

Data4RI Policy

Pavel Zbornik – DG Research & Innovation

Data4RI Policy Initiative

- Better evidence based policy making in Research and Innovation (R&I) requires more and better data.
- The digitalisation of science and innovation policy has improved our ability to operate large, linked dataset at a very detailed level, complementing already available macro data analyses.
- Several European countries have already adopted important initiatives to enrich their R&I administrative datasets for R&I policy design and evaluation.
- The European Commission has also embarked in building and enriching its Framework Programme R&I related dataset.

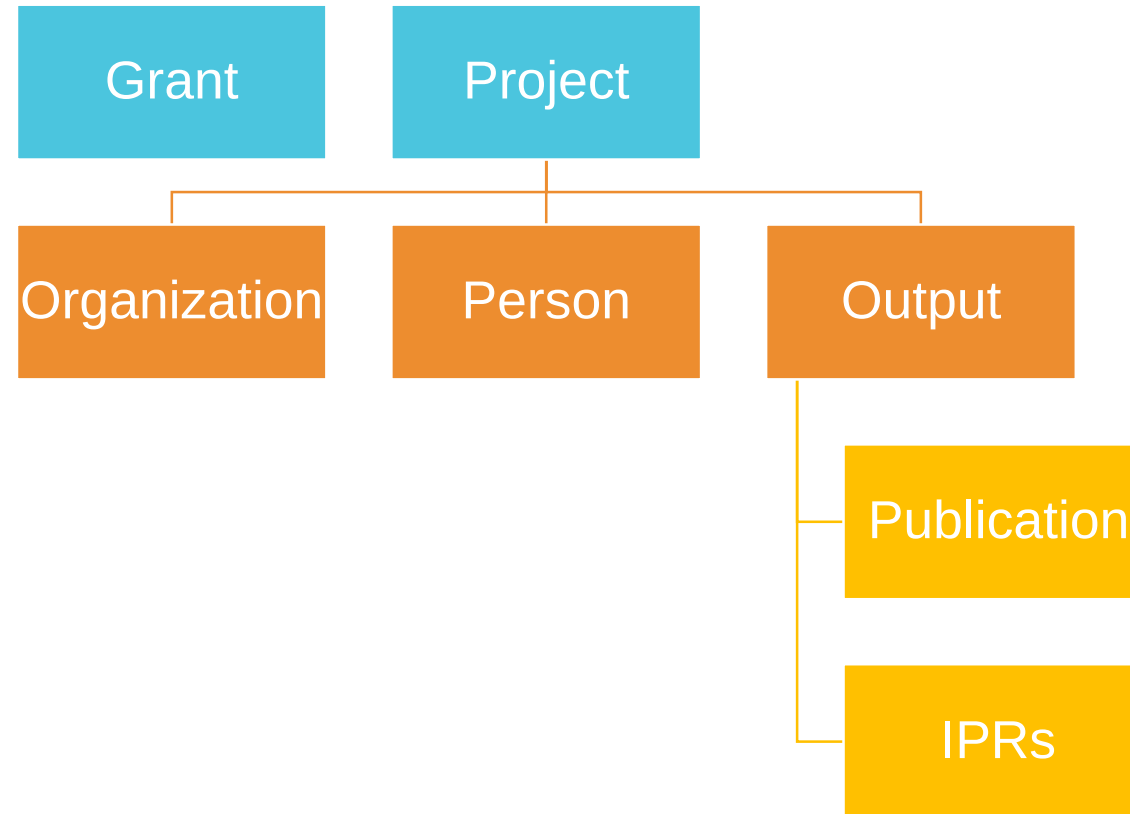
Data4RI Policy Initiative

The European Data for Research and Innovation Policy Initiative aims at fostering the **sharing** of Research and Innovation project **micro-data** at European scale, with mostly as objective to ease and refine data-informed policymaking in the ERA context.

