

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

Alessandro Mosca

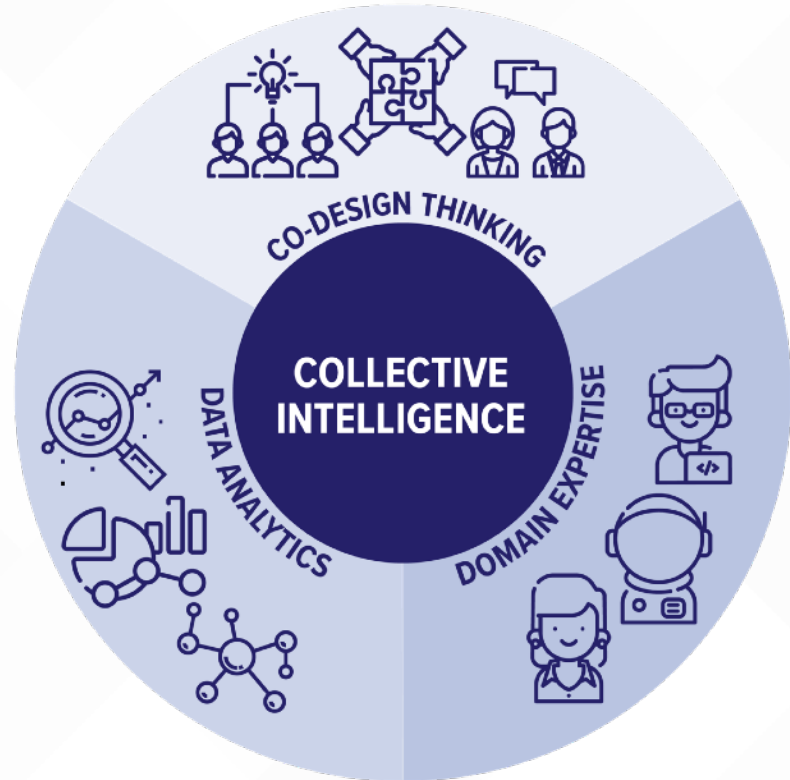
SIRIS LAB, Research Division of SIRIS Academic

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.





Regione Toscana



Barcelona Institute of
Science and Technology



Obra Social
Fundación "la Caixa"

UNIVERSITÀ DELLA CALABRIA



UNIVERSITY
OF CENTRAL ASIA



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.”

The academic structures of Boston, London and Paris: a comparison

Report prepared for CNRS

CLIENT:
Centre National de la Recherche Scientifique
(CNRS)

DATE:
October 2016

SIRI2
ACADEMIC

info@siriacademic.com - www.siriacademic.com

Finançament de la recerca en biomedicina a Espanya

Informe elaborat per a la Fundació "La Caixa"

CLIENT:
Fundació "La Caixa"

DATA:
04 / 08 / 2017

SIRI2
ACADEMIC

info@siriacademic.com - www.siriacademic.com

Uso efficace e complementare dei fondi destinati alla Ricerca, all'Innovazione e all'Alta Formazione nel contesto della programmazione dei fondi strutturali 2014-2020 in Calabria

Proposta Operativa

CLIENT:
Università della Calabria

DATA:
Agosto 2016

SIRI2
ACADEMIC

info@siriacademic.com - www.siriacademic.com

Characterisation of preliminary priority areas for smart specialisation in Moldova

Network analysis for the identification of key
stakeholders for preliminary priority areas
for smart specialisation in Moldova

CLIENT:
EUROPEAN COMMISSION - Joint Research Center

DATA:
28/02/2018
REF:
2018JRC01

SIRI2
ACADEMIC

info@siriacademic.com - www.siriacademic.com

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?”

