

Ontology-mediated Data Integration and Access in Research and Innovation Policy Making

Alessandro Mosca

SIRIS LAB, Research Division of SIRIS Academic

Who am I?



ALESSANDRO MOSCA
SIRIS Lab Researcher

At some point in life mathematical logic turned me into a computer scientist. Nowadays I'm a SIRIS Lab researcher working in the area of knowledge representation & reasoning, modal and description logics, conceptual modelling and data management.

I studied Philosophy in Milano and did my master's thesis on modal logic-based representations of first order quantification. In year 2006, I finished my PhD dissertation on a formal model for the design of knowledge-based systems for the chemical compound formulation problem. I have been a postdoc researcher at the Department of Computer Science, Systems, and Communication in Milano-Bicocca, working on Knowledge Representation formalisms based on non-classical logics. I've been an Assistant Professor in the KRDB Research Centre, at the Faculty of Computer Science, Free University of Bozen-Bolzano (Italy). I joined the SIRIS LAB Research Division of SIRIS Academic in 2014, in Barcelona.

Who am I?



At some point in life mathematical logic turned me into a computer scientist. Nowadays I'm a SIRIS Lab researcher working in the area of knowledge representation & reasoning, modal and description logics, conceptual modelling and data management.

Where do I come from today?



“We are a European consulting firm designing and implementing evidence-driven strategy and policy solutions for higher education, research and innovation”.

We are a consultancy and research team striving for providing tailor-made services in organisational diagnosis, strategy and change management.

- ❑ Human-based interpretations and analyses, reports, benchmarking
- ❑ Technological tools (mostly based on OBDA/I, Data Viz, and Text mining solutions)
- ❑ Education and training

Agenda of the talk



© European Commission, 2016

- ❑ From scenario description to needs and, possibly, requirements
- ❑ Ontology-mediated and Semantic Web technologies solutions: concept and experiences
- ❑ Everything done?
- ❑ Q&A and Open discussion

Open Innovation: where it comes from



Henry Chesbrough

Open Innovation: A New Paradigm for Understanding Industrial Innovation, 2006

- ❑ Knowledge creation, diffusion, and use ('origination', 'adoption', 'retention', in evolutionary economics terms).
- ❑ Open Innovation, in the original view: *“use of purposive inflows and outflows of knowledge to accelerate internal innovation”*.
- ❑ Linear, bilateral transactions and collaborations.

Need: Transferring knowledge, expertise and resources from one company or research institution to another.

Open Innovation: how to inflow and outflow

CLOSED INNOVATION PRINCIPLES	OPEN INNOVATION PRINCIPLES
The smart people in our field work for us.	Not all the smart people work for us. We need to work with smart people inside and outside our company.
To profit from R&D, we must discover it, develop it, and ship it ourselves.	External R&D can create significant value; internal R&D is needed to claim some portion of that value.
If we discover it ourselves, we will get it to market first.	We don't have to originate the research to profit from it.
The company that gets an innovation to market first will win.	Building a better business model is better than getting to market first.
If we create the most and the best ideas in the industry, we will win.	If we make the best use of internal and external ideas, we will win.
We should control our IP, so that our competitors don't profit from our ideas.	We should profit from others' use of our IP, and we should buy others' IP whenever it advances our own business model.

Henry Chesbrough
Open Innovation: The New Imperative for Creating and Profiting from Technology
Harvard Business Review Press, 2005

Open Innovation 2.0



Independent Expert Group Report on Open Innovation & Knowledge Transfer
Directorate-General for Research and Innovation, 2014

- ❑ Innovation is seen as the outcome of a complex co-creation process involving knowledge flows across the entire economic and social environment.
- ❑ Dynamic, networked, multi-collaborative innovation ecosystems.
- ❑ The [current] premise of Open Innovation is to *open up* the innovation process to *all active players* so that knowledge can circulate more freely and be transformed into products and services that create new markets, fostering a stronger culture of entrepreneurship.

Need: **Knowledge exchange** and **absorptive capacities** from all actors involved, whether businesses, academia, financial institutions, public authorities or citizens.

