@openaire_eu

# The OpenAIRE Research Graph

## Supporting Evidence-Based R&I indicators

Claudio Atzori
CNR

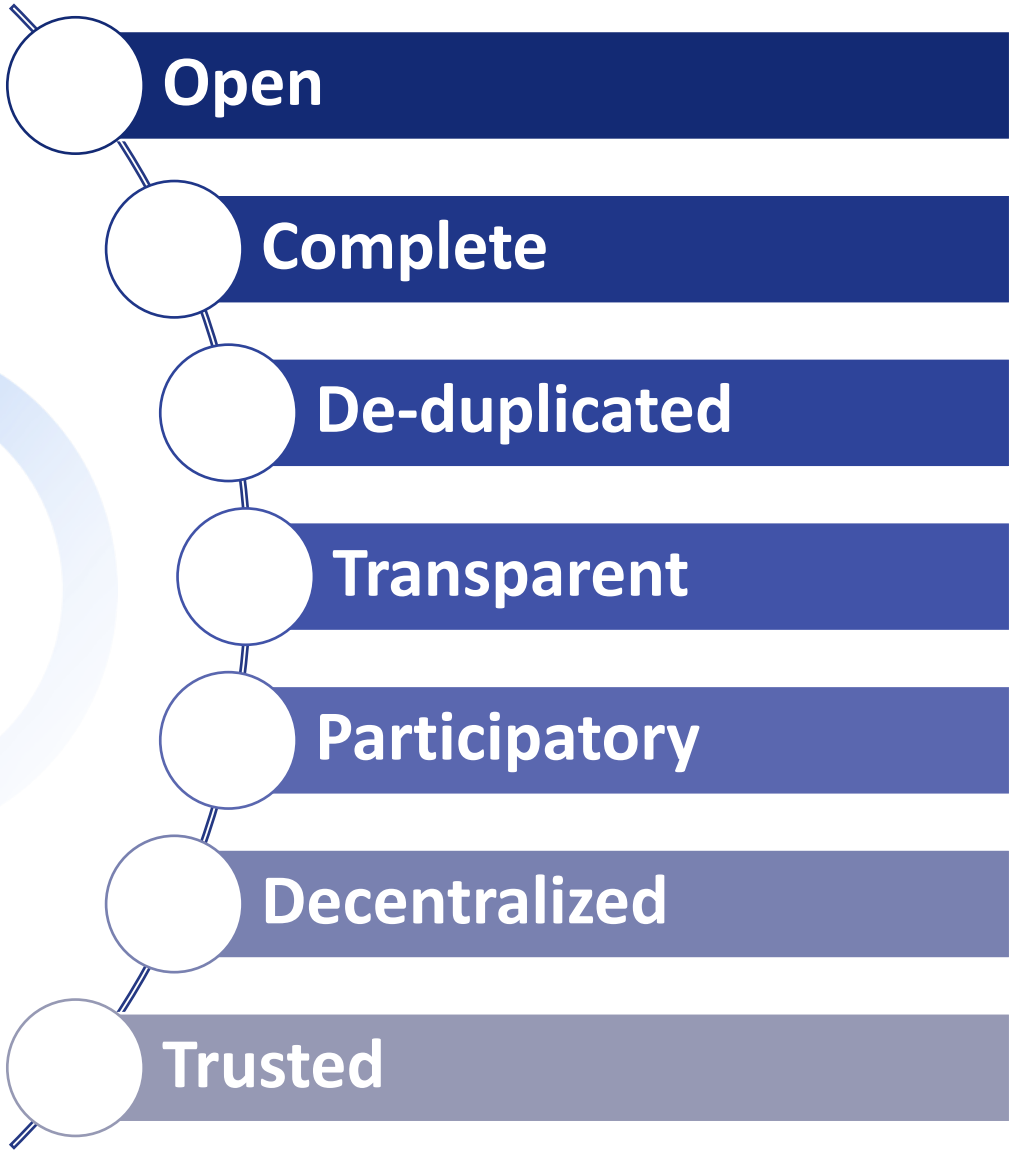
Ioanna Grypari
ATHENA Research Center

# Outline

1. The OpenAIRE Research Graph
2. IntelComp: evidence-based R&I Policy Making

# OpenAIRE Research Graph

*putting research into context, making the connections*




OpenAIRE ResearchGraph

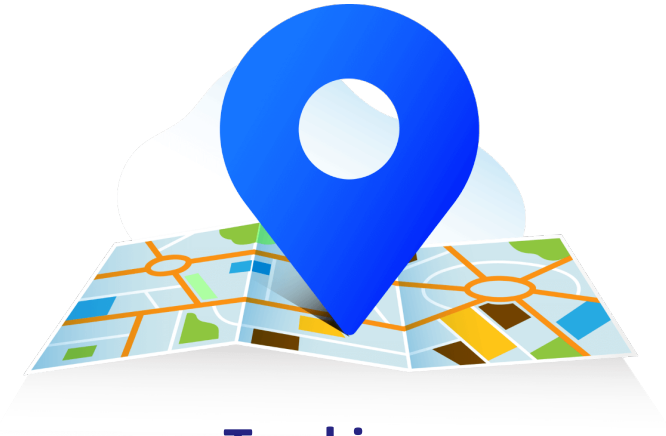
# What is it?

"A collection of metadata describing **objects** in the research lifecycle and **relationships** among them"