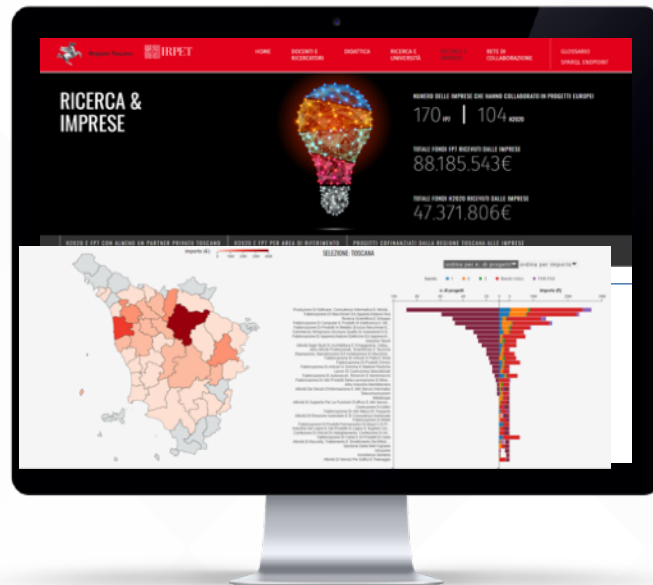
R&I Funding Bodies

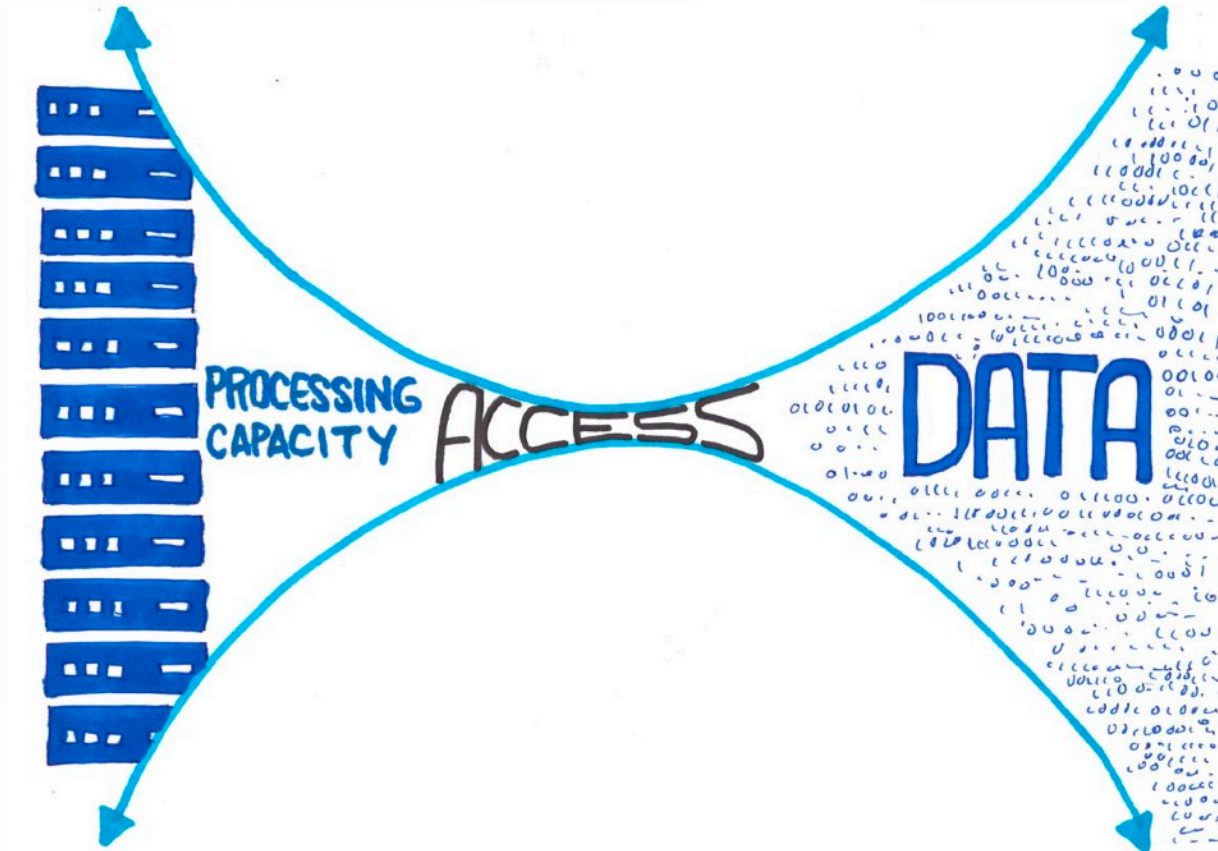


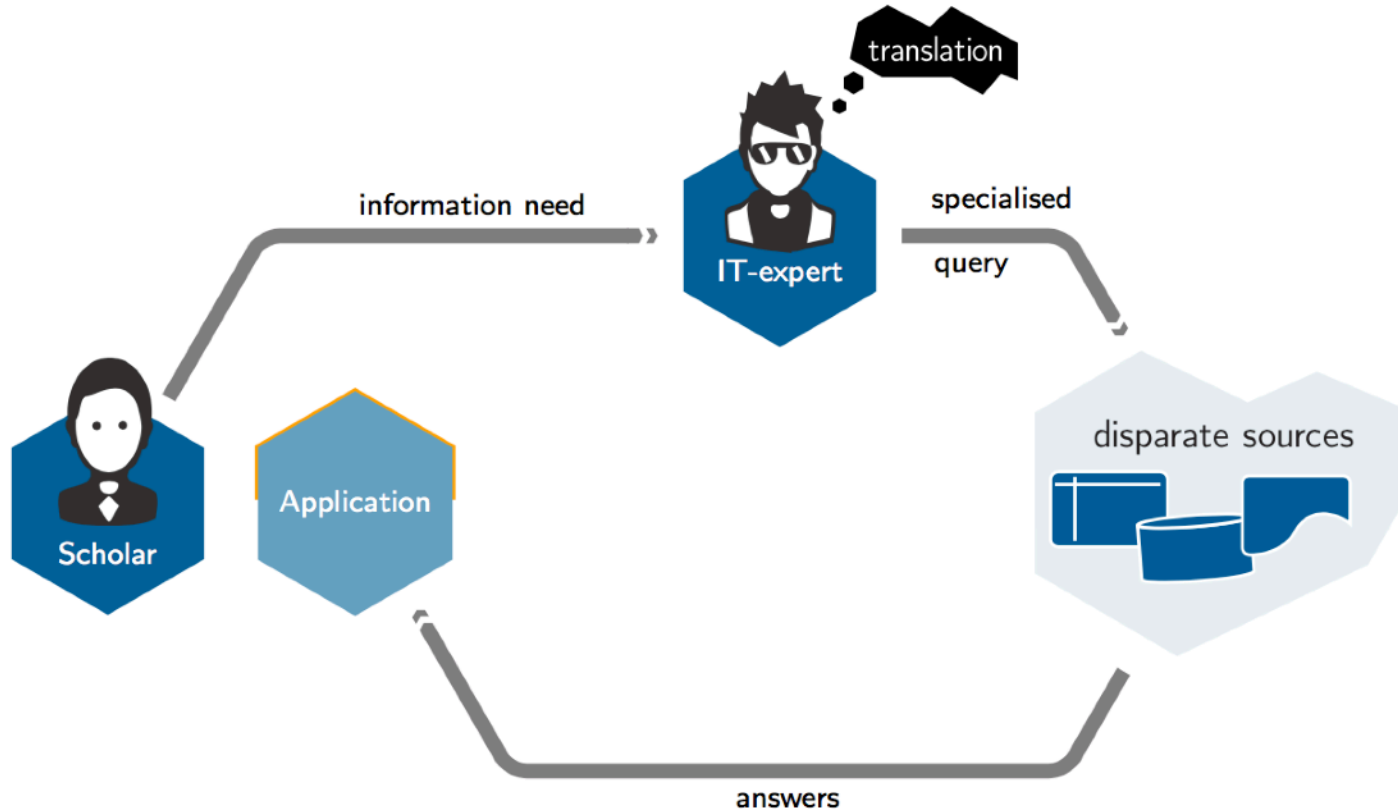
“**Research and innovation ecosystems** are fast changing and fast growing. I need to define our research and innovation **priorities**, but also inform the shaping our future **investments** and then monitor and learn (via specific **indicators**) to eventually reshape our existing policies.”

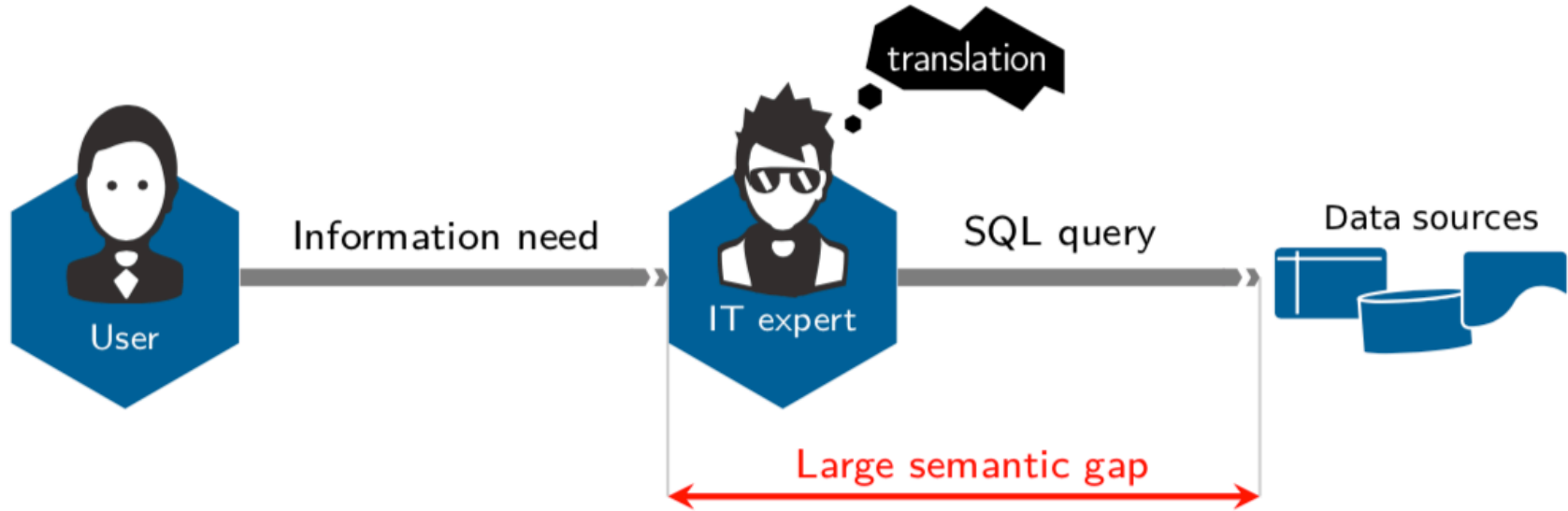
Our solutions provide end-users with:

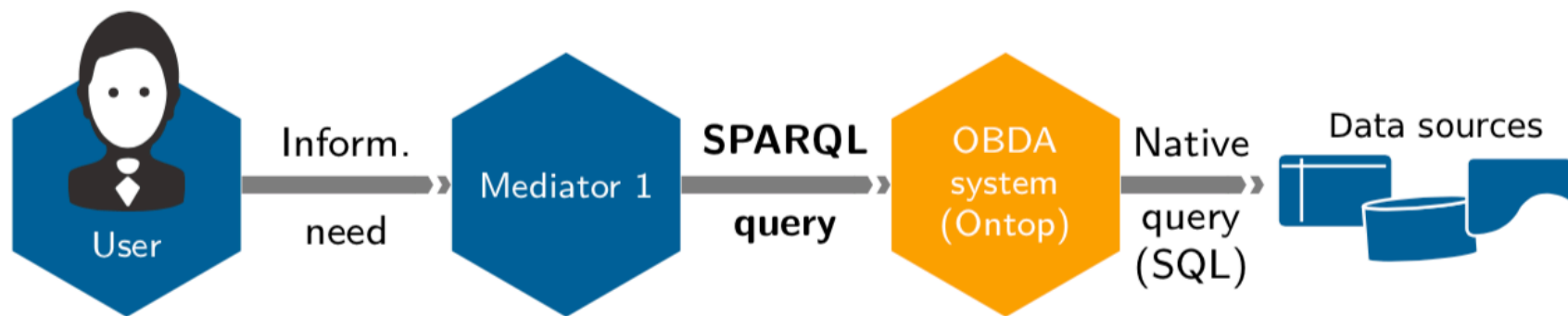
- a running technology for **integrating and accessing data** in a way that is conceptually sound with the **experts domain knowledge**
- a **semantically-transparent platform**, ready to acquire and be complemented with new **data from different sources**
- a theoretically grounded mechanism to **homogenise information** originally stored in **different formats** and according to **different conceptualisations**