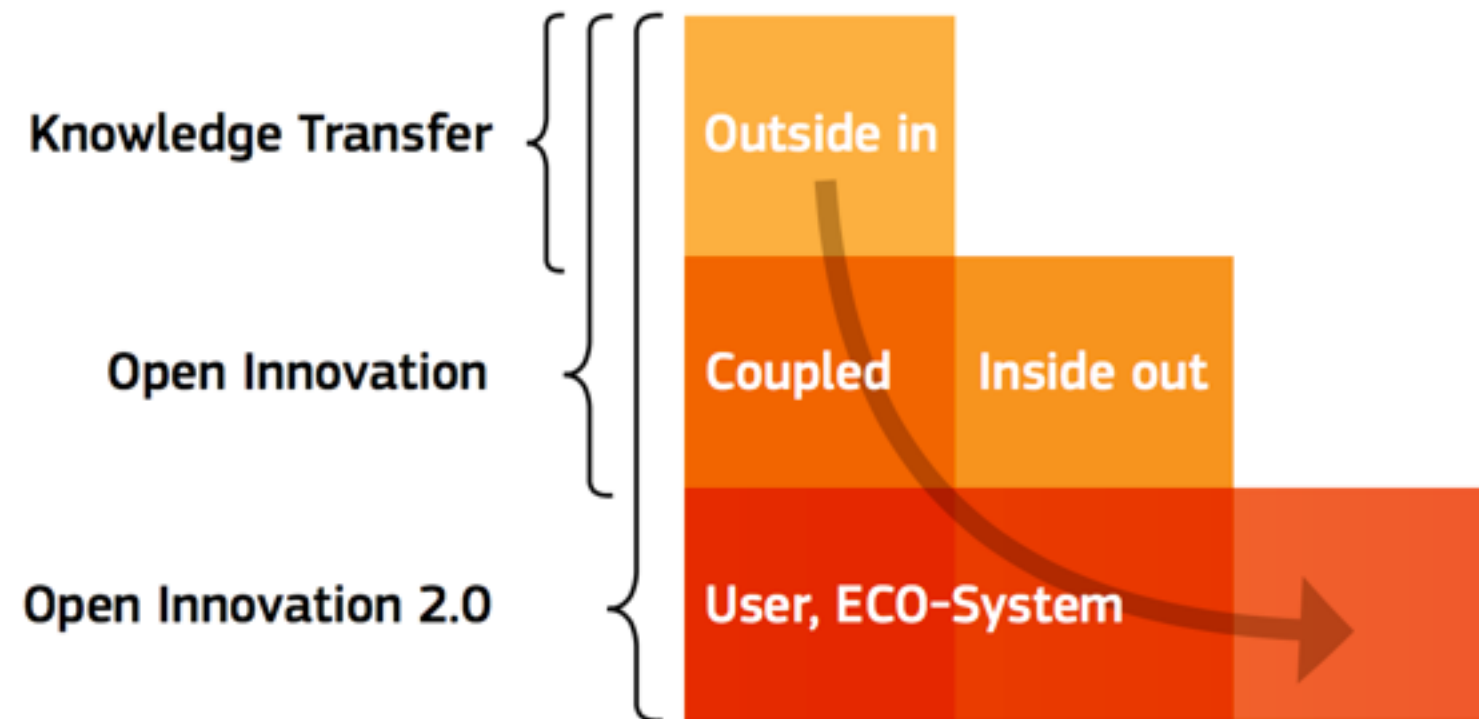
Open Innovation 2.0: two main elements

1. **Public Engagement**

The users are in the spotlight: an invention becomes an innovation only if users become a part of the value creation process.

2. **Ecosystem**

Creating a well-functioning eco-system that allows co-creation becomes essential for Open Innovation. In this eco-system relevant stakeholders are collaborating along and across industry and sector-specific value chains to co-create solutions to socio-economic and business challenges.



Open Innovation 2.0: actors and roles

THE PUBLIC SECTOR

The public sector has a central role to play in promoting Open Innovation. First and foremost it creates the regulatory environment in which all other actors operate. It puts in place rules and tools that can incentivise an open circulation of knowledge and cooperation among different actors with the aim to develop and market innovative solutions. Secondly, it offers better modes of coordination among the economic actors involved in order to enhance productivity and value. Thirdly, it can create a demand for innovation, both through the above-mentioned regulatory means and, for instance, through the procurement of innovative solutions.

THE FINANCIAL SECTOR

Innovation can be a risky business, therefore accessing funding and / or finance is not always easy for those who have innovative ideas. Building more innovation-friendly financial instruments and institutions and promoting the integration of existing funds and tools is essential to support Open Innovation. It is important that investors of all kinds find their interest in investing in innovation.

INNOVATIVE BUSINESSES

Businesses play a key role in innovating. In order to be able to bring innovations to the market, they must be able to maximise their returns on the resources allocated to innovating. This is the reason why it is important to reduce European market fragmentation, while fostering faster market access and development.

ACADEMIA

Universities, Higher Education Institutions, and Public Research Organisations / Research and Technology Organisations have a key role to play in the innovation eco-system, not only as knowledge producers, but also as co-creators and generators of skilled human capital. Challenges in this component of the eco-system include the co-creation capabilities of universities, the design of incentives for academics when working with users and the absorptive capacity of academic knowledge within firms.

CITIZENS

Citizens, users and Civil Society Organisations have a central and transversal role to play in bringing innovation to the market. They create a demand for innovative products and services, they can fund and / or finance projects that are relevant to them, they can be at the source of innovative ideas worth spreading and scaling up and they can have a say in what research is meaningful to them and can impact their lives.

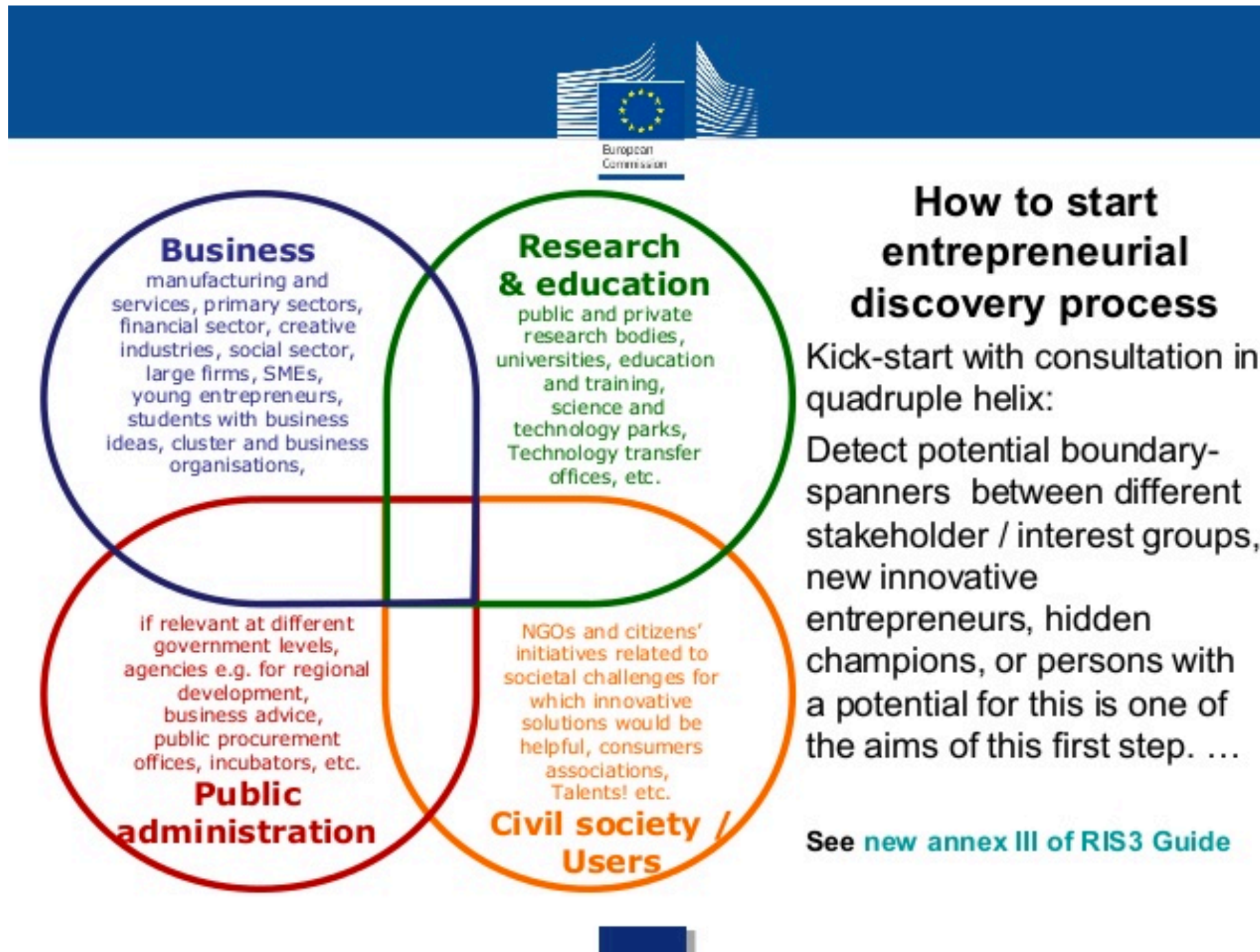
The *G*oal of an open innovation ecosystem

