTERNATIONAL STUDENTS REGISTERED TO BACHELOR'S DEGREES (ACADEMIC YEAR 2017/18)

4.678
Registrations
NUMBER OF INTERNATIONAL STUDENTS REGISTERED TO MASTER'S DEGREES



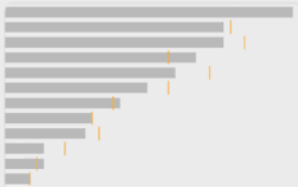
Distribution by age, role and CUN area of teachers and researchers

The graphs show the average age of teaching staff [...]



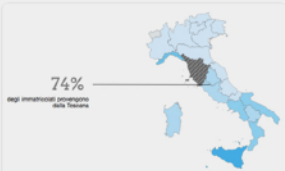
Distribution by gender, role and CUN area of teachers and researchers – Tuscany

The graph shows, for each year, the number of [...]



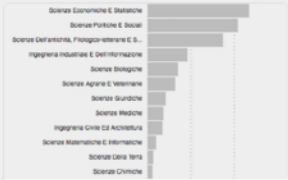
Number of students per teacher and per subject area

The graph shows the relationship between the number of students [...]



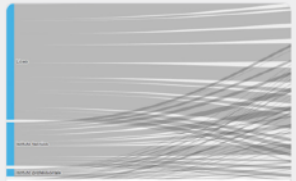
Provenance of enrolled students by province of residence

The figure shows provenance by geographical area of residence (Region [...]



Residents in Tuscany enrolled in universities away from home

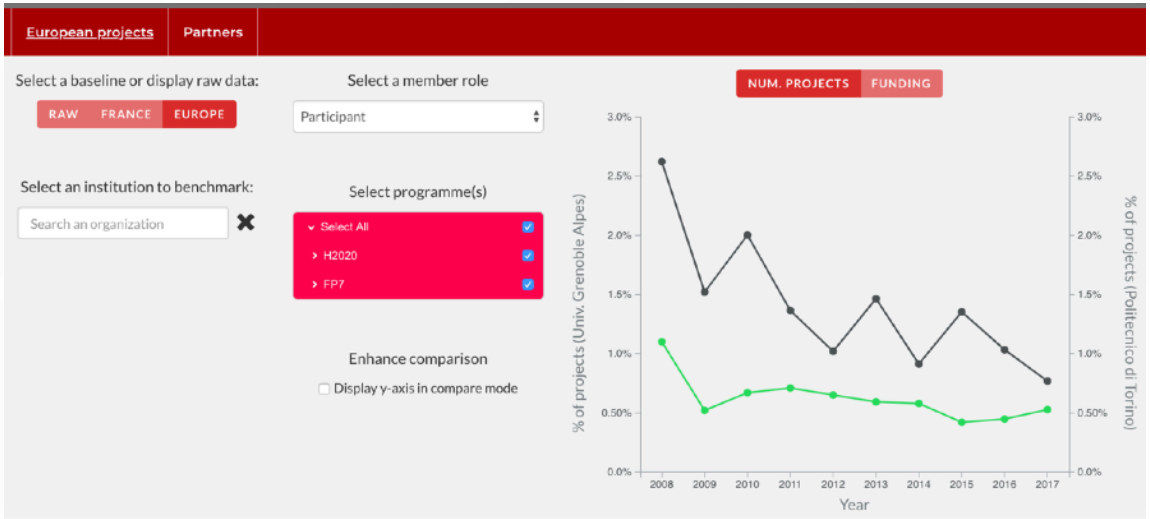
These three graphs explore the flow of residents in Tuscany [...]



Enrolment by type of diploma and degree class

The display shows the flow of enrolments by type of [...]

IDEX Université Grenoble Alpes “EU Projects Explorer”



The UNiCS Open Data portal

UNICS
UNIVERSITY ANALYTICS
The open platform for Higher Education and Research

QUERY

http://endpoint.unics.cloud/public/sparql

```
PREFIX : <http://www.semanticweb.org/ontologies/2016/4/untitled-ontology-69#>
SELECT
    ?acronym ?title ?costi ?shortName (count(distinct ?participant) as ?numberOfParticipants)
WHERE {
    ?prj :acronym ?acronym .
    ?prj :title ?title .
    ?prj :totalCost ?costi .
    ?prj :ecFundingScheme ?ecFundingScheme .
  FILTER regex(?shortName, "^ERC") .
    ?ecFundingScheme :shortName ?shortName .
    ?prj :ecParticipant ?participant .
}
GROUP BY ?acronym ?title ?costi ?shortName
limit 50
```

Table | Raw Response | Pivot Table | Google Chart | Geo | Execute & Download CSV

Showing 1 to 50 of 50 entries (in 0.955 seconds)