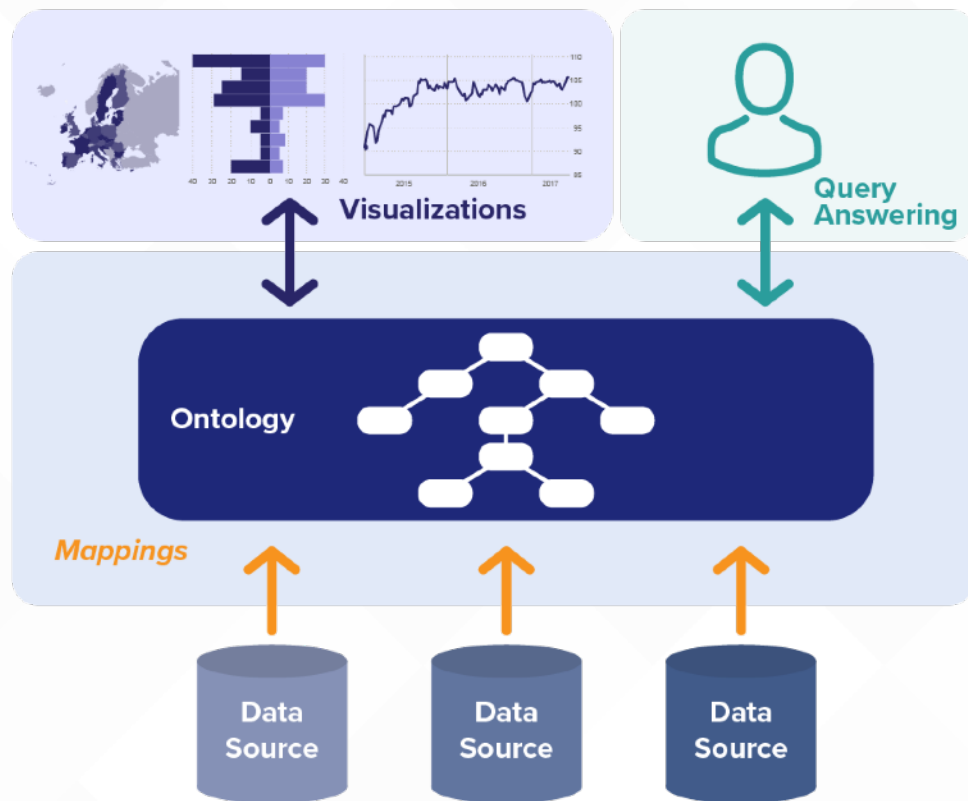


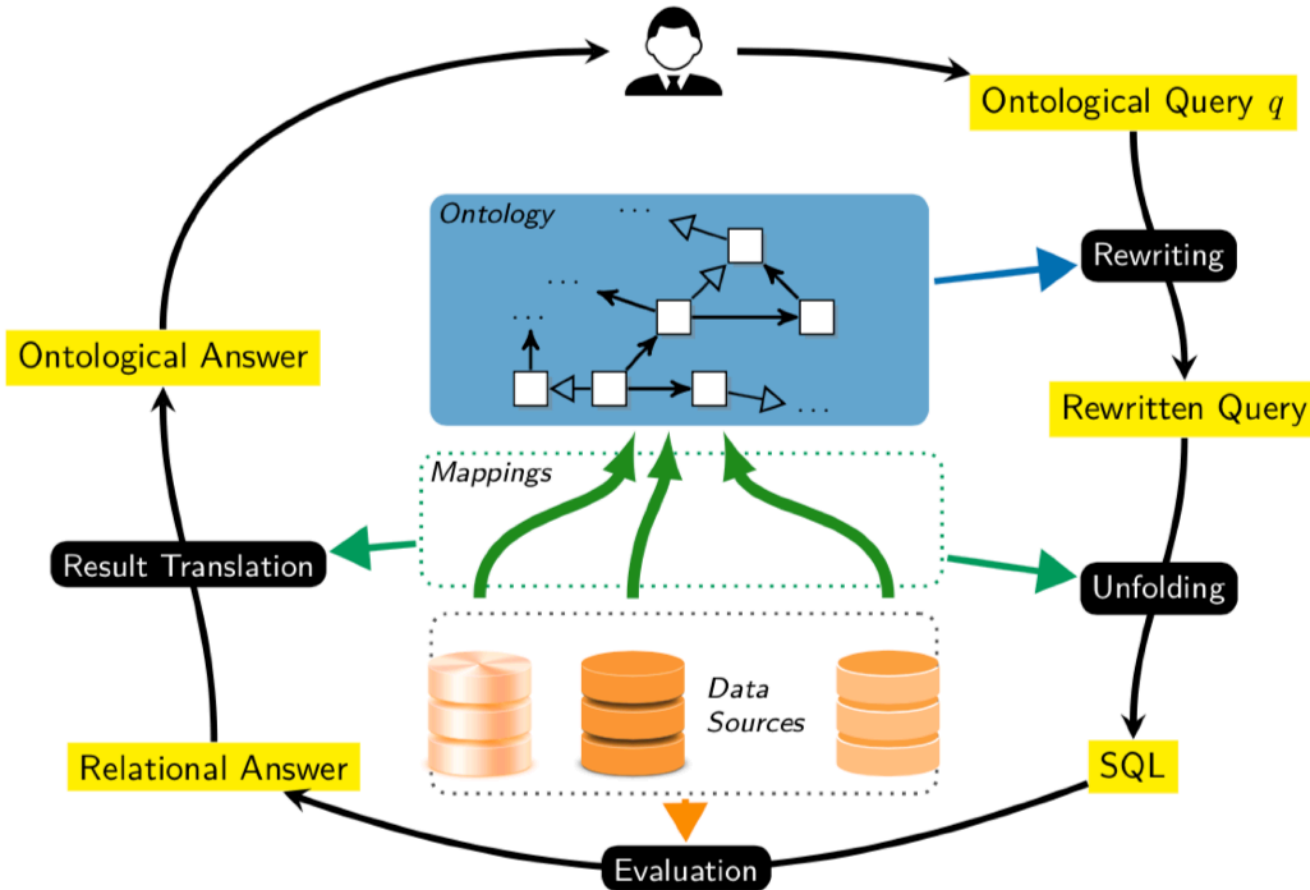
Exploits the **domain knowledge** to support the high-level translation of the information needs, and to enrich the extracted answers.

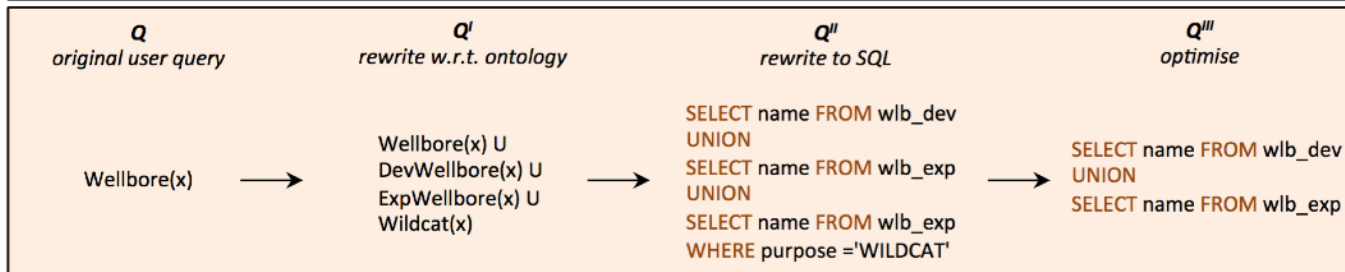
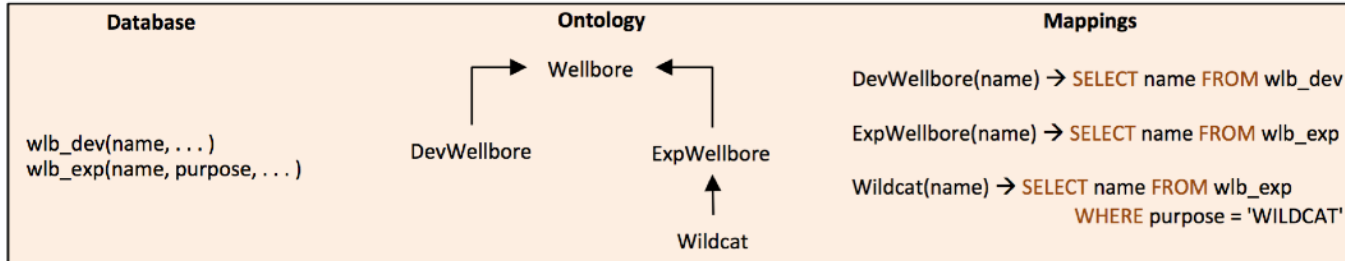
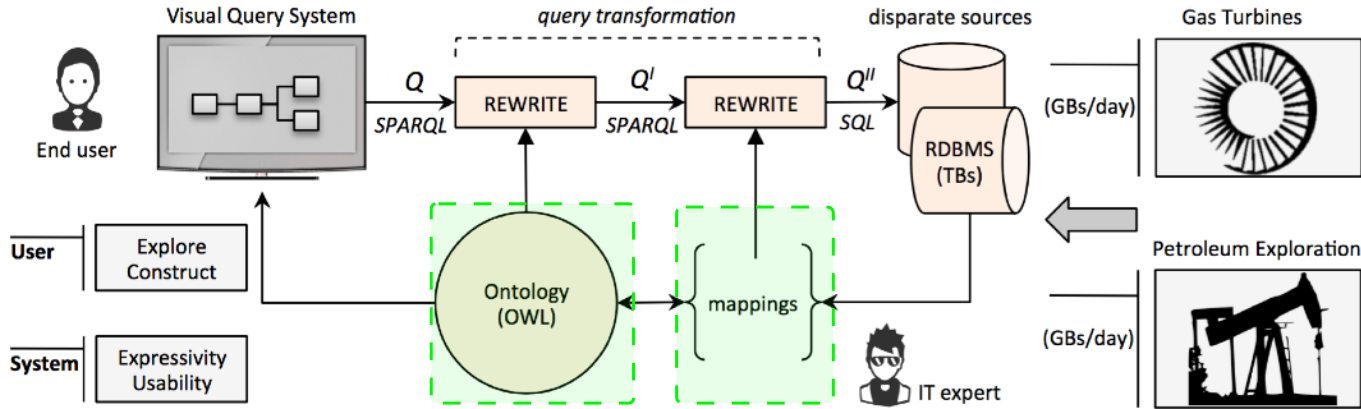
Accepts as input intuitive queries posed by the users on the basis of **domain-centred vocabulary**.