CC BY

# The OpenAIRE Research Graph: why?



## Tracking open research

**Reproducibility** and transparency require tracking of all outcomes of **science and related "context"**

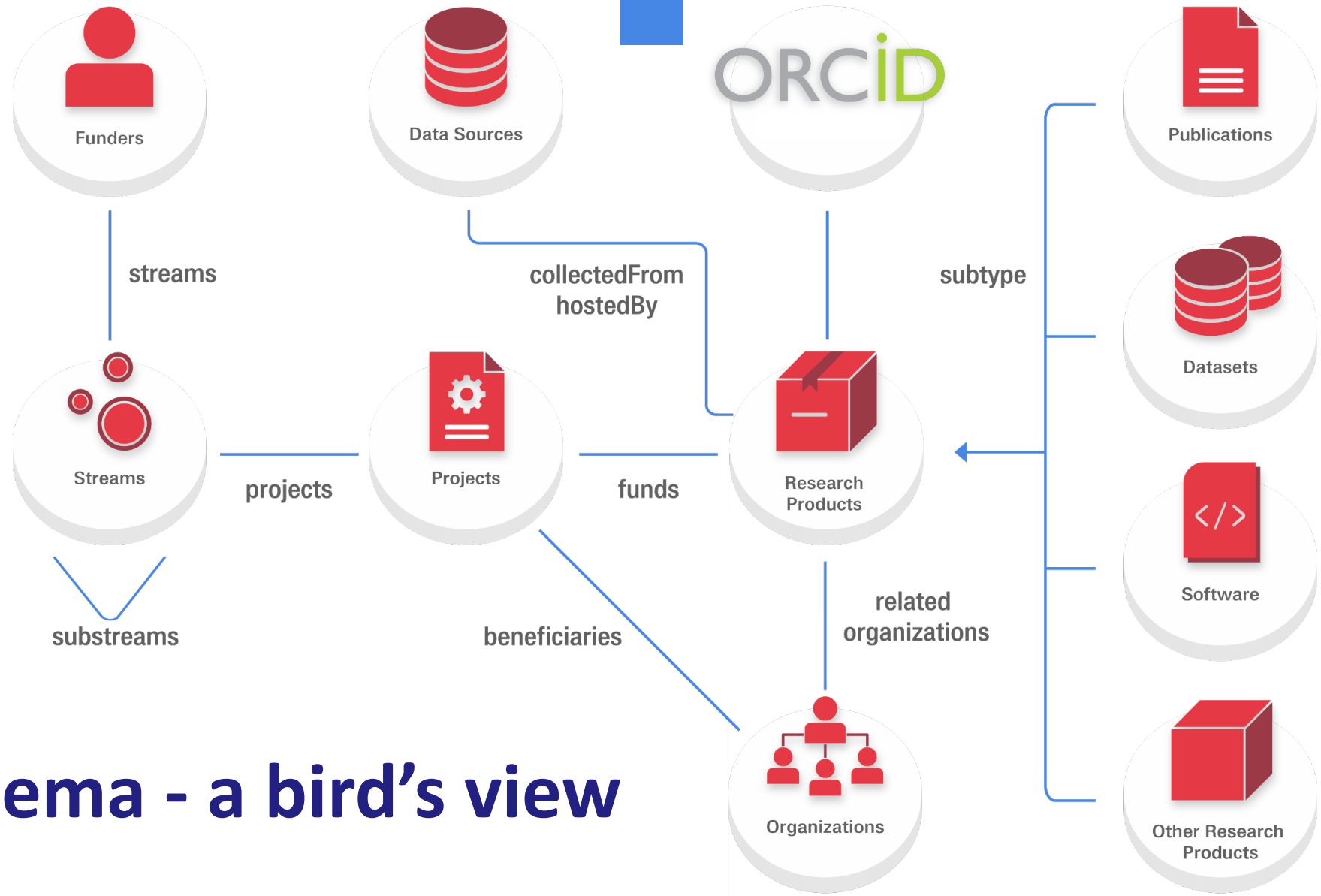
## Discovering open research

Discovery of **reproducible science** outcomes must find new ways, driven by "scientific intentions" that go beyond the "find articles related to a research topic"

## Monitoring (open) research

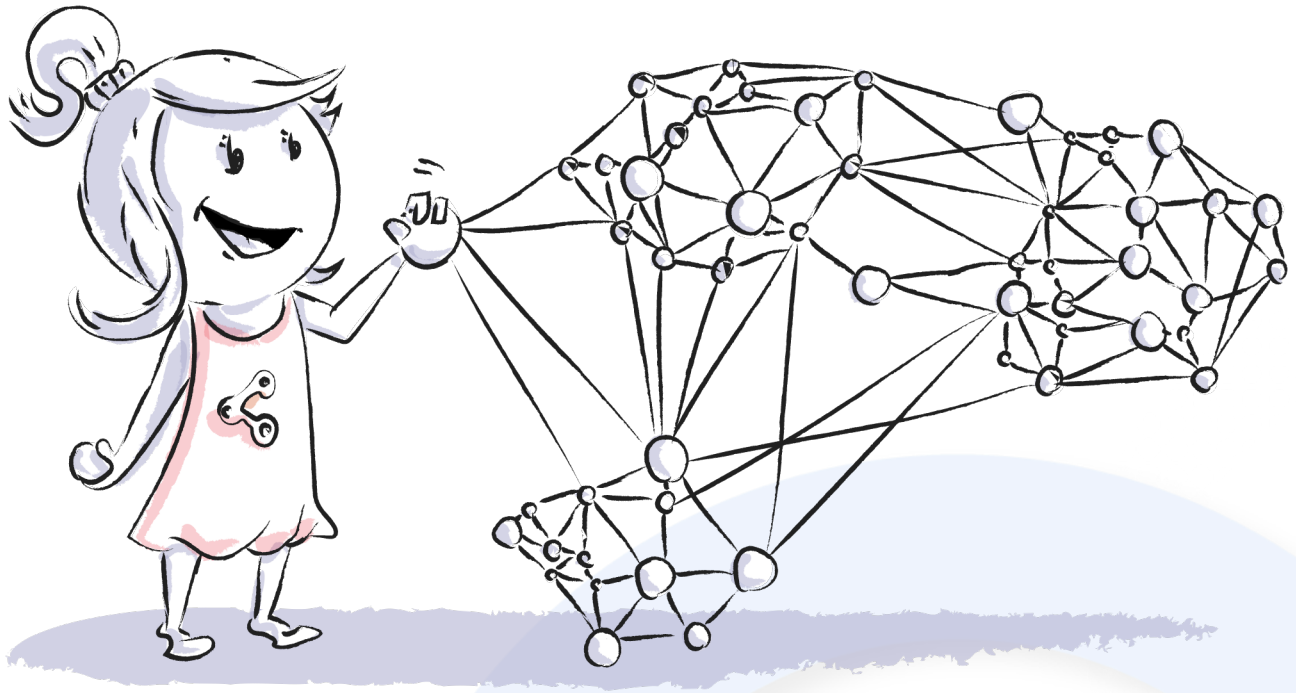
Monitoring quality, impact, and **openness** of science should be a transparent, reproducible process for all, inclusive of research "context"

# The schema - a bird's view



# OpenAIRE Research Graph in numbers

- 25 funders
- ~2k direct data sources
- 3Mi projects
- 144Mi publications
- 304k software
- 17Mi research data
- 6.7Mi other research products



OpenAIRE ResearchGraph

# (EOSC) Users of the Graph

## Direct

- Content providers: publishers, libraries
- Content consumers: data scientists and service providers

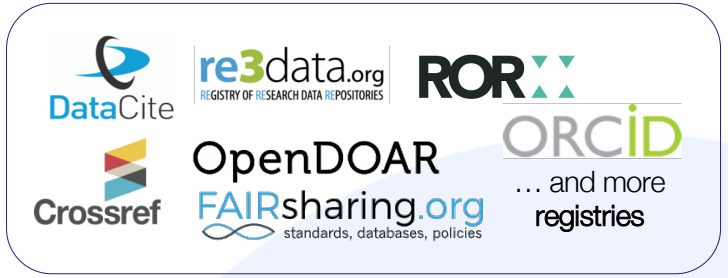
## Indirect (via services)

- Funders, institutions, research infrastructures, policy makers, industry, researchers

## Services

- EOSC marketplace, Commission Participant Portal
- Elsevier Scopus and SciVal, Springer, ORCID
- Institutional and thematic repositories
- MANY others…