“The translation of **knowledge** into socio-economic value”

Why (Open) Science in Open Innovation ecosystems?


- ❑ Open Science represents a new approach to the scientific process based on cooperative work and new ways of diffusing knowledge by using digital technologies and new collaborative tools. The discussion on broadening the science base and on novel ways to produce and spread knowledge gradually evolved from two global trends: open access and open source.
- ❑ Open Science has the potential to strengthen and enhance science by facilitating more transparency, openness, networking and collaboration, and by fostering interdisciplinary research.
- ❑ "In addition to the formal supply side elements such as research skills, excellent science, funding and Intellectual Property management, there is also a need to concentrate on the demand side aspects of knowledge circulation, making sure that scientific work corresponds to the needs of the users and that knowledge is **findable, accessible, interpretable** and **re-usable** (FAIR)”.

‘Quadruple Helix’ approach to R&I Policy Making



Quadruple helixes do exist!

newsabout useventsnetworkmembers



European Network of Living Labs

Reflection on OpenLivingLab Days 2017

Between 29th of August and 1st September 300 participants from 31 different countries gathered in Krakow at the 8th edition of the ENoLL Summer School - the OpenLivingLab Days. For the first time, the event was hosted in Central & Eastern Europe -...

[Read more](#)

Login or Sign up!

NEWS

Small Advanced Economies Initiative Innovation day

SUBMITTED BY ENOLL ON WED, 2017-09-13 16:06

Reflection on OpenLivingLab Days 2017

SUBMITTED BY ENOLL ON TUE, 2017-09-12 13:43

CALL for independent external controller auditor (registered in Belgium) - TALIA Interreg MED project - EXTENDED DEADLINE

SUBMITTED BY ENOLL ON THU, 2017-08-17 10:02

At the Open Innovation 2.0 Conference


SUBMITTED BY ENOLL ON WED, 2017-07-05 10:51

ENoLL projects contributing to IoT week

SUBMITTED BY ENOLL ON WED, 2017-07-05 10:07


[Read more news](#)

IoT PRE-MARKET CONSULTATION



Influence our Open Call now


SPOTLIGHT SLIDES





European Network of Living Labs

TWITTER

Tweets by @openlivinglabs

 ENoLL Retweeted

 SocialcreativeMED @MEDSocialcreat
We are happy to announce the 1st Social&Creative Innovation Week (Oct. 2-5). Discover the networking opportunities! bit.ly/2w4esMb



From 2nd to 5th October 2017
Social&Creative Innovation Week

**Generalitat de Catalunya**

**Unió Europea**

Balanç del Primer Camp d'Innovació de CatLabs

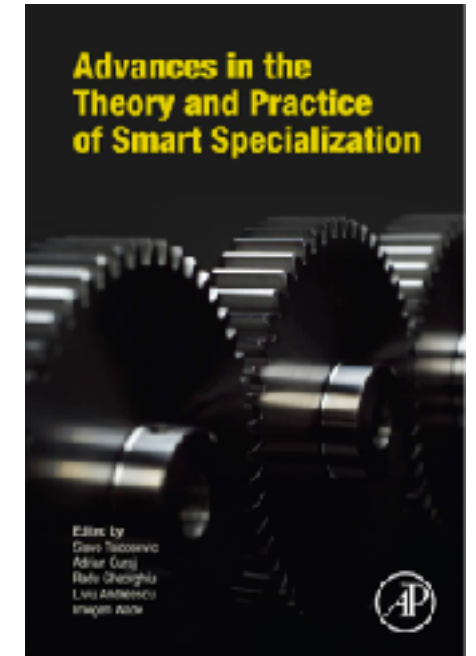
Tecnocampus-Mataró
(6, 7 i 8 de juny de 2017)



**CAT LABS**

RIS3

“Research and innovation strategies for smart specialisation”



“Smart specialization was seen as a way of increasing the effectiveness of spending on innovation and achieving the Europe 2020 objectives.”

Peter Berkowitz

Head of Unit G1 (Smart and Sustainable Growth)