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.







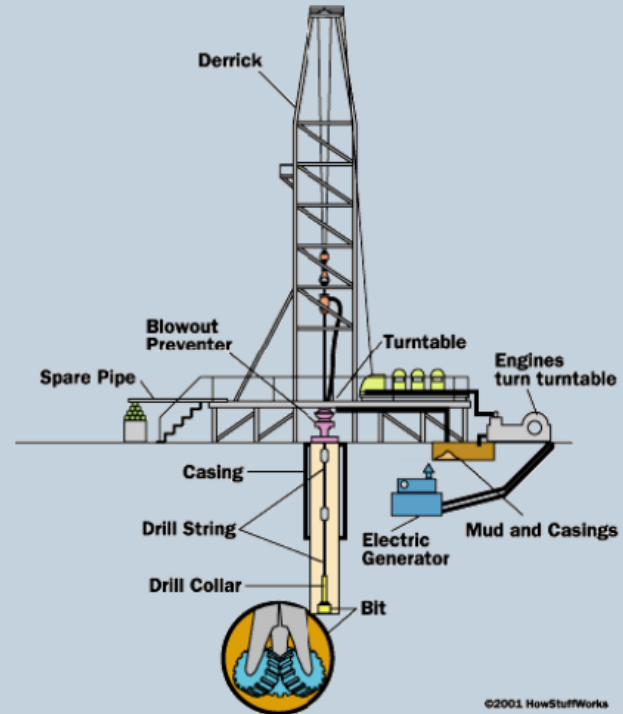
A conceptual model of (some aspect of) the world

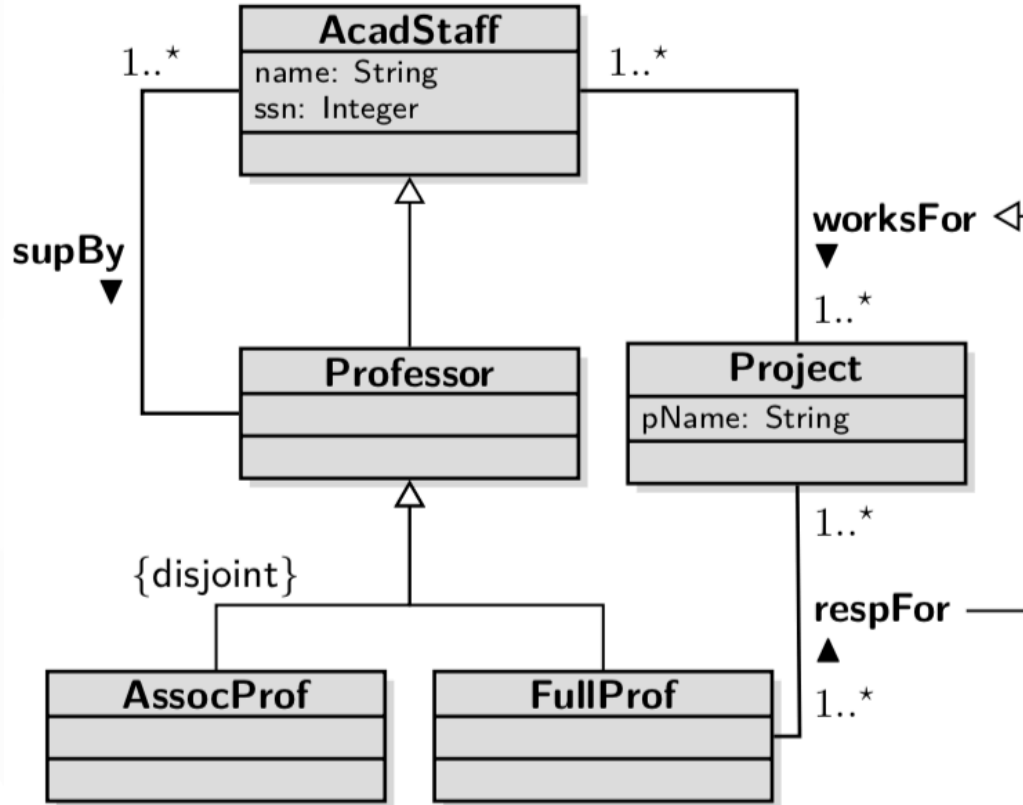
- Introduces **vocabulary** relevant to domain
- Specifies **meaning** (semantics) of terms

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

- Formalised** using suitable logic

OilPipeline \sqsubseteq Pipeline \sqcap
 \exists from Facility.OilFacility





Professor	⊑	AcadStaff
AssocProf	⊑	Professor
FullProf	⊑	Professor
AssocProf	⊑	¬FullProf
AcadStaff	⊑	∃ssn
∃ssn	⊑	AcadStaff
∃ssn ⁻	⊑	Integer
∃worksFor	⊑	AcadStaff
∃worksFor ⁻	⊑	Project
AcadStaff	⊑	∃worksFor
Project	⊑	∃worksFor ⁻
respFor	⊑	worksFor
	⋮	

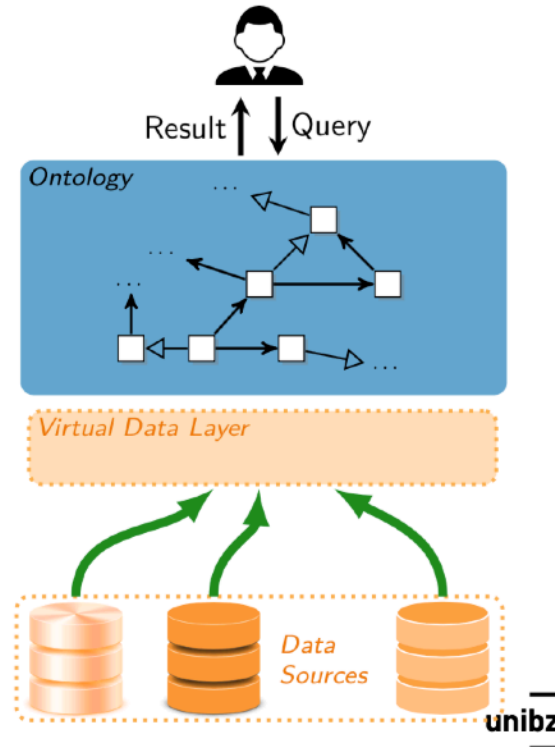
- **OWL 2 QL** is one of the three standard profiles of OWL 2.
- Derived from the **DL-Lite_R** Description Logic of the *DL-Lite-family*:
 - Groups the domain into **classes** of objects with common properties.
 - Binary relations between objects (**object properties**).
 - Binary relations from objects 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 $\mathcal{O} = \langle \mathcal{T}, \mathcal{M}, \mathcal{S}, \mathcal{D} \rangle$, 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} .