# Sources contributing to the Graph



# Register on OpenAIRE | PROVIDE

OpenAIRE Interoperability Guidelines

- Registries
- Repositories
- OA Journals
- Aggregators
- Publishers
- CRIS

**Data sources**

Data Sources

Link

CONNECT

EXPLORE

Metadata, relationships

Metadata, relationships

OpenAIRE UsageCounts

Metadata

EMBL-EBI

Crossref

DataCite

ScholExplorer

Relationships

ORCID

unpaywall

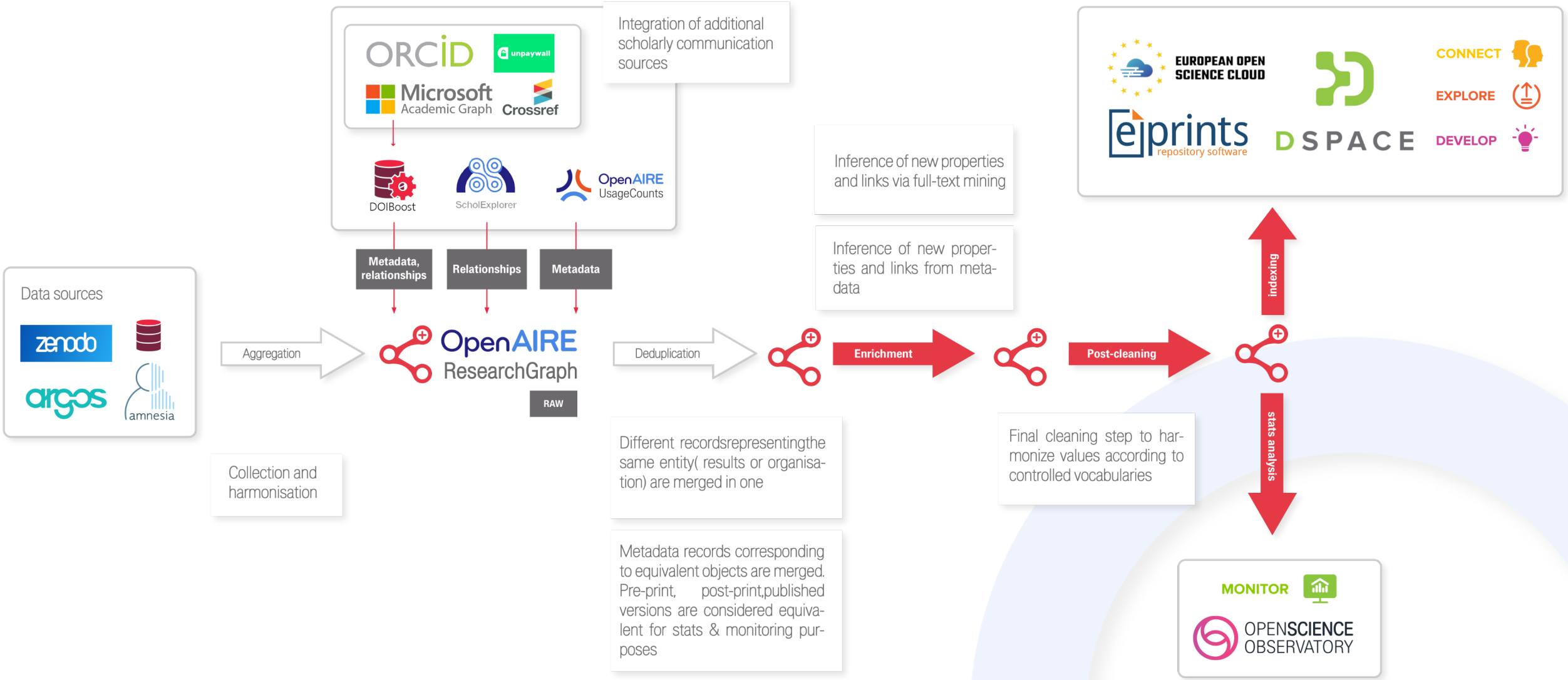
Microsoft Academic Graph

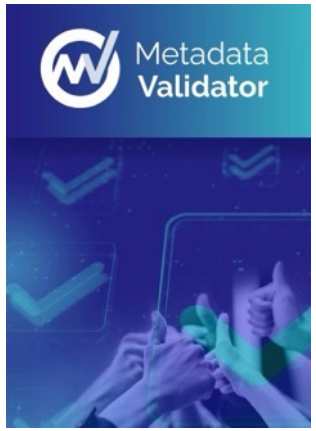
Metadata, relationships

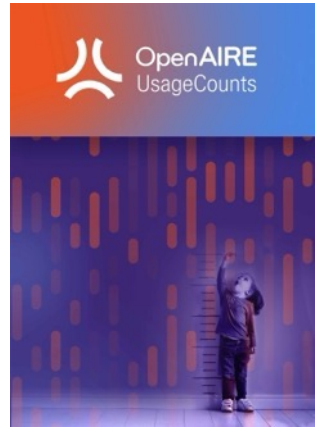
OpenAIRE ResearchGraph

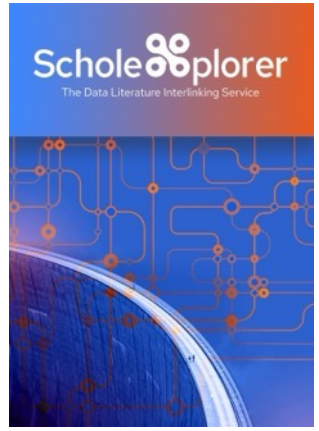
RAW

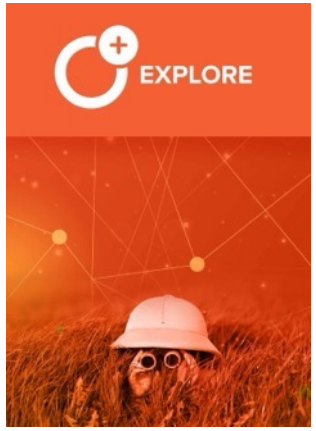
# OpenAIRE Research Graph: the supply chain