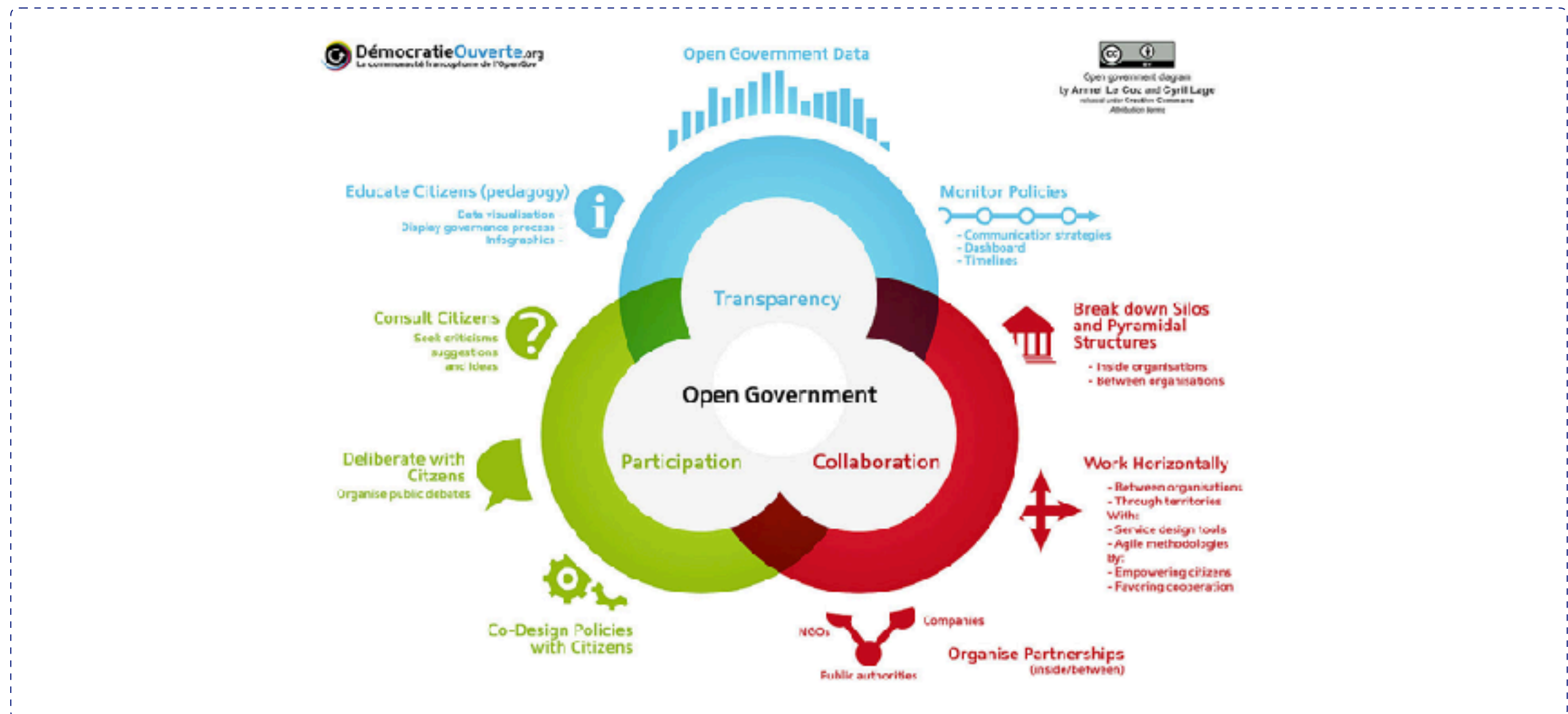
Directorate General for Regional and Urban Policy, European Commission

The simple intellectual project of smart specialisation strategies is that regions will build the capabilities that they need to achieve their own structural changes. A smart specialisation strategy thus has two facets (Foray et al., 2009) are as follows:

1. forming **capabilities** in a few strategic domains;
2. driving **structural changes**.

Smart specialisation involves businesses, research centres and universities working together to identify a Member State or region’s most promising areas of specialisation, but also the weaknesses that hamper innovation there

Techs @“Open Data Government”



- ❑ Foster **evidence- and data-driven policies** and context interpretations.
- ❑ Promote a culture of **accountability**, and **transparency**.
- ❑ Moving towards a ‘more scientific’ **impact analysis** of the investments.
- ❑ Create an effective **collaboration environment** for universities, research institutes, and firms (see, the EU vision on ‘Open Innovation and Open Science’)

Open Data Government: principles



- ❑ **Complete:** All public data are made available.
- ❑ **Primary:** Data are as collected at the source, with the highest possible level of granularity, not in aggregate or modified forms.
- ❑ **Timely:** Data are made available as quickly as necessary to preserve their value.
- ❑ **Accessible:** Data are available to the widest range of users for the widest range of purposes.
- ❑ **Machine processable:** Data are reasonably structured to allow automated processing.
- ❑ **Non-discriminatory:** Data are available to anyone, with no requirement of registration.
- ❑ **Non-proprietary:** Data are available in a format over which no entity has exclusive control.
- ❑ **License-free:** Data are not subject to any copyright, patent, trademark or trade secret regulation.

December 7-8, 2007, Thirty open government advocates
@Sebastopol, California

The technologies behind the design and the implementation of support platforms have to be open, capable of accommodating new strategic demands, new uses and new data sources, both internal and external.

Be FAIR!

Supporting knowledge discovery through good data management

[F] To be Findable:

- F1. (meta)data are assigned a globally unique and eternally persistent identifier.
- F2. data are described with rich metadata.
- F3. (meta)data are registered or indexed in a searchable resource.
- F4. metadata specify the data identifier.

[A] To be Accessible:

- A1 (meta)data are retrievable by their identifier using a standardized communications protocol.
 - A1.1 the protocol is open, free, and universally implementable.
 - A1.2 the protocol allows for an authentication and authorization procedure, where necessary.
- A2 metadata are accessible, even when the data are no longer available.

[I] To be Interoperable:

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles.
- I3. (meta)data include qualified references to other (meta)data.

[R] To be Re-usable:

- R1. meta(data) have a plurality of accurate and relevant attributes.
 - R1.1. (meta)data are released with a clear and accessible data usage license.
 - R1.2. (meta)data are associated with their provenance.
 - R1.3. (meta)data meet domain-relevant community standards.

“Digital Agenda for Europe”, May 2010

<https://data.europa.eu/euodp/en/linked-data>

<http://ec.europa.eu/digital-agenda/>



Linked Open Data is the standard to represent data on a wide range of topics which makes it easier for developers to connect information from different sources, resulting in new and innovative applications. Linked Open Data enables a "browsing" or "discovery" approach to finding information, as compared to the usual "search" practice.