Class instance (:Student)

	$q(s) \leftarrow \text{uni1-student}(s, f, l)$
Source	<pre>SELECT s_id FROM uni1.student</pre>
Target	<p>$\text{Student}(\text{URI}_1(s))$</p> <pre>ex:uni1/student/{s_id} a :Student .</pre>

Data property (foaf:firstName)

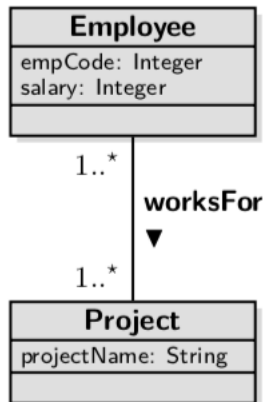
Source (SQL)	<pre>SELECT s_id, firstName, lastName FROM uni1.student</pre>
Target (RDF)	<pre>ex:uni1/student/{s_id} foaf:firstName "{firstName}"^^xsd:string ; foaf:lastName "{lastName}"^^xsd:string .</pre>

Object property (:teaches)

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

Magic number

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



D_1 :

<i>SSN</i>	<i>PrName</i>
23AB	optique
...	...

D_2 :

<i>Code</i>	<i>Salary</i>
e23	1500
...	...

D_3 :

<i>Code</i>	<i>SSN</i>
e23	23AB
...	...

m_1 :
SELECT SSN, PrName
FROM D_1

\leadsto Employee(**pers**(SSN)),
Project(**proj**(PrName)),
projectName(**proj**(PrName), PrName),
worksFor(**pers**(SSN), **proj**(PrName))

m_2 :
SELECT SSN, Salary
FROM D_2 , D_3
WHERE D_2 .Code = D_3 .Code

\leadsto Employee(**pers**(SSN)),
salary(**pers**(SSN), Salary)

Class membership:

Fact	Prof(<i>uni2/person/1</i>)
RDF triple	<code><uni2/person/1> a :Prof</code>

Note: This is an abbreviation for

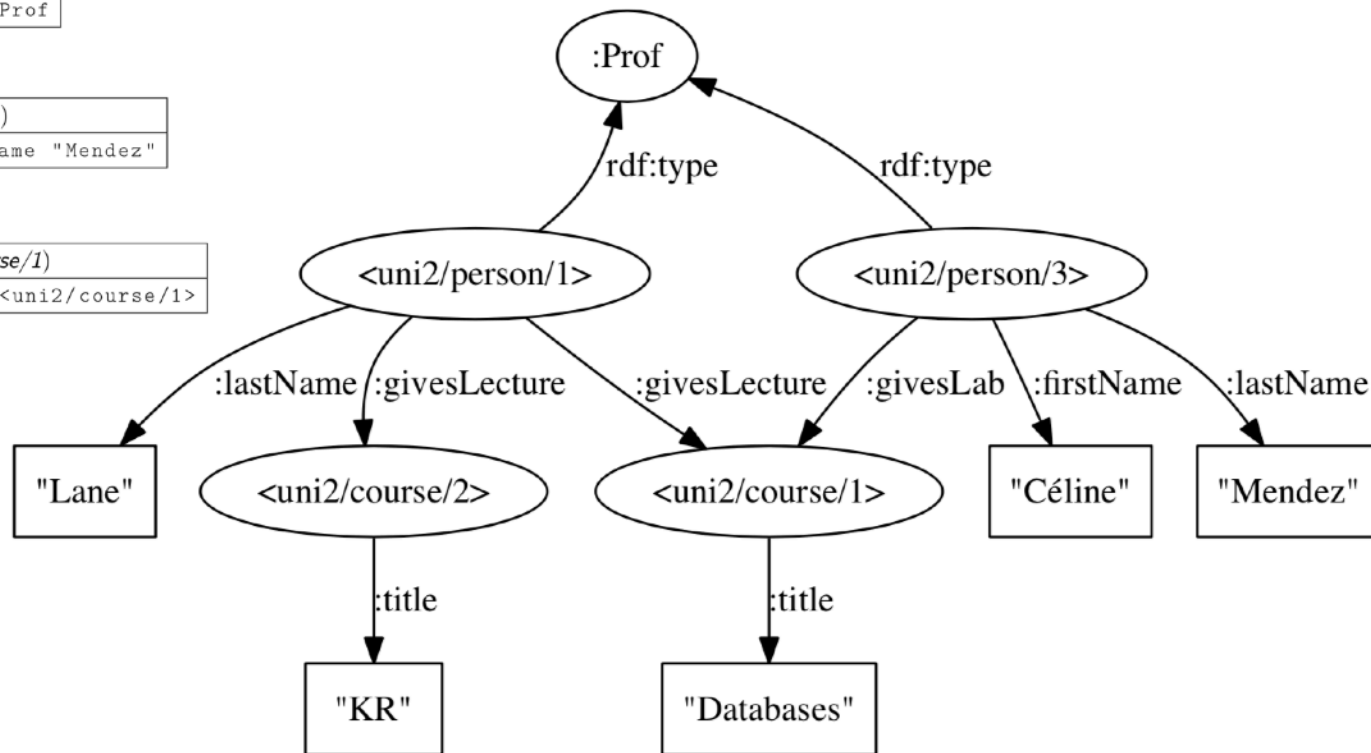
RDF triple	<code><uni2/person/1> rdf:type :Prof</code>
------------	---

Attribute of an individual:

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

Property of an individual:

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

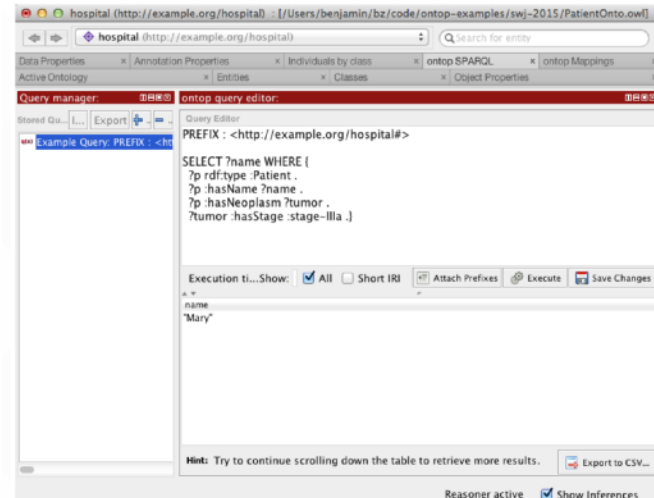
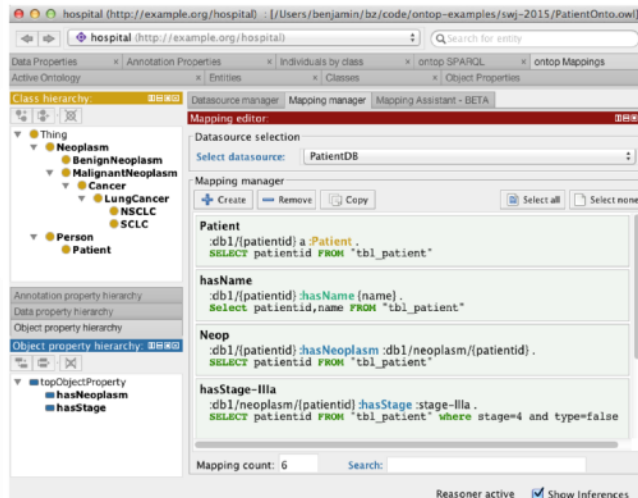
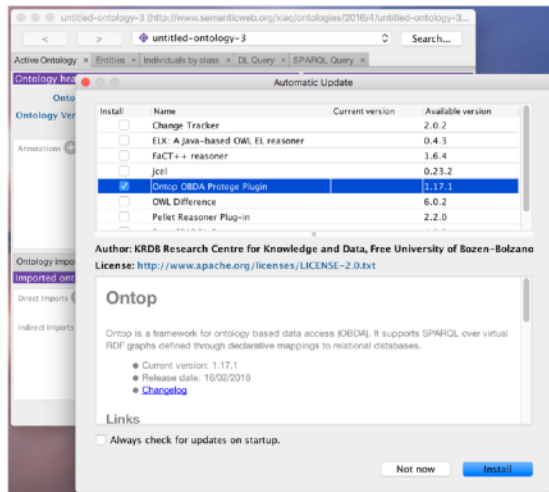