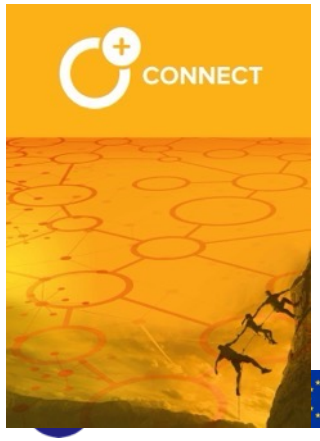


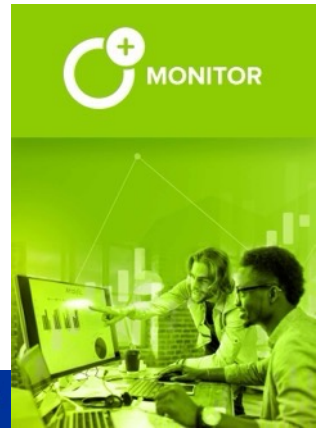



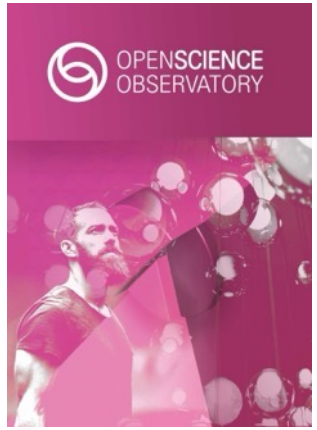



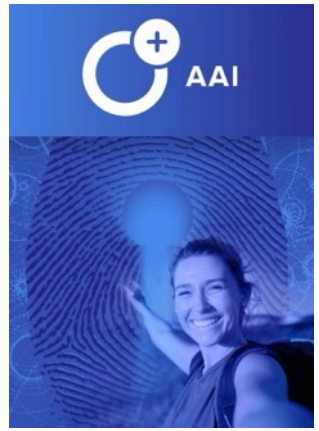



OpenAIRE
ResearchGraph



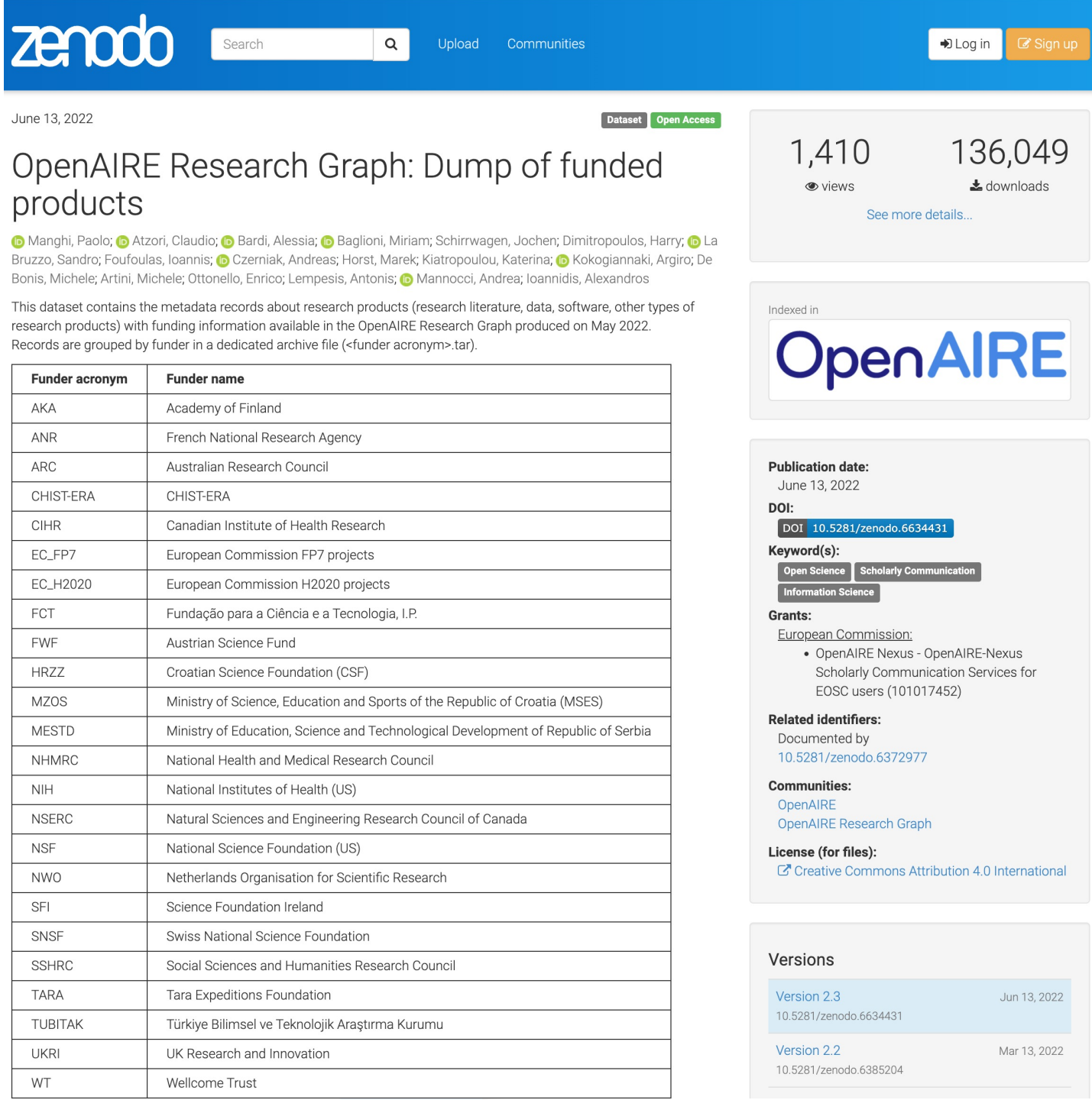


# Access to the ORG

## Data dumps on Zenodo

- Funded products
- RI & RC related products
- COVID-19 related products
- Publication – Dataset links in Scholix format
- Complete
- Linked Open Data

zenodo.org/communities/openaire-research-graph



zenodo Search Upload Communities Log in Sign up

June 13, 2022 Dataset Open Access

### OpenAIRE Research Graph: Dump of funded products

Manghi, Paolo; Atzori, Claudio; Bardi, Alessia; Baglioni, Miriam; Schirrwagen, Jochen; Dimitropoulos, Harry; La Bruzzo, Sandro; Foufoulas, Ioannis; Czerniak, Andreas; Horst, Marek; Kiatropoulou, Katerina; Kokogiannaki, Argiro; De Bonis, Michele; Artini, Michele; Ottonello, Enrico; Lempesis, Antonis; Mannocci, Andrea; Ioannidis, Alexandros

This dataset contains the metadata records about research products (research literature, data, software, other types of research products) with funding information available in the OpenAIRE Research Graph produced on May 2022. Records are grouped by funder in a dedicated archive file (<funder acronym>.tar).

| Funder acronym | Funder name |
|---|---|
| AKA | Academy of Finland |
| ANR | French National Research Agency |
| ARC | Australian Research Council |
| CHIST-ERA | CHIST-ERA |
| CIHR | Canadian Institute of Health Research |
| EC_FP7 | European Commission FP7 projects |
| EC_H2020 | European Commission H2020 projects |
| FCT | Fundação para a Ciência e a Tecnologia, I.P. |
| FWF | Austrian Science Fund |
| HRZZ | Croatian Science Foundation (CSF) |
| MZOS | Ministry of Science, Education and Sports of the Republic of Croatia (MSES) |
| MESTD | Ministry of Education, Science and Technological Development of Republic of Serbia |
| NHMRC | National Health and Medical Research Council |
| NIH | National Institutes of Health (US) |
| NSERC | Natural Sciences and Engineering Research Council of Canada |
| NSF | National Science Foundation (US) |
| NWO | Netherlands Organisation for Scientific Research |
| SFI | Science Foundation Ireland |
| SNSF | Swiss National Science Foundation |
| SSHRC | Social Sciences and Humanities Research Council |
| TARA | Tara Expeditions Foundation |
| TUBITAK | Türkiye Bilimsel ve Teknolojik Araştırma Kurumu |
| UKRI | UK Research and Innovation |
| WT | Wellcome Trust |

1,410 views 136,049 downloads

See more details...