Data4RI Policy Initiative – Policy questions

- *Directionality - Strategic investments*
- *Connectivity - Knowledge flows/ Collaboration Networks*
- *Inclusiveness - Comparability and complementarity of results*

Data focus



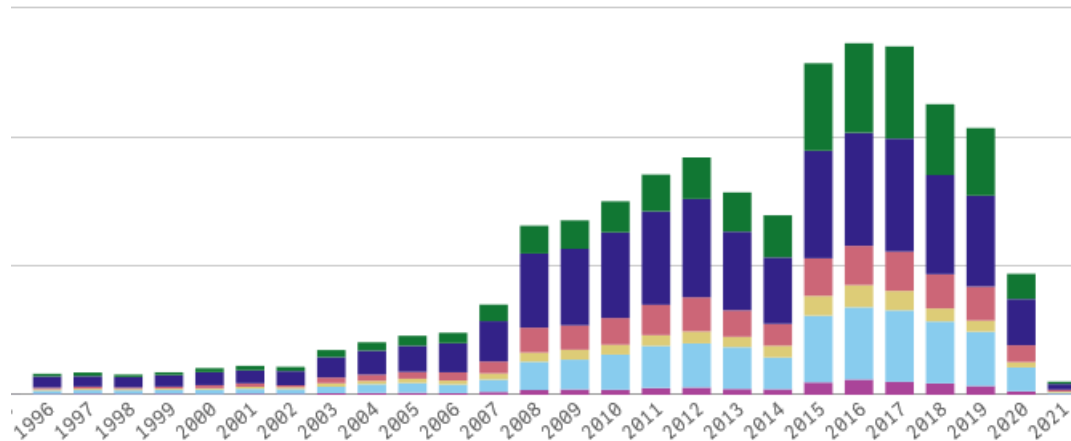
Data4RI Policy Initiative

Prioritized interest and shared views from MS workshop:

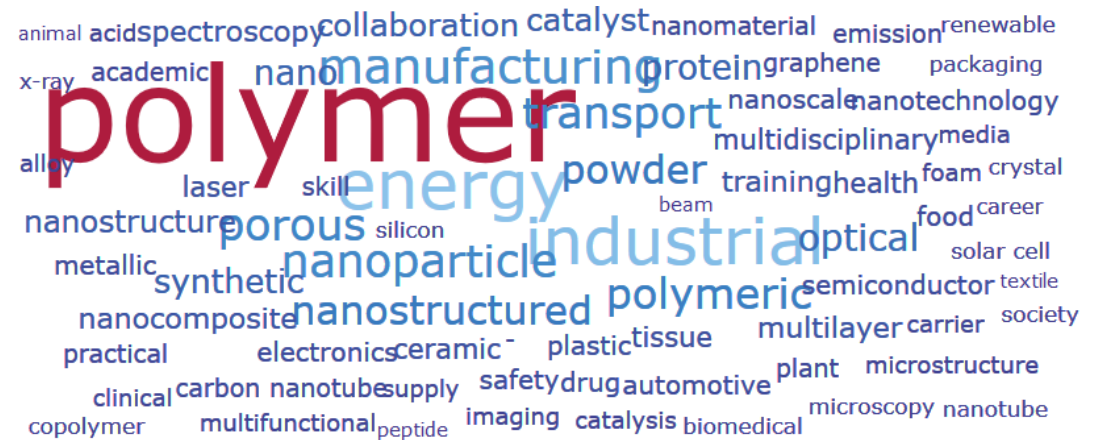
- 1) **Impact evaluation of R&I Investments**
- 2) **Directionality and complementarity of R&I Investments** at regional, national and European levels:
- 3) **Networks of knowledge and excellence in Europe**
- 4) Use of **common identifiers** for organizations, projects, researchers
- 5) **Classification and tools**
- 6) Use of **advanced techniques to increase DQ and interoperability, integrate multiple data sources**