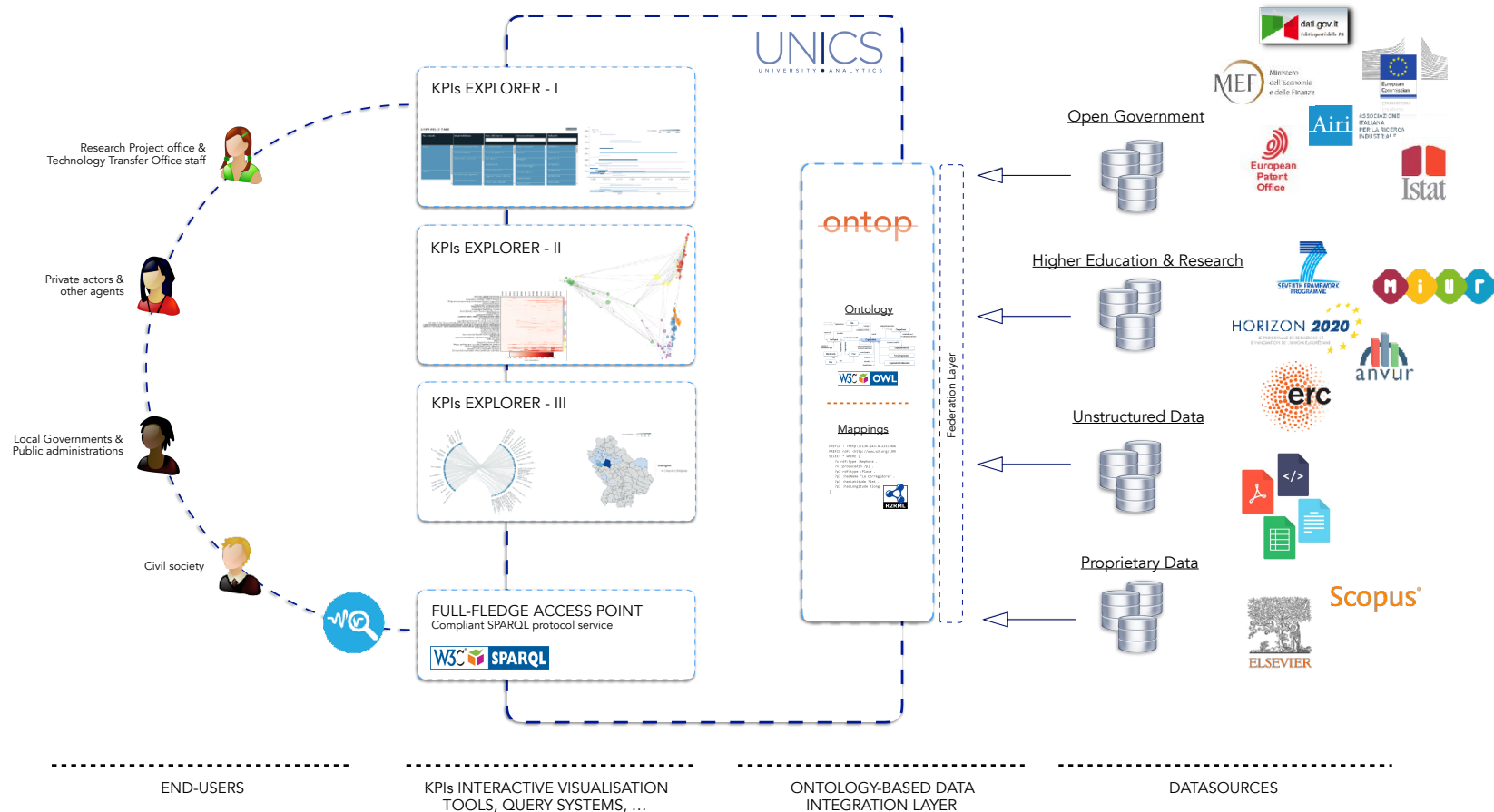


- State-of-the-art OBDA system
- Compliant with the **RDFS**, **OWL 2 QL**, **R2RML**, and **SPARQL standards**
- Supports all major relational DBs (and...)
- **Open-source** and released under *Apache 2* license
- Development of Ontop:
 - development started 6 years ago already well established:
 - +200 members in the mailing list +7000 downloads in last 18 months
 - main development carried out in the context of the EU project **Optique**



- Ontop website: <http://ontop.inf.unibz.it/>
- Github: <http://github.com/ontop/ontop/>
- Facebook: <https://www.facebook.com/obdaontop/>
- Twitter: [@ontop4obda](https://twitter.com/ontop4obda)



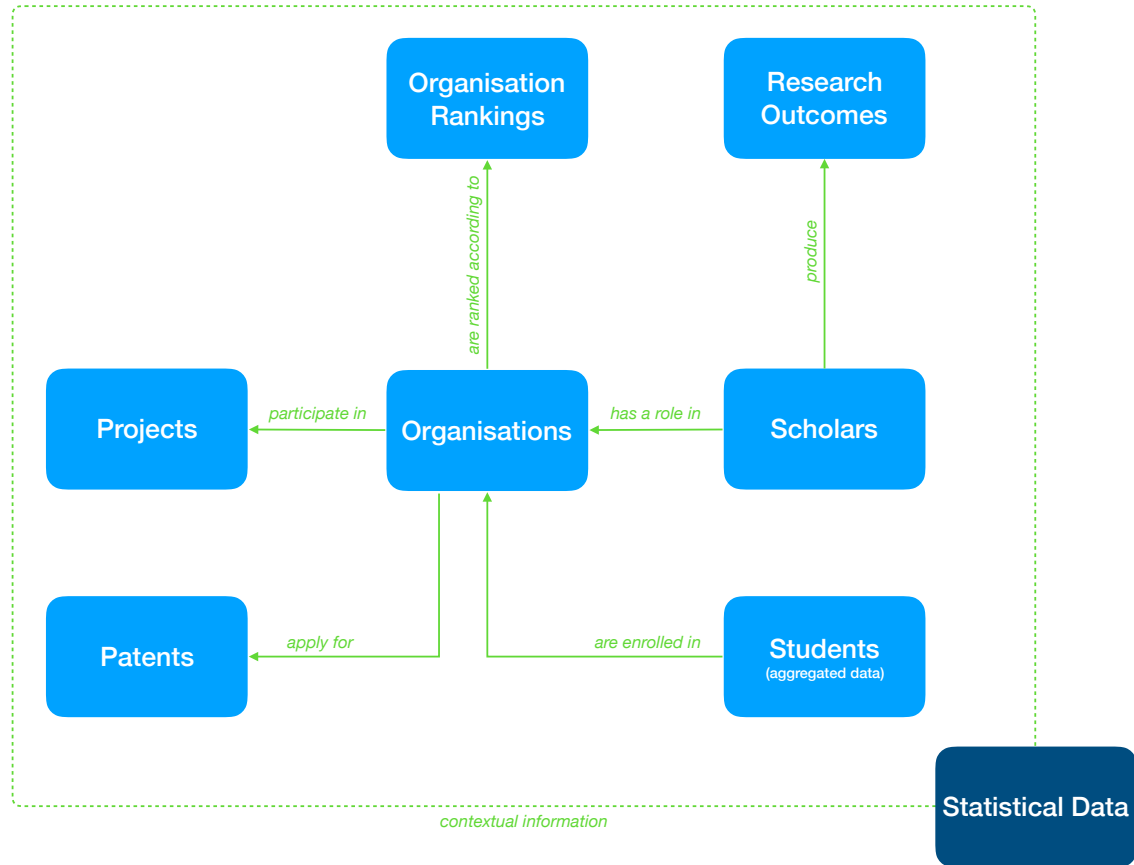


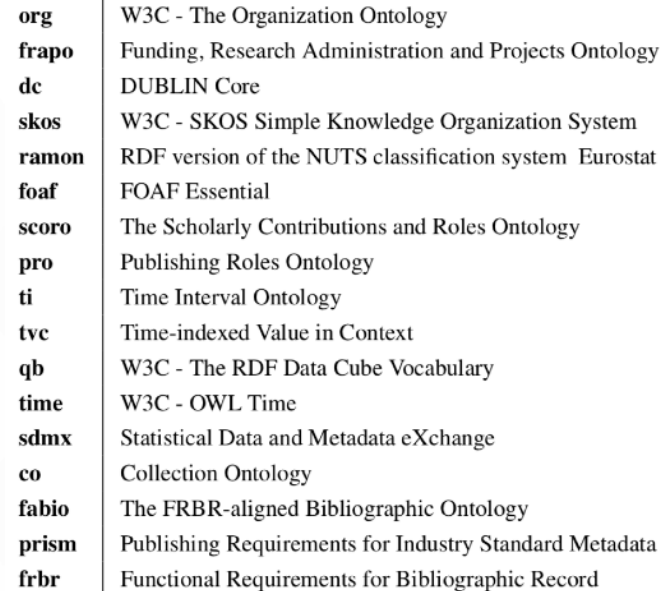


1711 owl:ObjectProperty

1174 owl:DatatypeProperty

3096 owl:Class





```
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
```