Indexed in OpenAIRE

**Publication date:** June 13, 2022

**DOI:** DOI 10.5281/zenodo.6634431

**Keyword(s):** Open Science, Scholarly Communication, Information Science

**Grants:**
European Commission:
- OpenAIRE Nexus - OpenAIRE-Nexus Scholarly Communication Services for EOSC users (101017452)

**Related identifiers:**
Documented by 10.5281/zenodo.6372977

**Communities:**
OpenAIRE
OpenAIRE Research Graph

**License (for files):**
Creative Commons Attribution 4.0 International

**Versions**

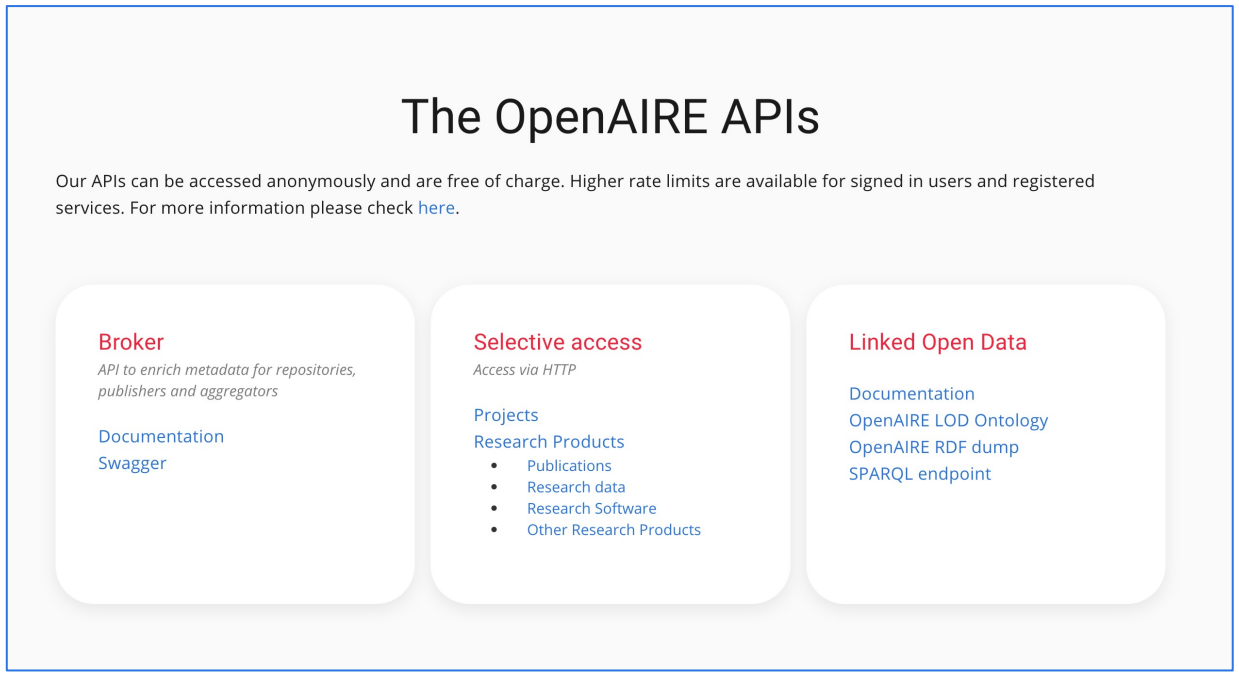
Version 2.3 Jun 13, 2022
10.5281/zenodo.6634431

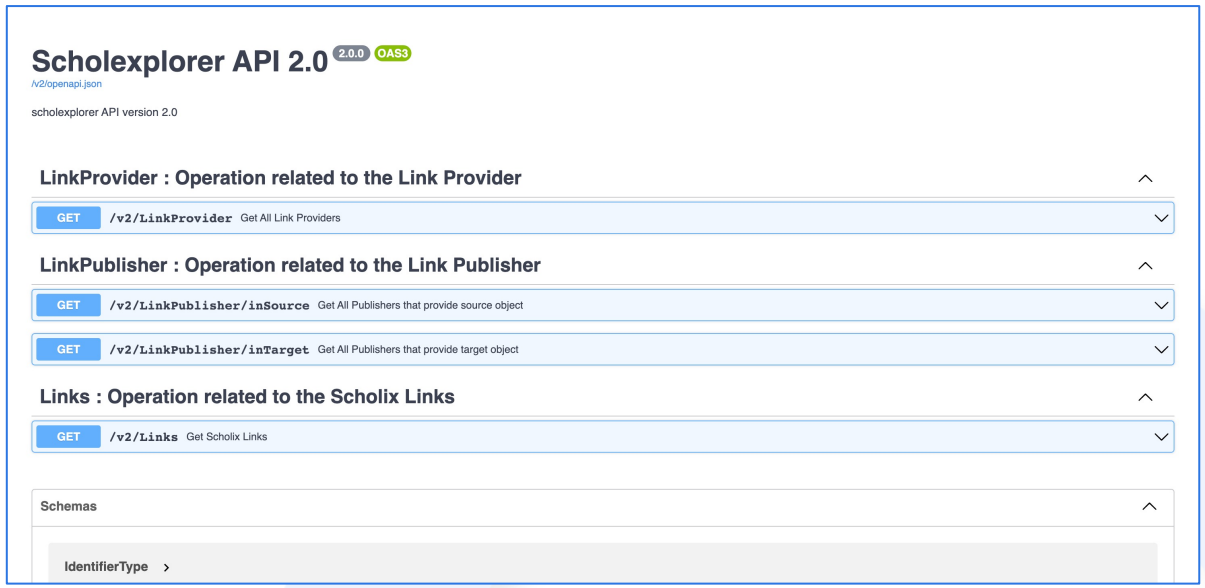
Version 2.2 Mar 13, 2022
10.5281/zenodo.6385204