The formal languages behind the concrete realisation of a Linked Open Data initiative are:

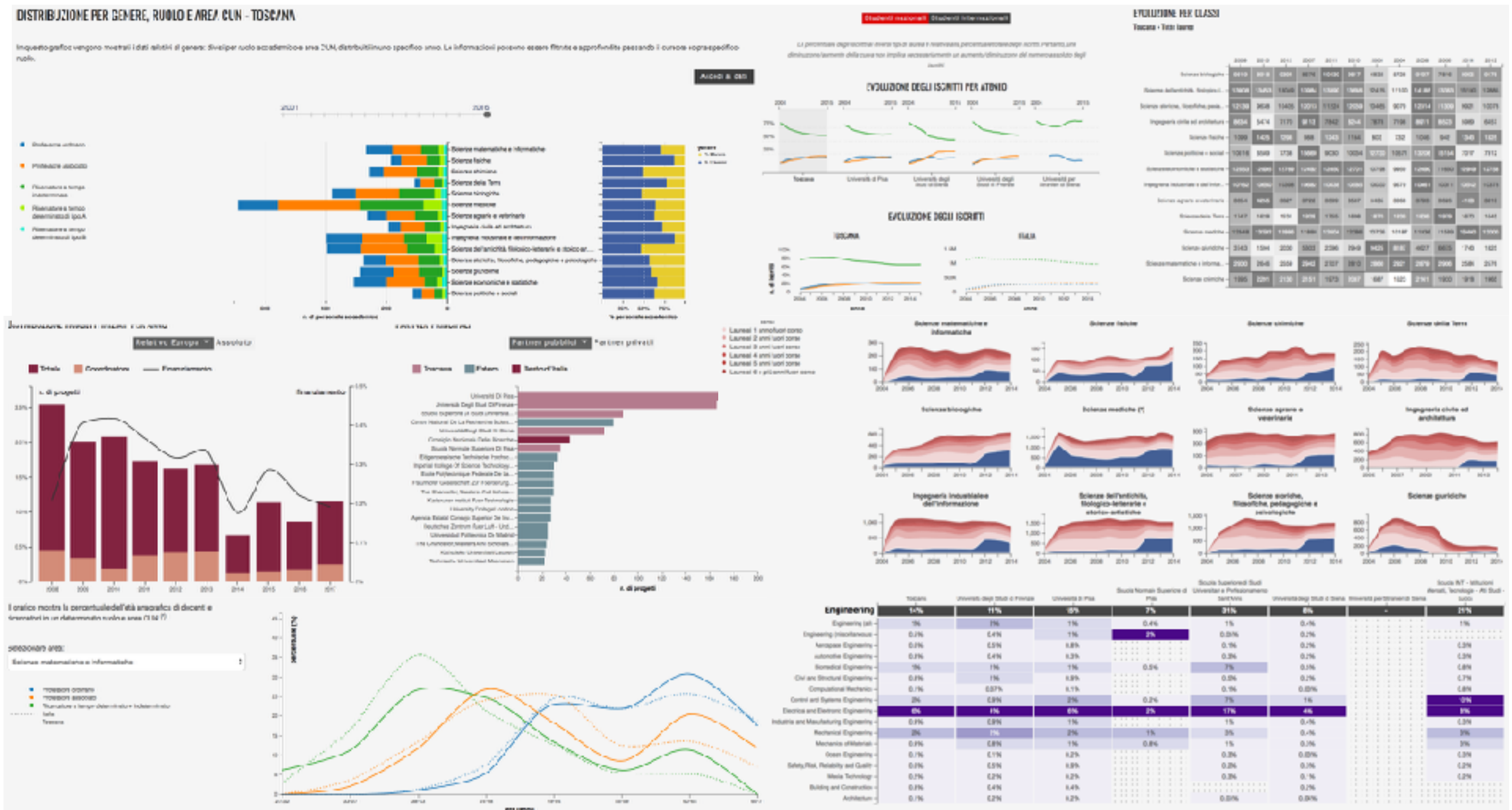
- ❑ **RDF** (“Resource Description Framework”): the flexible data model based upon the idea of making statements about resources in the form of subject–predicate–object expressions, known as triples;
- ❑ **RDFS/OWL2** (“Resource Description Framework Schema”/“Web Ontology Language”): the schema and ontology languages for describing concepts and relationships;
- ❑ **SPARQL** (“SPARQL Protocol and RDF Query Language”): the query language;
- ❑ **RIF** (“Rule Interchange Format”): a rules language originally designed to exchange rules among different existing rules dialects;
- ❑ **RDFa** (“Resource Description Framework in Attributes”): the language for marking up data inside HTML-based Web pages;
- ❑ **HTTP** communication protocol (“Hypertext Transfer Protocol”): the application protocol for distributed, collaborative, and hypermedia information systems, at the foundations of the so-called World Wide Web.