Funding Directionality / Complementarity

Time Evolution



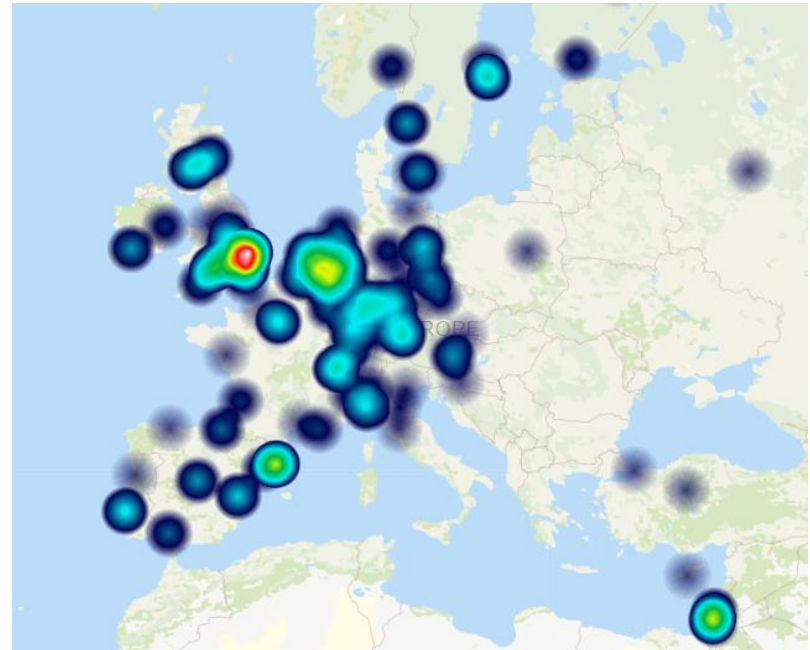
Keyword cluster



Scientific Production

Data Needs Outputs

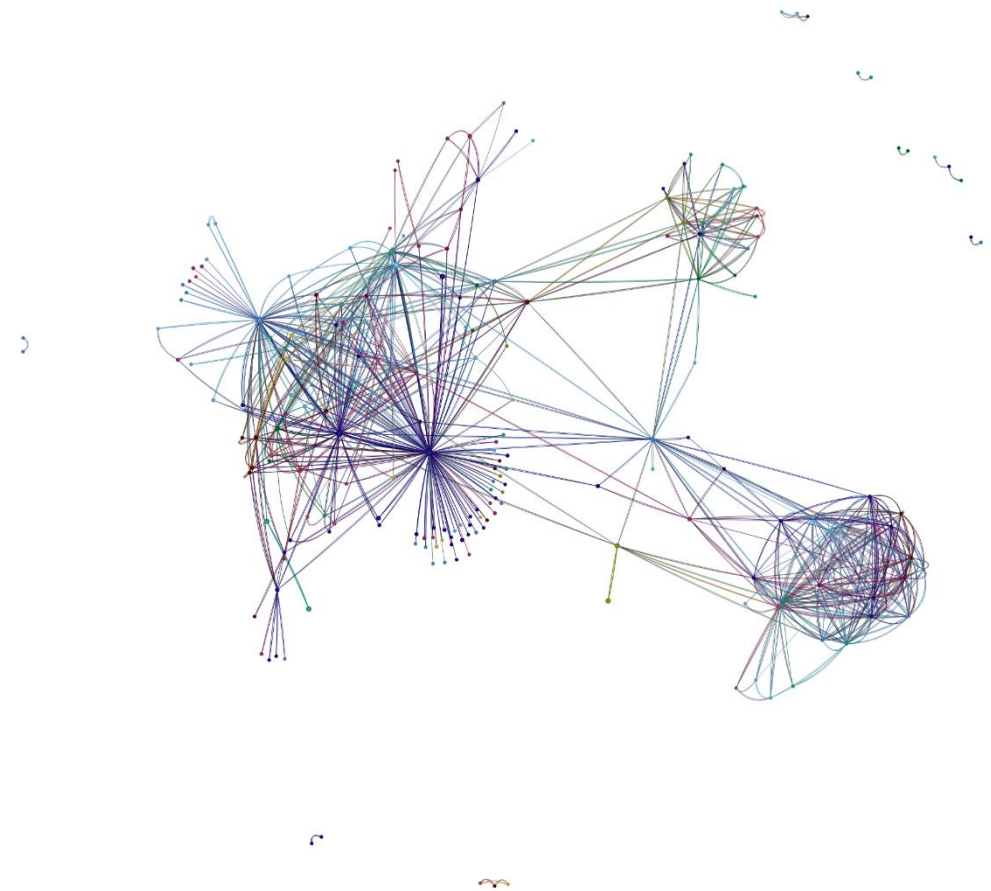
- Publication DOI
- Patent Application
- Organization Affiliation