SEARCH: SHOW 50 ENTRIES

	acronym	title	costi	shortName	numberOfParticipants
1	MOTHERTONGUE	The Evolution of Human Languages	"1999973"^^xsd:decimal	ERC-AG	"1"^^xsd:integer
2	Hyperpolarized MRI	Citicoline and deoxyglucose as new molecular imaging probes of DNP hyperpolarized MRI for cancer and neuroimaging	"1650000"^^xsd:decimal	ERC-SG	"1"^^xsd:integer
3	ODMIR	The origins and development of the human mirror neuron system	"1208400"^^xsd:decimal	ERC-SG	"1"^^xsd:integer
4	PHILOQUANTUMGRAVITY	Philosophy of Canonical Quantum Gravity	"1478386"^^xsd:decimal	ERC-SG	"1"^^xsd:integer
5	LOBENA	Long Beamtime Experiments for Nuclear Astrophysics	"1476075"^^xsd:decimal	ERC-SG	"1"^^xsd:integer

QUERY

http://endpoint.unics.cloud/public/sparql

```
    ?cp :ecParticipant ?cMember .
    ?cMember :organization ?org .

    ?org :nuts2 ?nuts2 .
    ?nuts2 :identifier ?nuts2Code .
    ?nuts2 :extendedName ?nuts2Name .

    ?org :extendedName ?orgName .

    FILTER (?nuts2Code = "FR71") #Rhône-Alpes
}
GROUP BY ?orgName ?nuts2Name
ORDER BY DESC(?eurekaProjects)
```

Table | Raw Response | Pivot Table | Google Chart | Geo | Execute & Download CSV

Showing 1 to 8 of 8 entries (in 1.533 seconds)

SEARCH: SHOW 50 ENTRIES

	orgName	nuts2Name	eurekaProjects	cordisProjects
1	Université Lyon 1 - Claude Bernard	Rhône-Alpes	"5"^^xsd:integer	"123"^^xsd:integer
2	Université Grenoble 1 - Joseph Fourier	Rhône-Alpes	"4"^^xsd:integer	"110"^^xsd:integer
3	Université Jean Monnet	Rhône-Alpes	"3"^^xsd:integer	"11"^^xsd:integer
4	Grenoble Institute of Technology	Rhône-Alpes	"2"^^xsd:integer	"85"^^xsd:integer
5	École centrale de Lyon	Rhône-Alpes	"2"^^xsd:integer	"38"^^xsd:integer
6	University of Lyon System	Rhône-Alpes	"1"^^xsd:integer	"13"^^xsd:integer
7	Université de Savoie Mont Blanc	Rhône-Alpes	"1"^^xsd:integer	"10"^^xsd:integer
8	Institut National des Sciences Appliquées de Lyon (INSA)	Rhône-Alpes	"1"^^xsd:integer	"46"^^xsd:integer

Semantic technologies & OBDA: we're thinking about…

Virtual integration over multiple dataset formats (triple store, RDB, no-SQL, for instance)
Statistical/Multidimensional data (is the 'RDF Data Cube' the right solution?)
Semi-automatic generation and testing of mappings
Access control management, distributed ontologies and mappings

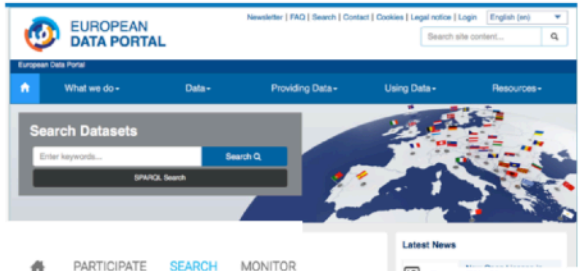
---

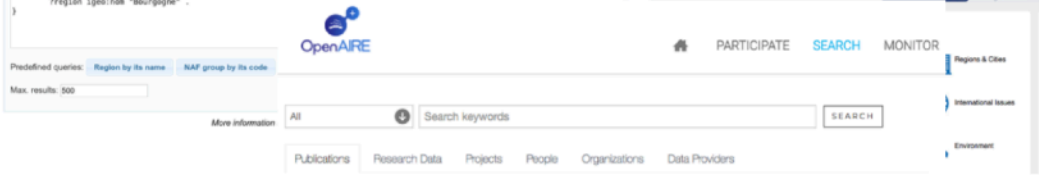
Data access and interoperability (SPARQL 1.1, OWL-DL… they are out there!)
Subject Classification Systems and context sensitive data exploration
Data quality & overall (Linked Data) system resilience

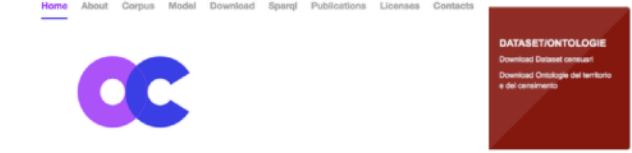
Ontologies specification: costs and benefits
Ontology understanding and use by non-technical users