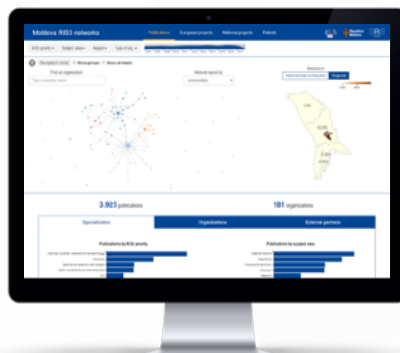
Regional Observatory for Research and Innovation



Research Information System @PSL



H2020 regional explorer



Open Contracting Open Basento



Smart Specialisation (S3) Monitoring Tool



UNiCS The Open Platform for HERI



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

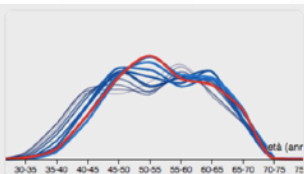
Years

AVERAGE AGE OF TEACHERS AND RESEARCHERS (2014)

37

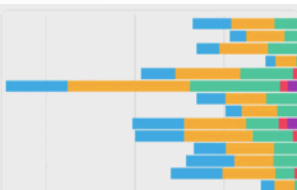
%

PERCENTAGE OF FEMALE TEACHERS AND RESEARCHERS
(2017)



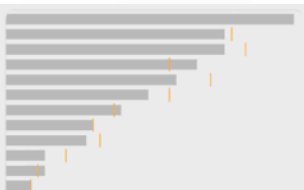
Distribution by age, role and CUN area of teachers and researchers

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 [...]

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 MASTERS 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 BACHELORS DEGREES (ACADEMIC YEAR 2017/18)

2.003

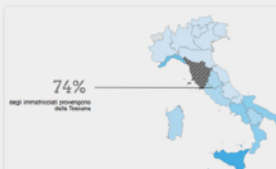
Registrations

NUMBER OF INTERNATIONAL STUDENTS REGISTERED TO BACHELORS DEGREES (ACADEMIC YEAR 2017/18)

4.678

Registrations

NUMBER OF INTERNATIONAL STUDENTS REGISTERED TO MASTERS DEGREES



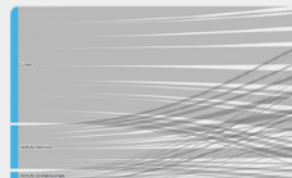
Provenance of enrolled students by province of residence

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



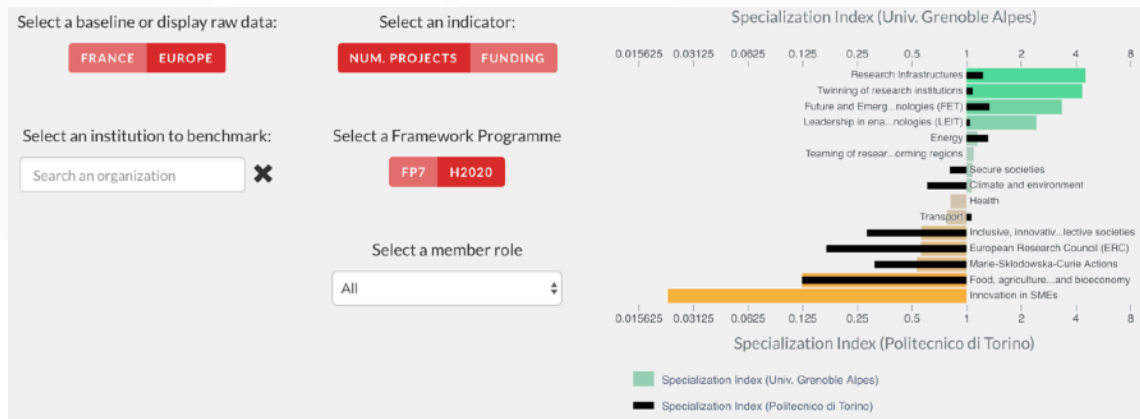
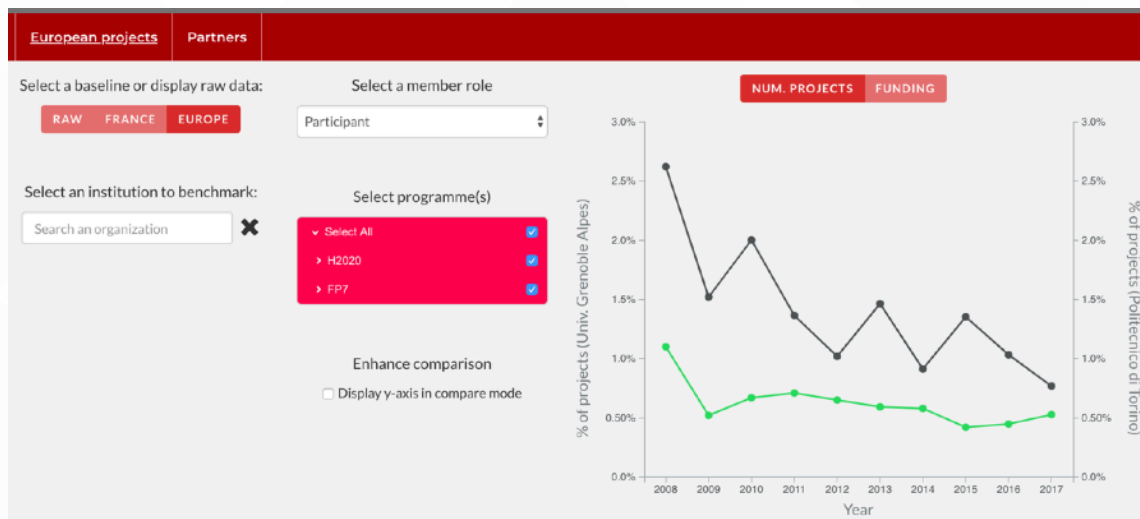
Residents in Tuscany enrolled in universities away from home

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 [...]





QUERY X +

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

```

1 PREFIX : <http://www.semanticweb.org/ontologies/2016/4/untitled-ontology-69#>
2 SELECT
3   ?acronym ?title ?costi ?shortName (count(distinct ?participant) as ?numberOfParticipants)
4 WHERE {
5   ?prj :acronym ?acronym .
6   ?prj :title ?title .
7   ?prj :totalCost ?costi .
8   ?prj :ecFundingScheme ?ecFundingScheme .
9   FILTER regex(?shortName, "ERC") .
10  ?ecFundingScheme :shortName ?shortName .
11  ?prj :ecParticipant ?participant .
12 }
13 GROUP BY ?acronym ?title ?costi ?shortName
14 limit 50
15

```

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

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

acronym	title	costi	shortName	numberOfParticipants
1 MOTHERTONGUE	The Evolution of Human Languages	¹ 99973 ⁺	ERC-AG	¹ ⁺
2 Hyperpolarized MRI	Citicoline and deoxyglucose as new molecular imaging probes of DNP hyperpolarized MRI for cancer and neuroimaging	¹ 650000 ⁺	ERC-SG	¹ ⁺
3 ODMIR	The origins and development of the human mirror neuron system	¹ 208400 ⁺	ERC-SG	¹ ⁺
4 PHILOQUANTUMGRAVITY	Philosophy of Canonical Quantum Gravity	¹ 478386 ⁺	ERC-SG	¹ ⁺
5 LOBENA	Long Beamtime Experiments for Nuclear Astrophysics	¹ 476075 ⁺	ERC-SG	¹ ⁺

QUERY X +

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

```

17 ?ecParticipant ?cMember .
18 ?cMember :organization ?org .
19
20 ?org :nuts2 ?nuts2 .
21 ?nuts2 :identifier ?nuts2Code .
22 ?nuts2 :extendedName ?nuts2Name .
23
24 ?org :extendedName ?orgName .
25
26 FILTER (?nuts2Code = "FK71") #Rhône-Alpes
27 }
28 GROUP BY ?orgName ?nuts2Name
29 ORDER BY DESC(?eurekaProjects)
30

```

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

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

orgName	nuts2Name	eurekaProjects	cordisProjects
1 Université Lyon 1 - Claude Bernard	Rhône-Alpes	⁵ ⁺	¹ 23 ⁺
2 Université Grenoble 1 - Joseph Fourier	Rhône-Alpes	⁴ ⁺	¹ 10 ⁺
3 Université Jean Monnet	Rhône-Alpes	³ ⁺	¹ 1 ⁺
4 Grenoble Institute of Technology	Rhône-Alpes	² ⁺	⁸ 5 ⁺
5 École centrale de Lyon	Rhône-Alpes	² ⁺	³ 8 ⁺
6 University of Lyon System	Rhône-Alpes	¹ ⁺	¹ 3 ⁺
7 Université de Savoie Mont Blanc	Rhône-Alpes	¹ ⁺	¹ 0 ⁺
8 Institut National des Sciences Appliquées de Lyon (INSA)	Rhône-Alpes	¹ ⁺	⁴ 6 ⁺



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?!