# Access to the OpenAIRE Graph

Graph: graph.openaire.eu/develop

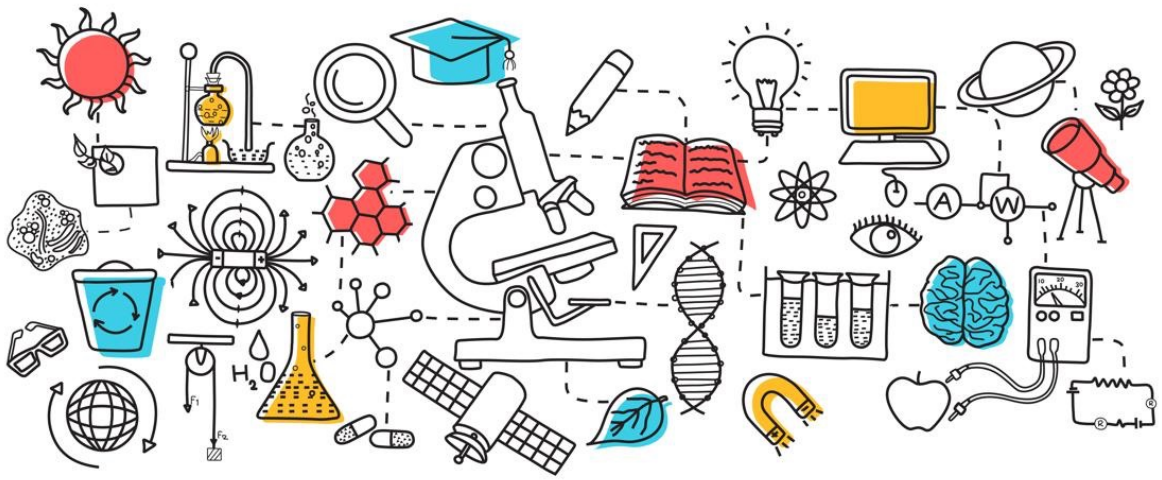
**Scholexplorer:**
api.scholexplorer.openaire.eu/v2/ui/



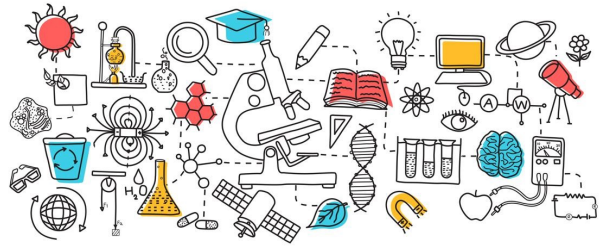
# Novelties

## Fields of Science



SUSTAINABLE DEVELOPMENT GOALS

# FoS Classification in the Graph



## Why

- Growth of # scientific publications
- Need for policy-makers, funders, publishers, scholars, companies to classify it according to Field-of-Science Taxonomies
  - organization, search engines
  - monitoring (evolution of trends, identification of new fields, ..)
  - evaluation (funding allocation, weak spots, opportunities, ..)
  - policy-making
  - recommendation systems

## What

- ~15Mi DOIs classified
- Classification based on a 3 level hierarchy taxonomy
- Based on “SciNoBo: A Hierarchical Multi-Label Classifier of Scientific Publications” 10.1145/3487553.3524677

| Level of FOS | Number FOS Labels |
|---|---|
| Level 1 | 6 |
| Level 2 | 42 |
| Level 3 | 174 |

# SDG classification in the Graph

## Why

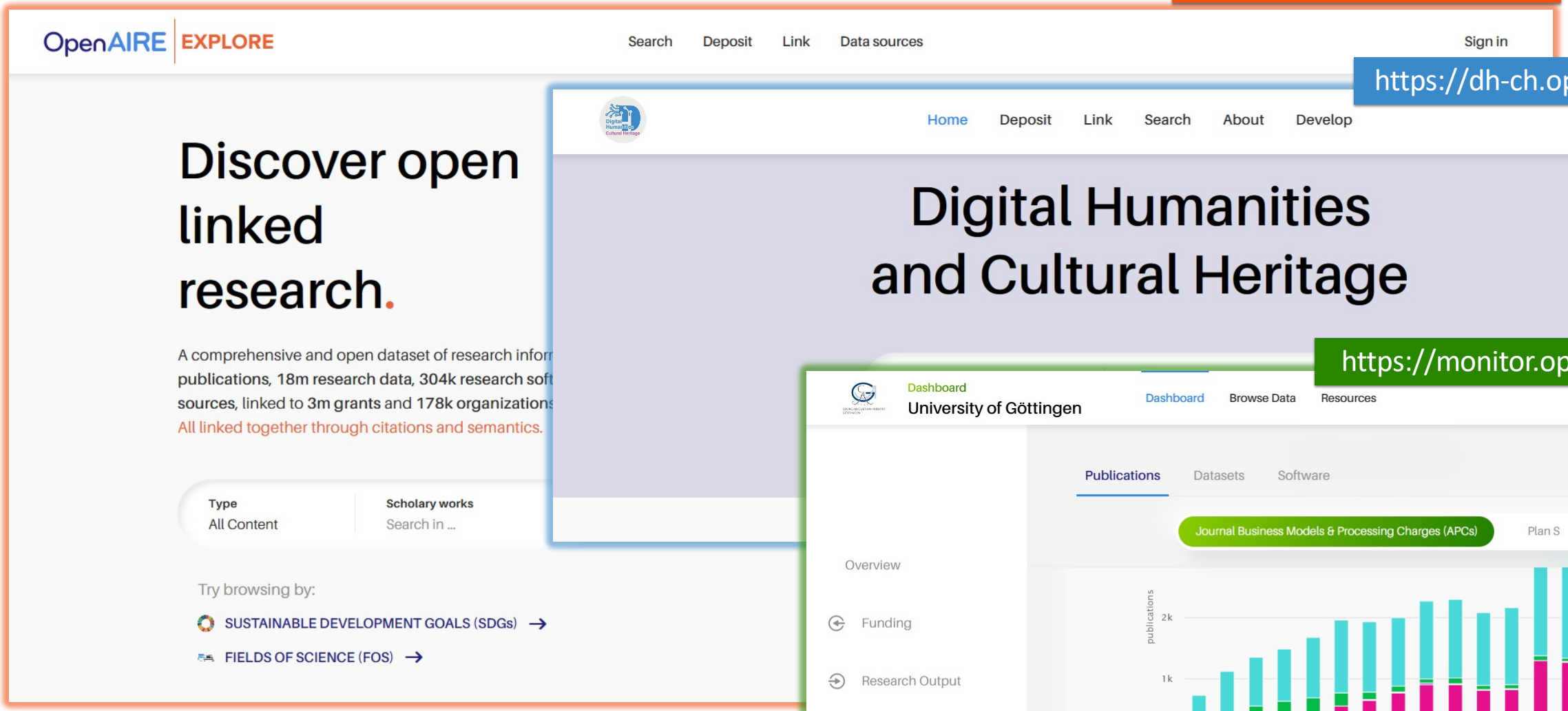
- Societal priorities (partly) set by UN Sustainable Development Goals
- Need for policy-makers to classify research according SDGs
  - monitoring
  - evaluation (societal impact)
  - policy-making

## What

- ~8Mi DOI classified
- SDG Classification system
  - Build an SDG Silver Corpus of Keywords & Keyphrases
  - uses publication title and abstract to classify it to (potentially multiple) SDGs