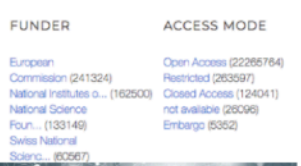
Education and training

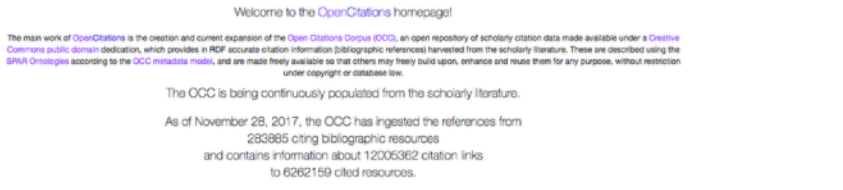
Opportunities?!

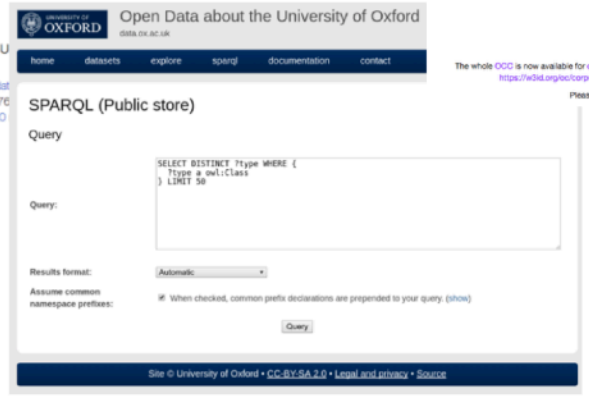


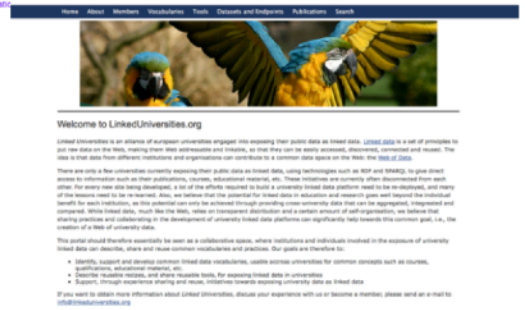






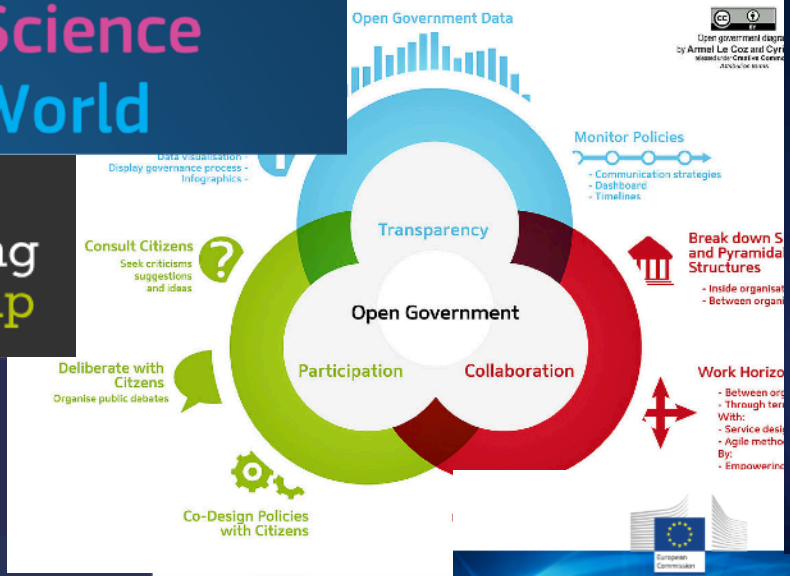


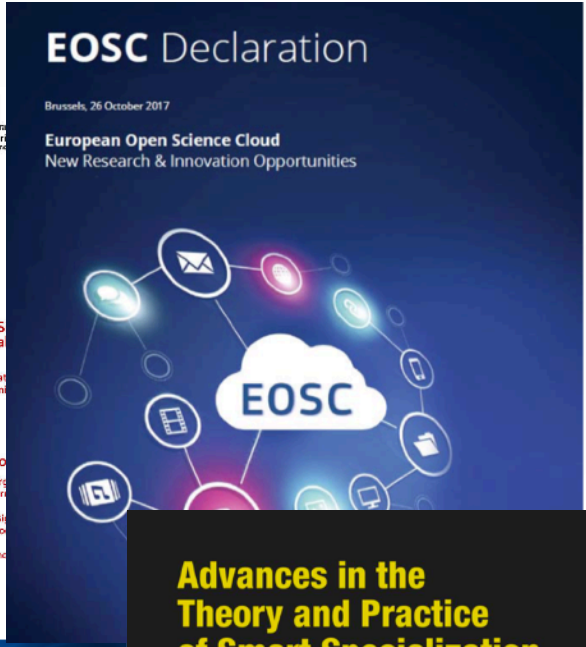





Open Innovation
Open Science
Open to the World


Open
Contracting
Partnership


Open Government Data
Monitor Policies
- Communication strategies
- Dashboard
- Timelines
Transparency
Consult Citizens
Seek criticisms
suggestions
and ideas
Open Government
Participation
Collaboration
Deliberate with
Citzens
Organise public debates
Co-Design Policies
with Citizens