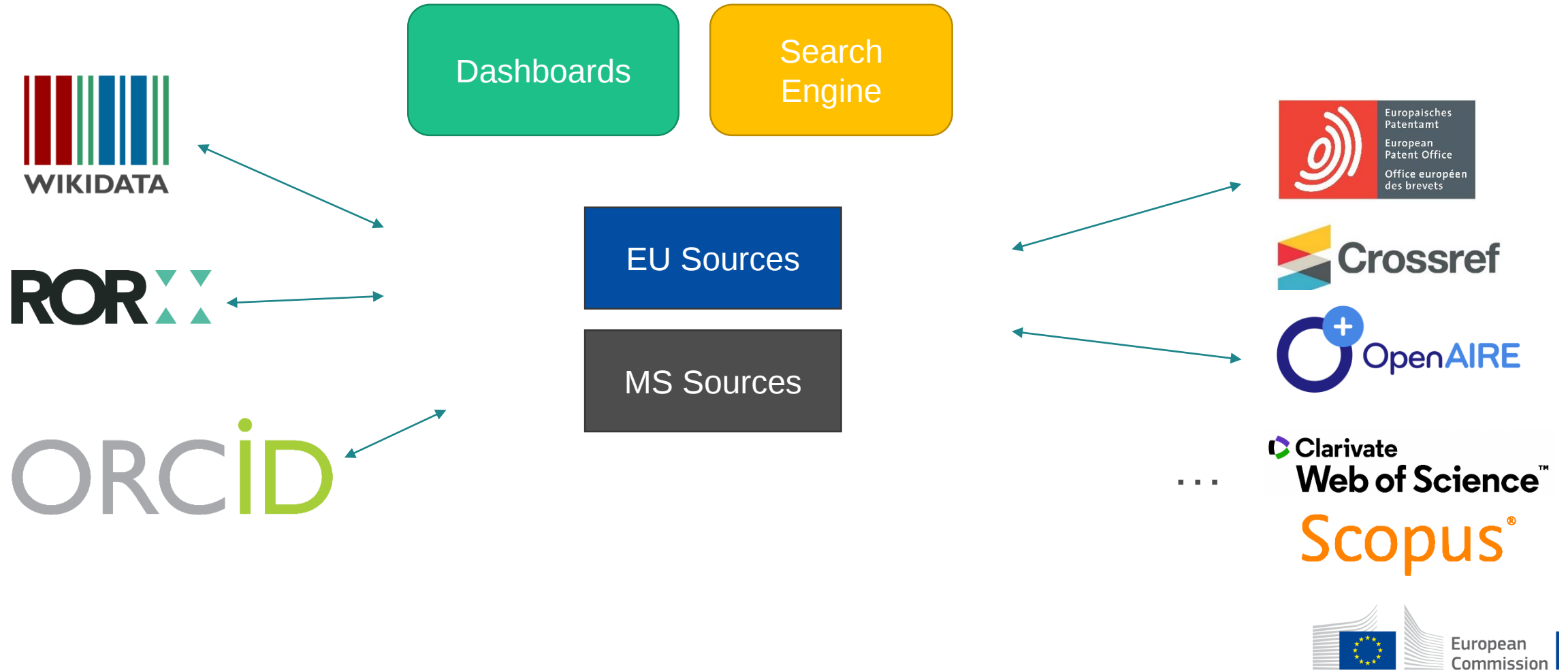
Collaborative Networks

Data Needs Organization

- Name
- Identifier
- Project role
- Location



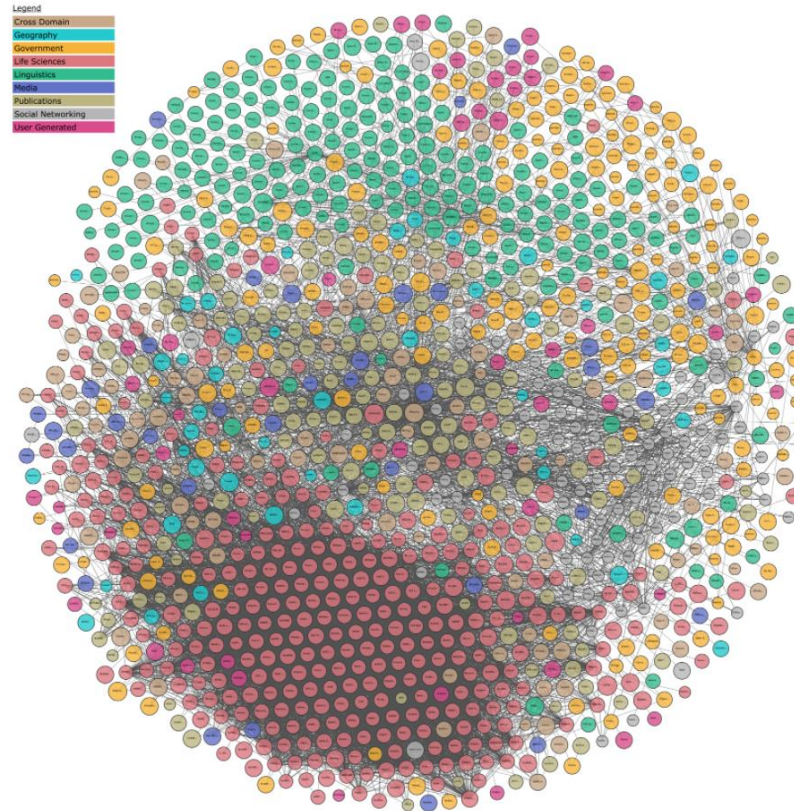
Vision



Can Open Data help?

- 31 sources identified, not from all countries in EU
 - **4** ☆ Data is available on the Web, in whatever format.
 - ☆☆ Available as machine-readable structured data, (i.e., not a scanned image).
 - **26*** ☆☆☆ Available in a non-proprietary format, (i.e, CSV, not Microsoft Excel).
 - ☆☆☆☆ Published using open standards from the W3C (RDF and SPARQL).
 - **1*** ☆☆☆☆☆ All of the above and links to other Linked Open Data.

Linked Open Data Cloud as a solution?



<https://lod-cloud.net/versions/2020-05-20/lod-cloud.png>

grantDOI or RAiD

- grantDOI or RAiD
 - with funding classification could be possible future option
- Complemented with wide spread of PIDs as ROR, ORCID and output DOIs
- Horizon Europe projects will get grantDOI in 2022

Data Process

DATA



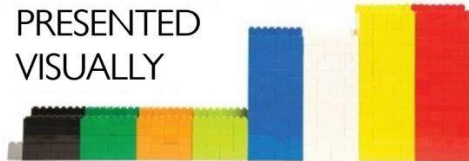
SORTED



ARRANGED



PRESENTED
VISUALLY



EXPLAINED
WITH A STORY



- Collect the data from CRIS/funder
- Harmonize in common structure
- Link to the external references
- Translate & Classify
- Resolve DQ issues
- Visualize
- Support R&I Policy

Effort

- Data Visualization
- Classification & Linking
- Harmonization in common structure
- Processing of raw data
- Loading of raw data



Challenges

- Source format variety
 - json, xml, csv, xlsx
- Duplication
- Language
- Use common PIDs

Data scope

Data availability

Summary of Data Availability

- Project

- Title 100%
- Funding 50%
- Abstracts 60%
- Start/End Dates 80%
- Coordinators 50%
- Consortium 30%
- Researcher(s) 30%

- Organizations

(if indicated – 50%)

- Name 100%
- PID Identifier(s) 10%
- Location 50%

Summary of Data Availability

- Project Outputs

(if indicated – 40%)