<https://data.europa.eu/euodp/en/linked-data>

username/password: dao2017/dao2017

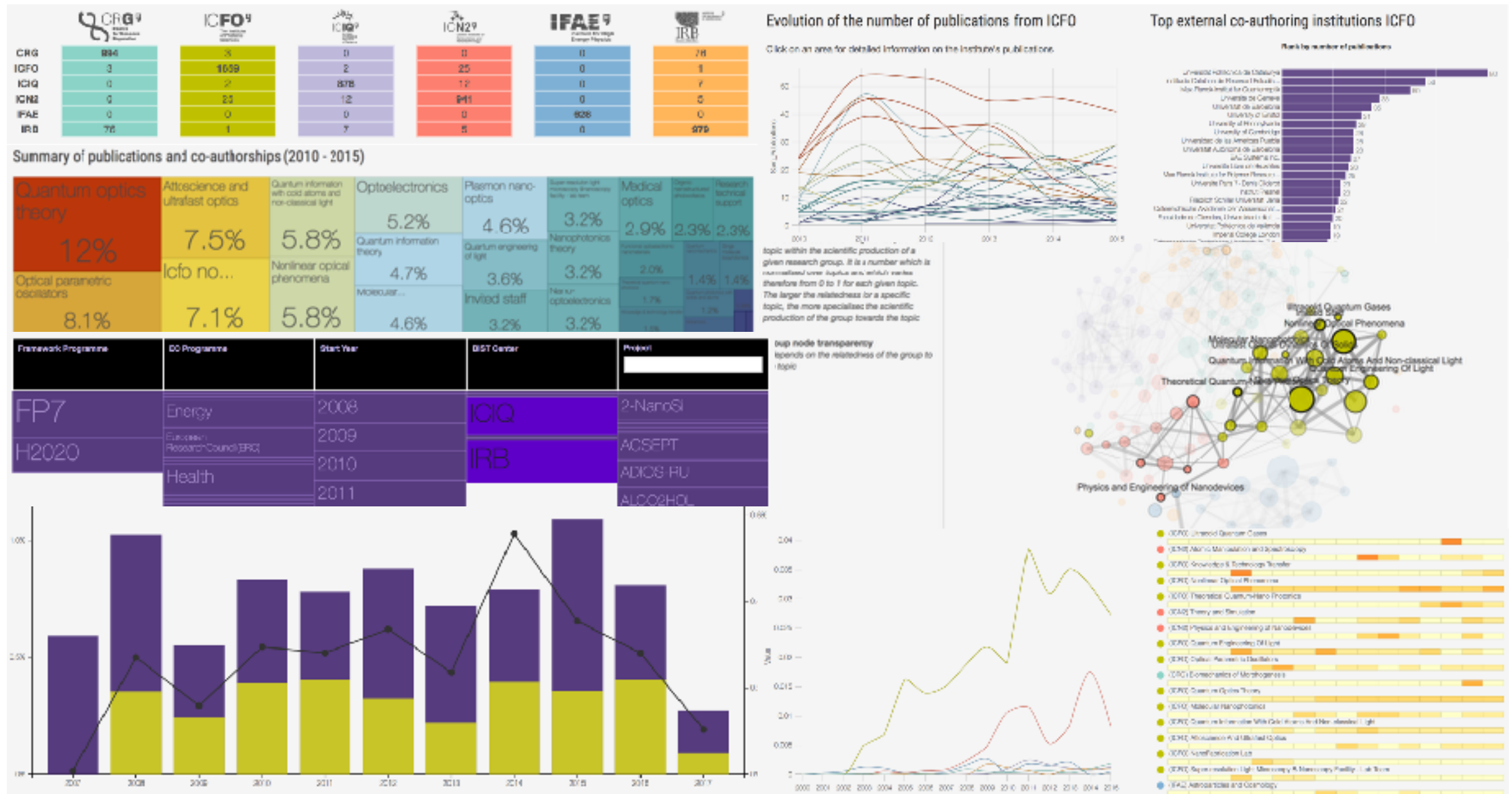
External communication: Valorise the Tuscan R&I system.

Internal Policy making: Support for the implementation of evidence-driven policy making.



BIST: Barcelona Institute of Science and Technology

BIST launched in October 2016 the development of an interactive dashboard with information on the aggregate activity of its six institutes, to: (i) improve the **self-knowledge** of this new institution, (ii) measure its potential for **scientific collaboration**, (iii) provide a **continuous information service** to its six research centres.