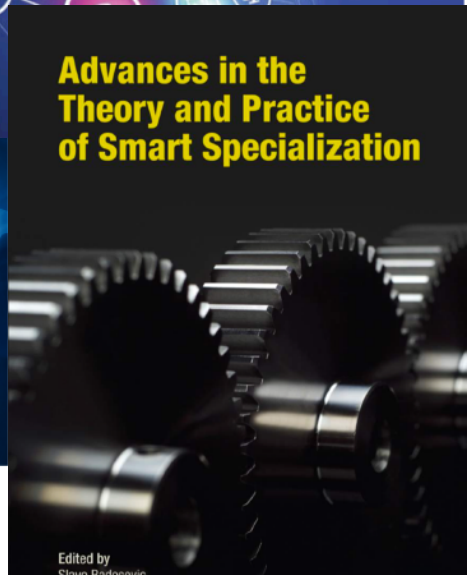

EOSC Declaration
Brussels, 26 October 2017
European Open Science Cloud
New Research & Innovation Opportunities
EOSC


Digital Agenda
2010–2020 for Europe

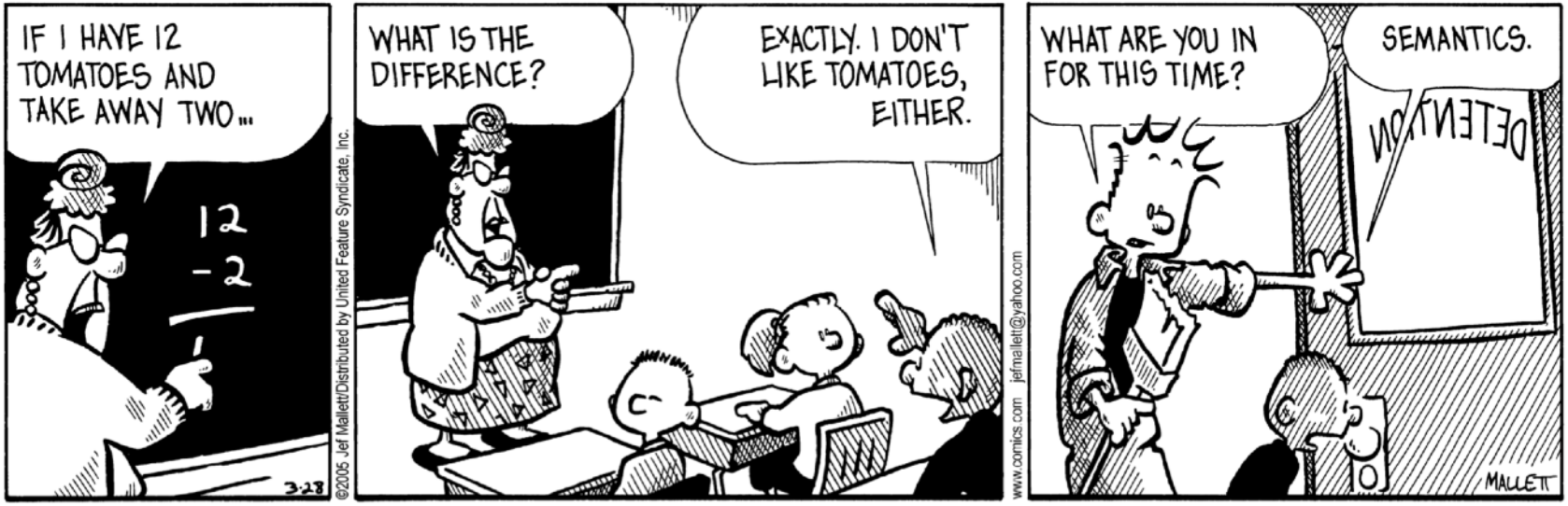

ESFRI
STRATEGY REPORT
ON RESEARCH
INFRASTRUCTURES


Business
Research & education
Public administration
Civil society / Users


Creating Value through Open Data
EUROPEAN DATA PORTAL


Advances in the Theory and Practice of Smart Specialization
Edited by

Questions?



FRAZZ: © Jeff Mallett/Dist. by United Feature Syndicate, Inc.