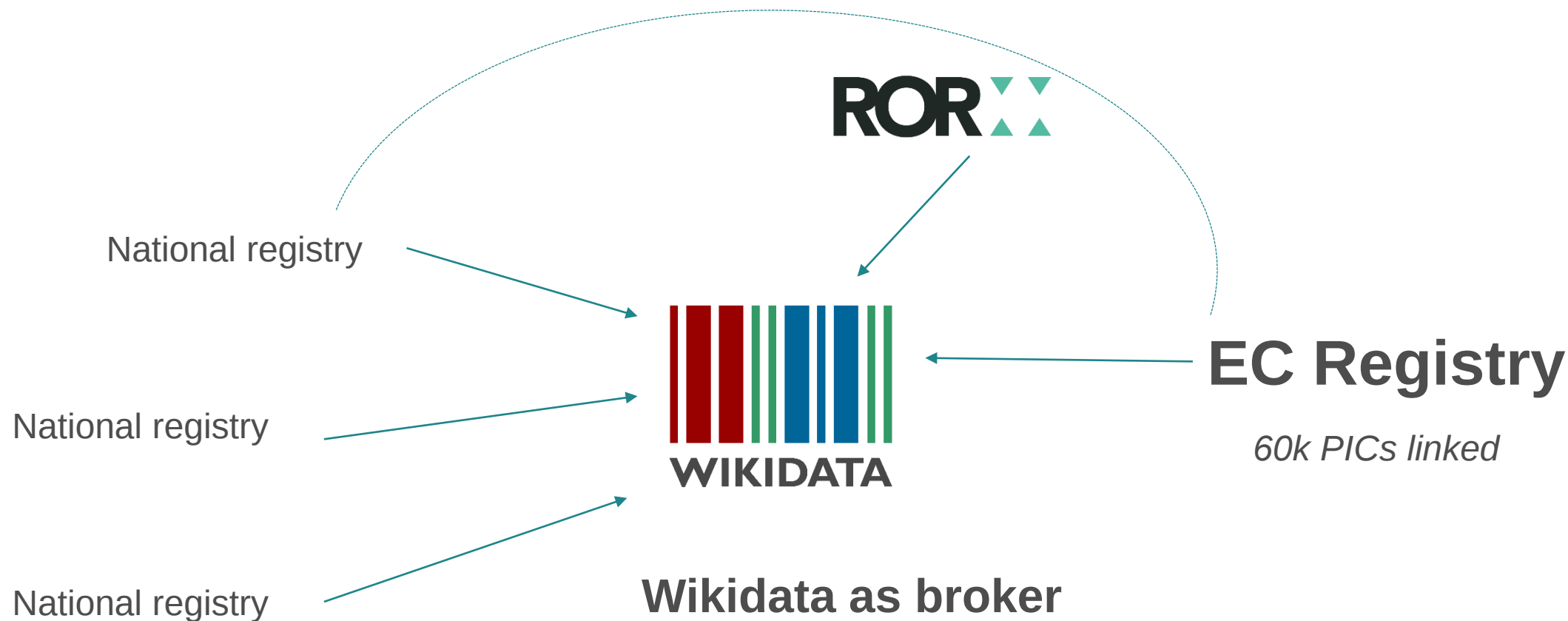
- Publications 100%
- IPRs 50%
- Further Funding 2%
- Datasets 2%

- Researcher(s)

(if indicated – 30%)

- Name 100%
- PID Identifier(s) 20%
- Affiliation 90%

Interoperability – Organizations PIDs



Interoperability - Directionality

<https://op.europa.eu/en/web/eu-vocabularies/euroscivoc>

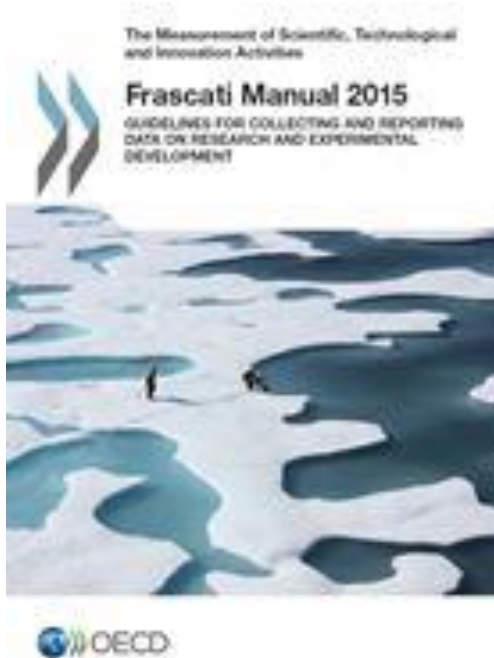
- Abstract classification by EuroSciVoc

EU Vocabularies

[Home](#) [Controlled vocabularies](#) [Models](#) [Business collection](#)

[Home](#) > [EU Vocabularies](#) > [Controlled vocabularies](#) > [Taxonomies](#) > [EuroSciVoc](#)

[Permanent link](#)