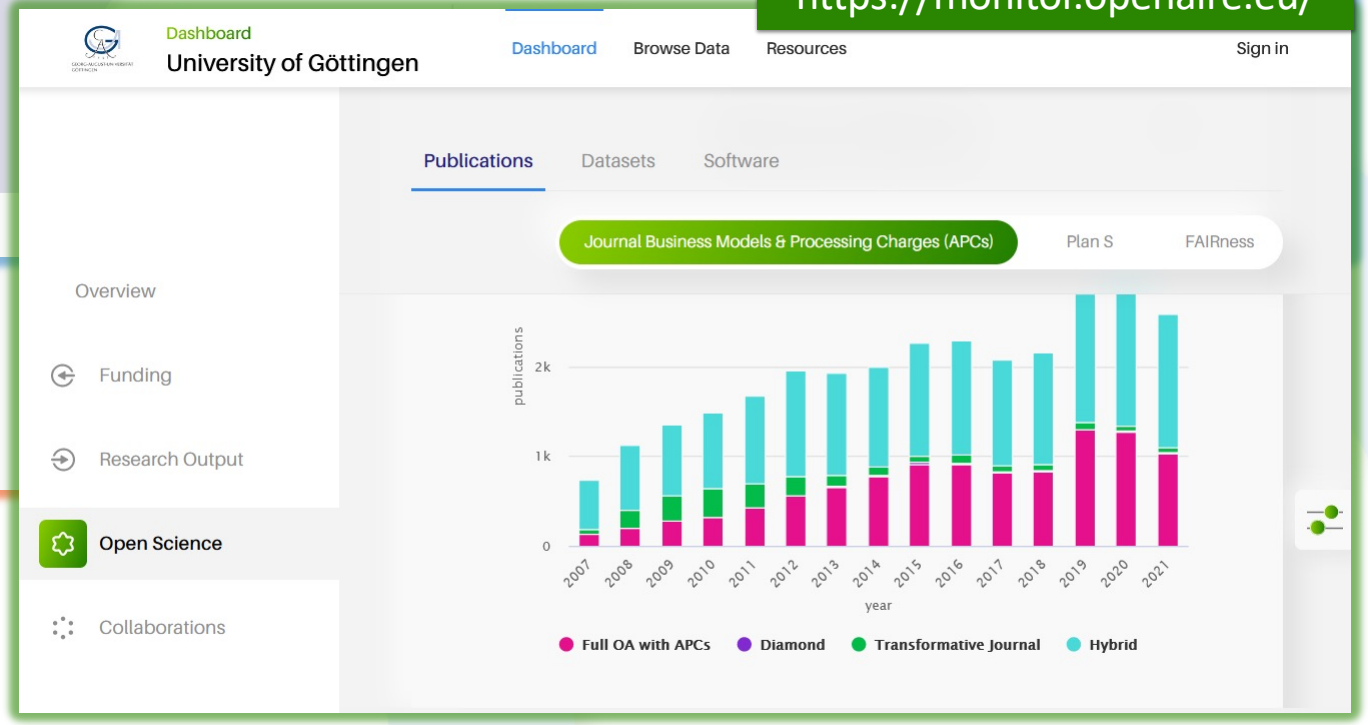
# Related OpenAIRE Services

https://explore.openaire.eu/




https://dh-ch.openaire.eu/

https://monitor.openaire.eu/



# Can I trust the indicators?

**a.k.a can I trust the processes for their calculation?**
**a.k.a can I trust the underlying data?**

"Not everything that counts can be counted, and not everything that can be counted counts."

## Challenges

- Data incompleteness
- Data format inconsistencies
- Default values
- Duplicate data
- Old / stale data
- Inconsistent keys
- Coverage

## ORG as a trustable resource

- Vocabulary based cleaning
- Automation of the aggregation processes
- Provenance information
- Transparency of the processes
  - Documentation (WIP)
- Repeatability

# Thank you for your attention so far!

intelcomp

# Evidence-based R&I Policy Making

*The Role of the OpenAIRE Research Graph*

intelcomp

A Competitive Intelligence Cloud/HPC Platform for AI-based STI Policy Making

Horizon 2020 – SOCIETAL CHALLENGES - Europe In A Changing World - Inclusive, Innovative And Reflective Societies

*(Jan 2020 – Dec 2023)*

# Why?

## R&I activities

- 310Bi EUR: EU expenditure in R&D in 2020 *(EC)*
- a priority across different players

- drive large share of Europe’s economic growth
- new and better jobs
- key in addressing societal challenges

# Why?

## R&I policy making

- align with priorities
  - sustainable development goals (social, economic, environmental dimensions)
- open and inclusive
  - “transparent, evidence-driven, accessible and responsive to as wide a range of citizens as possible” *(OECD)*
- up-to-date with state of play in R&I activities
  - Speed, complexity, interconnectedness, growth and size

# Why?

From **agenda setting**

- where should I invest in next?
  - research topic, organization, country, etc.
  - opportunities or room for improvement
  - alignment societal goals

to **evaluation**

- what is the impact of R&I activities on the society
  - different sectors/areas
  - timing (short, medium, long-term)
  - enabling factors and pathways
- how did (my) funding/policy/approach contribute

navigate the R&I data and knowledge space and make evidence-based decisions.

# How? Methodological Needs

**Tracking and evaluating** R&I activities must be

- **data-driven & relevant –** use key aspects of big data available
- **comprehensive & granular** – 360º view across multiple facets of R&I activities & their links, down to a fine level
- **automated & timely**
- **sustainable –** tracking long-term & repeatedly
- **transparent & replicable** → trustworthy

# How? Methodological Approach

- **Policy Intelligence**
  - leverage big data with AI-based systems
    - dynamic, multilingual and heterogeneous data
  - human-in-the-loop
  - state of the art technology
    - in computational power (High Performance Computing-HPC-, EOSC, commercial clouds),
    - AI techniques (e.g., NLP, ML, data mining)
  - frequent updates

- **Open & FAIR Data and Open, Transparent & Reproducible Methodologies**

  → discover, link, track propagation of R&I activities in the society and automate the process

  → replicable assessment

# What? The IntelComp Platform

An **end-to-end** platform, on a HPC environment, for *evidence-based R&I assessment and policy-making*