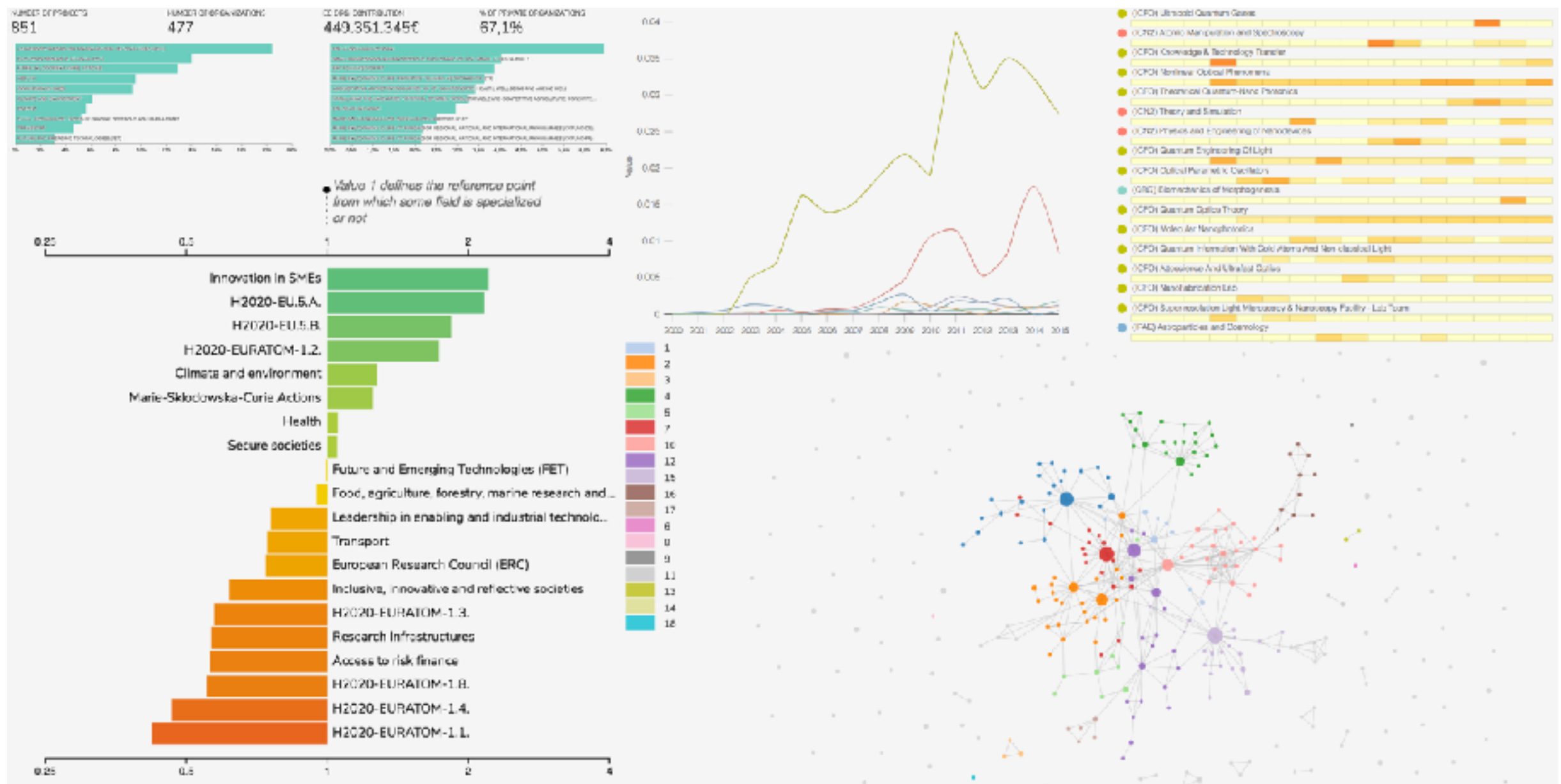


RIS3CAT Monitoring @Generalitat de Catalunya

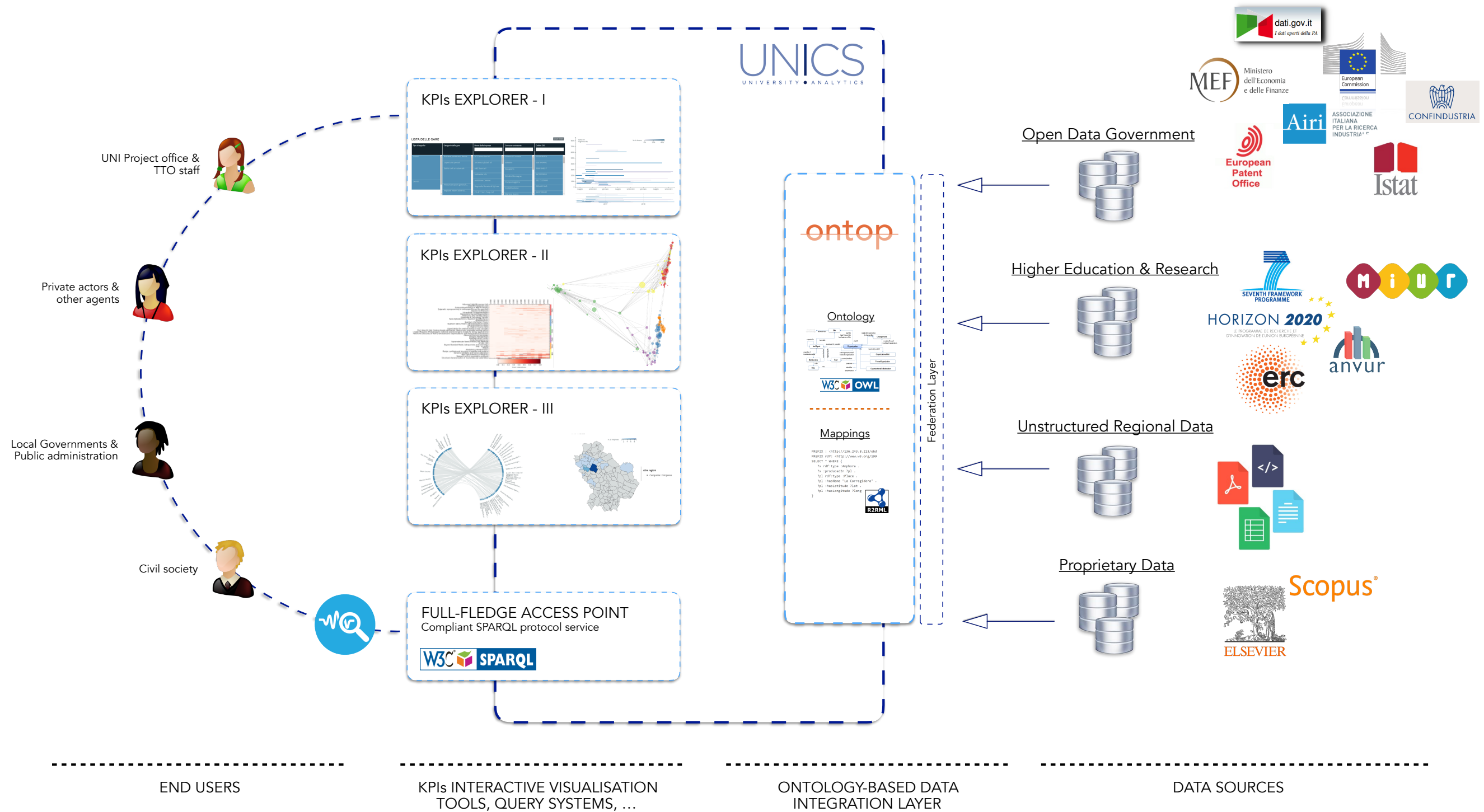
Connect and render interoperable public datasets about **research and innovation policies** supported with European funding (H2020, FP7 and ERDF).

Map the activities and **collaboration links of public and private stakeholders**.

Provide **open access** to the compiled data.



OBDA/I: A proposal of architecture



Go beyond the pre-compiled interactive views

[illegible]

The (current?...) UNiCS datasets

Fully integrated (SPARQL accessible)

European/World data:

- Educational Institutions
- Rankings (ARWU,THE,QS,NTU) (Not open-data)
- CORDIS projects (FP7,H2020)
- EUROSTAT (selected datasets)
- World countries

Italian data:

- Universities budget
- Scholars (Aggregated data)
- Students (Aggregated data)
- PRIN Projects (Aggregated data)

Stored in database (not mapped)

European/World data:

- ETER Database
(Universities Budget, Staff, Students,PhDs 2012-2015)