DataLift Insee Institut national de la statistique et des études économiques
Mesurer pour comprendre

SPARQL Query

Response format: **HTML** RDF/XML N3/Turtle N-Triples TriX CSV

Query/Label: # newIacer "Bourgoigne" par le non cherché
PREFIX rdf:<http://www.w3.org/1998/02/22-rdf-syntax-ns#>
PREFIX isgo:<http://rdf.insee.fr/def/isgo#>

SELECT ?Region WHERE {
?Region rdf:type isgo:Region ;
?Region isgo:nom "Bourgoigne" .
}

OpenAIRE

Predefined queries: Region by its name NAF group by its code

Max. results: 500

More information

Search keywords

Publications Research Data Projects People Organizations Data Providers

EUROPEAN DATA PORTAL

Newsletter | FAQ | Search | Contact | Cookies | Legal notice | Login | English (en)

Search site content...

European Data Portal

What we do - Data - Providing Data - Using Data - Resources -

Search Datasets

Enter keywords... Search Q

SPARQL Search

Participate SEARCH Monitor

Regions & Cities

International Issues

Environment

Latest News

OpenCitations

PSI 1

linkedopendata PIATTAFORMA SPERIMENTALE PER L'USO DELL'ISTITUTO NAZIONALE DI STATISTICA

Dati del Censimento della popolazione e delle abitazioni 2011

La piattaforma Linked Open Data (LOD) dell'Istat consente di accedere e navigare dati in formato open, sulla base di tecnologie e standard del web semantico. I LOD, interrogabili direttamente da qualsiasi applicazione, rispondono alle esigenze espresse dalle comunità di utilizzatori di disporre di dati standardizzati e interoperabili.

Home About Corpus Model Download Spargi Publications Licenses Contacts



DATASET/ONTOLOGIE
Download Dataset dataset
Download Ontologie del territorio e del censimento

Welcome to the OpenCitations homepage!

The main work of OpenCitations is the creation and current expansion of the **Open Citations Corpus (OCC)**, an open repository of scholarly citation data made available under a **Creative Commons public domain dedication**, which provides in RDF accurate citation information harvested from the scholarly literature. These are described using the **SIRI2 Ontologies** according to the **3CC metadata matrix**, and are made freely available so that others may freely build upon, enhance and reuse them for any purpose, without restriction under copyright or database law.

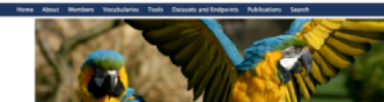
The OCC is being continuously populated from the scholarly literature.

As of November 28, 2017, the OCC has ingested the references from 283885 citing bibliographic resources and contains information about 12005362 citation links to 6262159 cited resources.

The whole OCC is now available for querying (via SPARQL) and for browsing by means <https://linked.org/corpus/for1/>. Additional and more user-friendly interfaces

Please follow us on [Twitter](#) and read the [OpenCitations](#)

Linked Universities



Welcome to LinkedUniversities.org

Linked Universities is an attempt of European universities engaged in exposing their public data as linked data. Linked data is a set of principles to put new data on the Web, making them Web-addressable and reusable, so that they can be easily accessed, discovered, connected and reused. The data is then able from different institutions and organizations can contribute to a common data space on the Web, the Web of Data.

There are only a few universities currently exposing their public data as linked data, using technologies such as RDF and SPARQL, to give direct access to information such as their publications, courses, educational materials, etc. These universities are currently often disconnected from each other. For every new data being developed, a lot of the efforts required to build a university linked data platform need to be re-discovered, and many of the lessons need to be re-learned. Also, we believe that the potential for linked data in education and research goes well beyond the individual benefit for each institution, as this potential can only be achieved through providing open university data that can be aggregated, integrated and compared. While linked data, such as the Web, relies on transparent distribution and a certain amount of self-organization, we believe that sharing practices and collaborating in the development of university linked data platforms can significantly help towards this common goal, i.e., the creation of a Web of university data.

This portal should therefore essentially be seen as a collaborative space, where institutions and individuals involved in the exposure of university linked data can identify, share and reuse common vocabularies and practices. Our goals are therefore to:

- Identify, suggest and develop common linked data vocabularies, usable across universities for common concepts such as courses, qualifications, educational material, etc.
- Search reusable resources, and share reusable tools, for exposing linked data in universities
- Support, through experience sharing and reuse, initiatives towards exposing university data as linked data

If you want to obtain more information about Linked Universities, discuss your experience with us or become a member, please send an e-mail to info@linkeduniversities.org

Springer Nature SciGraph
A Linked Open Data platform for the scholarly domain

We are pleased to introduce Springer Nature SciGraph, the new Linked Open Data platform aggregating data sources from Springer Nature and key partners from the scholarly domain. The Linked Open Data platform will initially contain information from across the research landscape, such as funders, research projects, conferences, affiliations and publications. Additional data, such as citations, patents, clinical trials and usage numbers will follow over time. This high quality data from trusted and reliable sources provides a rich semantic description of how information is related, as well as enabling innovative visualizations of the scholarly domain.

By doing so, Springer Nature SciGraph overcomes former boundaries by relating comprehensive information about the research landscape. It represents a further step in data integration and it will continue to grow organically. This platform will increase the discoverability of high quality data as larger parts of our datasets will be made freely available under a CC-BY-NC 4.0 license.

Any questions?
Please contact us: [Contact](#)

Dataset Download [Download](#)

Licensing Information [Information](#)

Further Info

Conference Presentation 2018
(PDF: 22.98 MB)

EGI Foundat
FET FP7 (76
FET H2020

Open Data about the University of Oxford
DATA.DC.AC.UK

home datasets explore spargi documentation contact

SPARQL (Public store)

Query

SELECT DISTINCT ?type WHERE {
?type a owl:Class
} LIMIT 50

Query:

Results format:

Automatic

Assume common namespace prefixes:

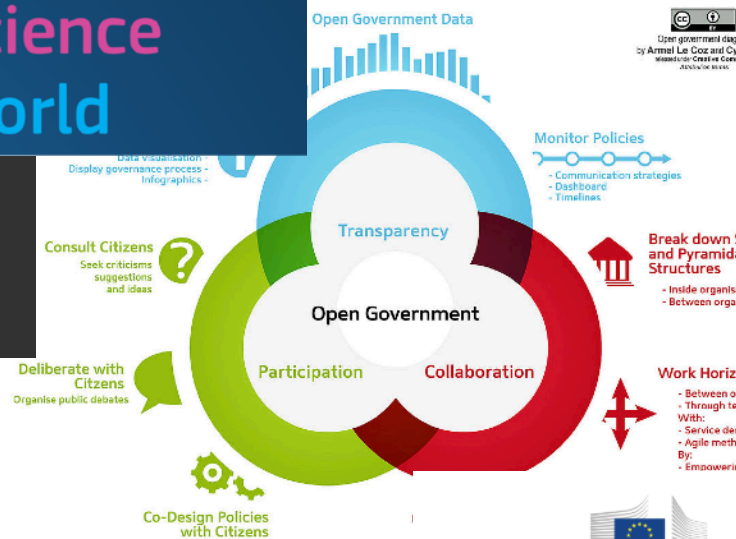
☒ When checked, common prefix declarations are prepended to your query. (show)

Query

Open **Innovation**
Open **Science**
Open to the **World**



Open
Contracting
Partnership



EOSC Declaration

Brussels, 26 October 2017

European Open Science Cloud
New Research & Innovation Opportunities



Digital Agenda

2010-2020

for Europe



Advances in the Theory and Practice of Smart Specialization



Edited by
Simon Richardson

Questions?