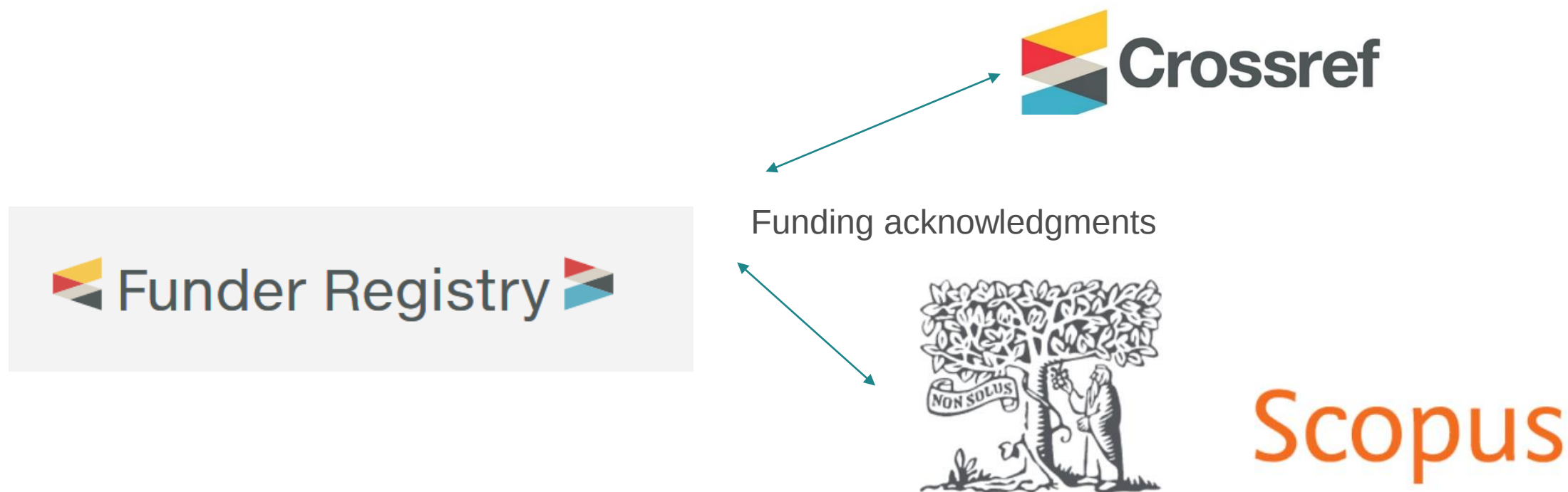


[Tree view](#) [Table view](#) [List view](#)

Filter by:

- [+ agricultural sciences](#)
- [+ engineering and technology](#)
- [+ humanities](#)
- [+ medical and health sciences](#)
- [+ natural sciences](#)
- [+ social sciences](#)

Interoperability – Funders PIDs



Building the bridges to identify outputs

State of the play

- 31 sources ingested *
 - 5.3 million projects (200+ funders)
 - 400 thousand organizations
 - 750 thousand researchers
 - 11 million outputs