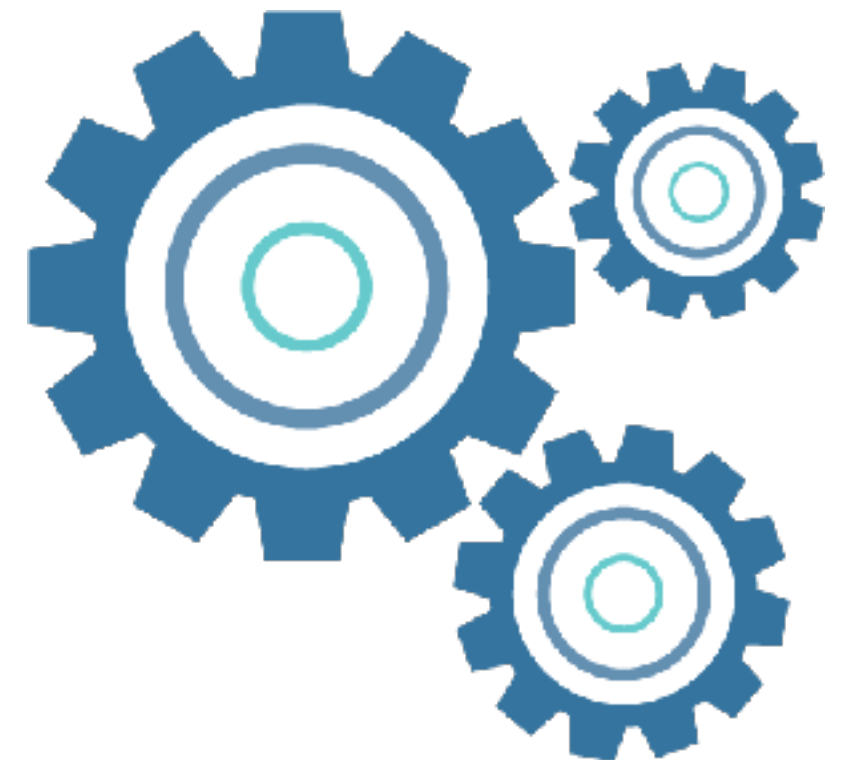
Not in database

Italian data:

- Rankings IISole24Ore, CENSIS
- VQR Indicators

Internal use only

- GRID database (World institutions)
- NUTS conversion dataset (postal code, lau code)



- owl:Thing
 - :Affiliation-Reified
 - :BIST-staff-Reified
 - :Budget-Reified
 - :Collection
 - :Dataset
 - :DataVoc
 - :ActivityType
 - :Budget-Voc
 - :Eurostat-field
 - :Gender
 - :MetricCategory
 - :MIUR-Citizenship
 - :PRIN-Voc
 - :Sorting
 - :UnityOfMeasure
 - :VQR-field
 - :DisciplinaryClassification
 - :ANVUR-GEV
 - :BIST-Classification
 - :EC-Classification
 - :MIUR-Classification
 - :PatentClassification
 - :SCOPUS-Classification
 - :WOS-Category
 - :EC-ParticipantContribution-Reified
 - :Eurostat-Reified
 - :Evaluation
 - :Item
 - :ListItem
 - :Metric
 - :MIUR-Reified
 - :MIUR-ScholarAggregate
 - :MIUR-StudentAggregate
 - :PersonRole-Reified
 - :PRIN-Reified
 - :Project
 - :EC-Project
 - :ResearchProject
 - :Ranking
 - :RankingValue
 - :PositionValue
 - :ScoreValue
 - :ResearchOutput
 - :BibliographicEntity
 - :Patent
 - :ScientificPublication-Reified
 - :Role
 - :TemporalEntity
 - :TerritorialUnit
 - :UnitOfMeasure
 - :VQR-Reified
 - :WorksFor-Reified
 - dc:Agent
 - :Organization
 - :Person

- owl:topObjectProperty
 - :academicRank
 - :activityType
 - :affiliation-reified
 - :agent
 - :anvurGev
 - :bistClassification
 - :budgetDescriptor
 - :budgetField
 - :composedOf
 - :creator
 - :dataSource
 - :degreeClass
 - :ecClassification
 - :ecParticipantContribution
 - :ecProject
 - :element
 - :elementOf
 - :eurostat
 - :eurostat-field
 - :evaluation
 - :followedBy
 - :fundingAgency
 - :gender
 - :geoCoverage
 - :globalMetric
 - :inUse
 - :italianStudent
 - :item
 - :itemContent
 - :itemContentOf
 - :itemOf
 - :location
 - :mapsTo
 - :metric
 - :metricCategory
 - :metricScore
 - :miurCitizenship
 - :miurClassification
 - :partOf
 - :patentApplicant
 - :patentInventor
 - :personRole-ref
 - :personRole-reified
 - :positionValue
 - :precededBy
 - :prinDescriptor
 - :prinField
 - :publisher
 - :ranking
 - :researchOutput
 - :role
 - :roleContext
 - :roleIn
 - :scopusClassification
 - :scoreValue
 - :sorting
 - :unitOfMeasure

- owl:topDataProperty
 - :abstract
 - :date
 - :birthdate
 - :endingDate
 - :publicationDate
 - :startDate
 - :year
 - :birthyear
 - :publicationYear
 - :description
 - :ecContribution
 - :ecMaxContribution
 - :ecObjective
 - :ecRef
 - :endingPage
 - :error
 - :geoLat
 - :geoLong
 - :identifier
 - :alpha2
 - :alpha3
 - :cooperativeClassificationLabel
 - :cordisRef
 - :DOI
 - :grantAgreementNo
 - :InternationalClassificationLabel
 - :ISSN
 - :orcidId
 - :patentApplicationNo
 - :patentPriorityNo
 - :patentPublishedAs
 - :picNumber
 - :scopusIdentifier
 - :scopusAuthorId
 - :scopusOrganisationId
 - :scopusPaperId
 - :scopusSourceId
 - :scopusSubjectAreaCode
 - :WoS-id
 - :index
 - :name
 - :acronym
 - :alternativeName
 - :editionLabel
 - :extendedName
 - :firstName
 - :lastName
 - :shortName
 - :numberOfProjects
 - :pageRange
 - :postalCode
 - :publisher
 - :rangeEnd
 - :rangeStart
 - :rank
 - :size
 - :startingPage





Open Innovation, Open Science, and a few, related...
Personally perceived, 'OBDA Open Issues'

1. Virtual integration over multiple dataset formats (RDF triple store, RDB, no-SQL, for instance)
2. Semi-automatic generation and testing of mappings
3. Statistical data ('RDF Data Cube' is not a solution)
4. Access control management, distributed ontologies and mappings
5. Data publication and interoperability (ETL processes should be dismissed)
6. Data quality (homogeneity, redundancy, inconsistency checks, incompleteness, etc.)
7. Classifications and conceptualisations alignment
8. Proper ontologies specification (should it be more than translating a UML CD?): costs and benefits
9. Ontology understanding and use by non-technical users (visualisation?)
10. Education and training (please, give me back the 'man in the middle')

“Science more generally has a critical role across many areas of policy in providing evidence that helps understand the risks and benefits of different policy choices.”

Jean-Claude Juncker

President of the European Commission