*Raw data not unique

AL-E-CRIS

EE-ETIS

NO-CRISTIN

AT-FFG

ES-FECYT

OpenAIRE

AT-FWF

EU-Cordis

PL-RADON

BG-BulCRIS

EU-Kohesio

Re-searchAlps

BH-E-CRIS.BH

FI-ResearchFI

RS-E-CRIS

BH-E-CRIS.RS

FR-scanR

SE-SWECRIS

CH-ARAMIS

HR-HRZZ

SI-SICRIS

CH-P3

HU-NFKI

UK-UKRI-GTR

CZ-IsVaVAI

HU-NRDI

XK-E-CRIS

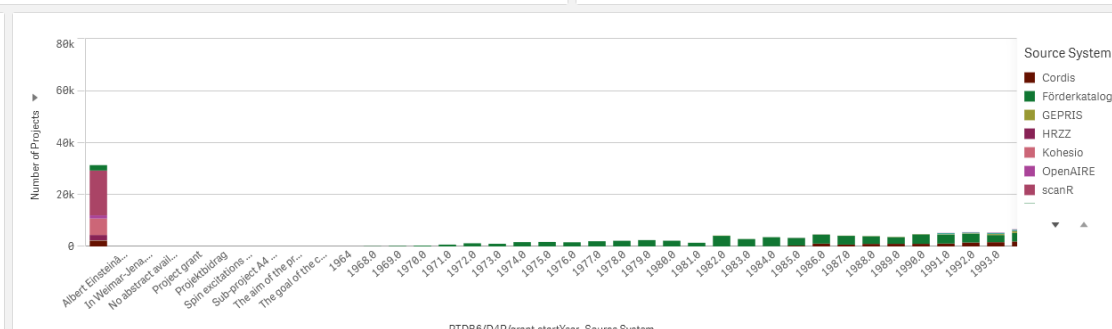
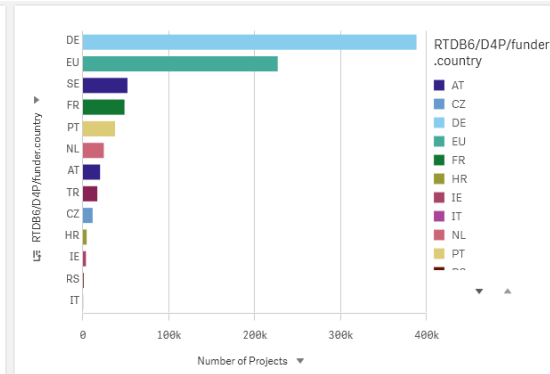
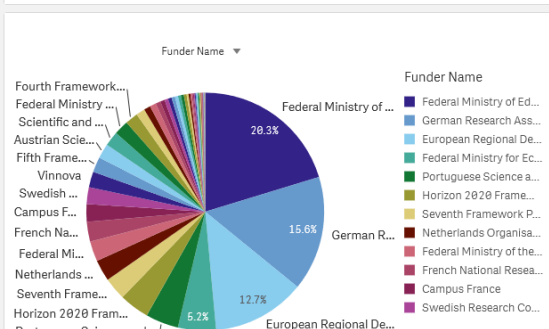
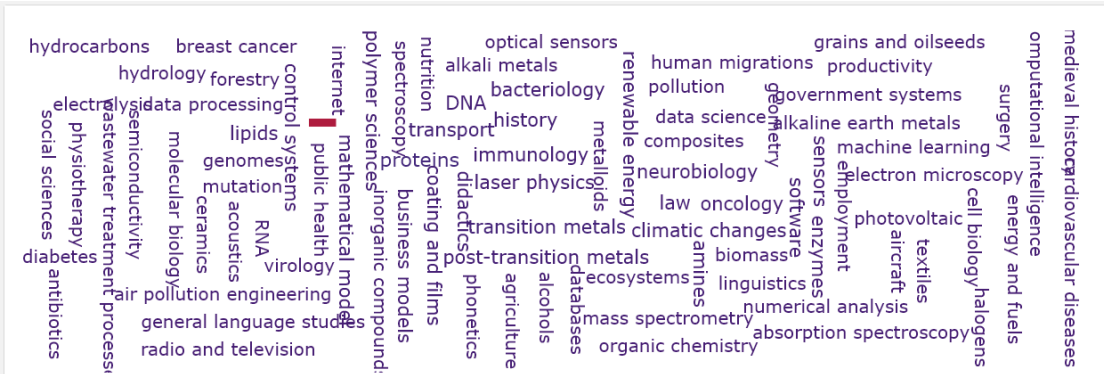
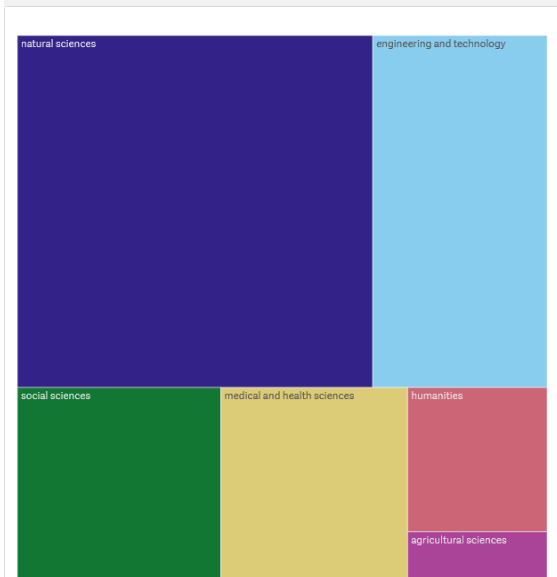
DE-FoKat

LV-ScienceLatvia

DE-GEPRIS

MK-E-CRIS

Number of Projects
1,17M



Next Steps

- Presentation on EU Open Data Days on 23th November
- Engage with MS/AS to analyze the landscape and needs -> Build a community
- Co-create the initiative
- Process the data in tiered approach
 - Projects – Organizations – Outputs - Persons
- Have first outcomes in 2022 as open data

Thank you

Use of PIDs in CRIS systems?

Adoption of grant DOIs?

Output standardization of CERIF based systems? As Open Data?

Use of authority tables, taxonomies and controlled vocabularies?