# What? The IntelComp Platform

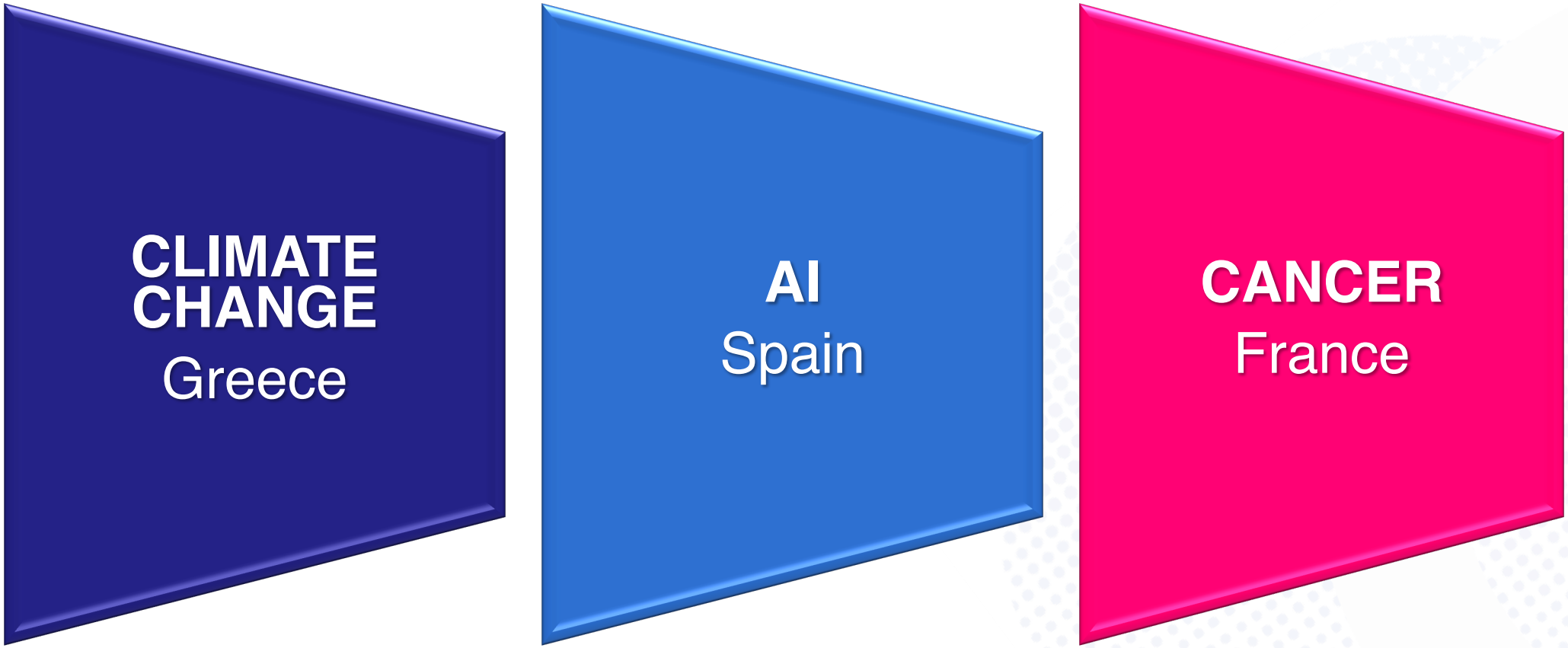
- Set of tools providing a cooperative environment where actors (mainly policy makers, funders, analysts and public administrators) can **visualize**, **interact** and **analyze** information

- Business intelligence R&I dashboards
- Search & browsing tools
- Project proposal evaluation tool
- Supported by
  - data lake
  - NLP pipelines & analytics workflows
  - interactive model trainer
  - catalogue of resources

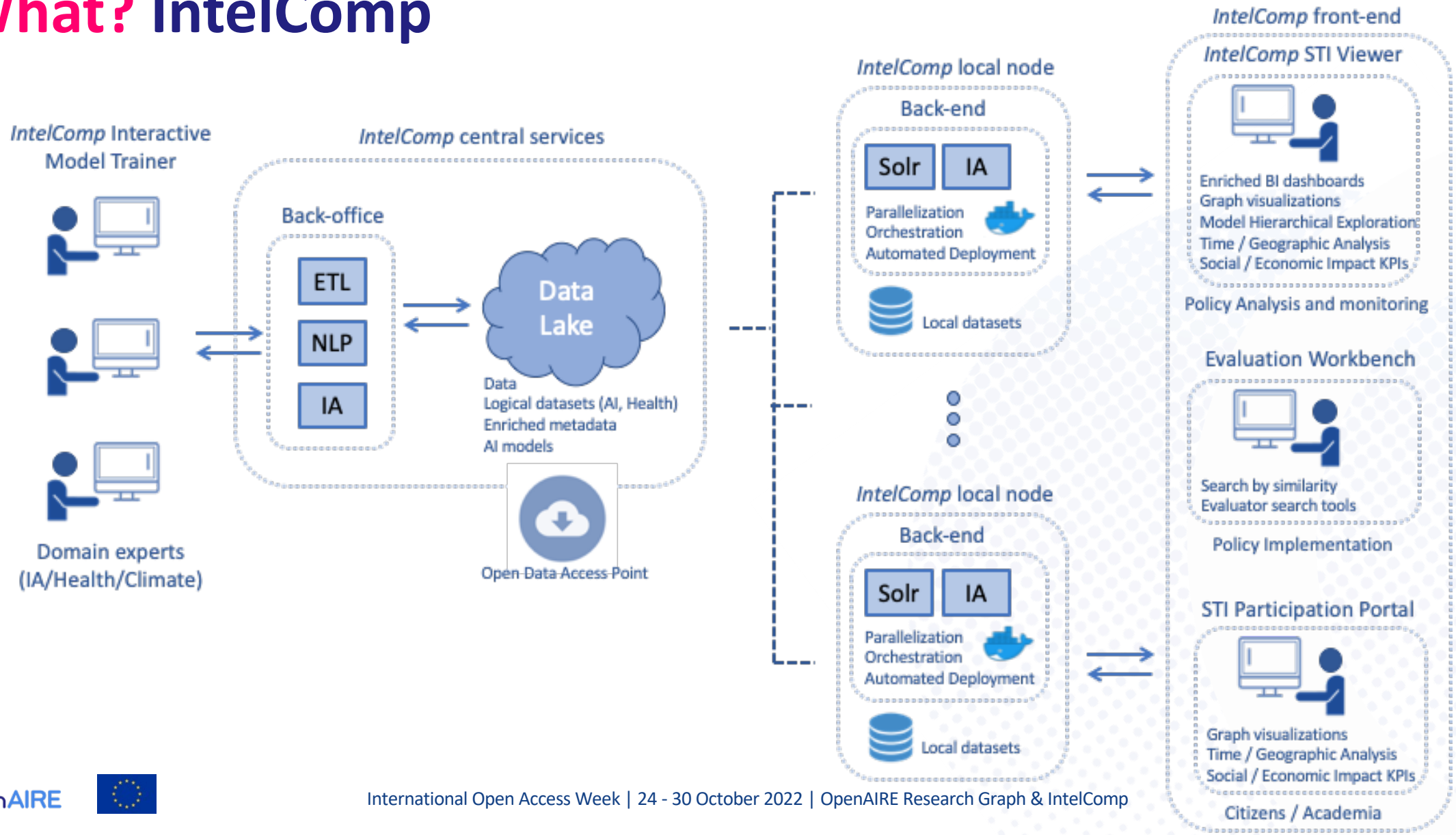
intelcomp

# What? The IntelComp Platform

- CO-CREATED using a Living Labs approach



# What? IntelComp



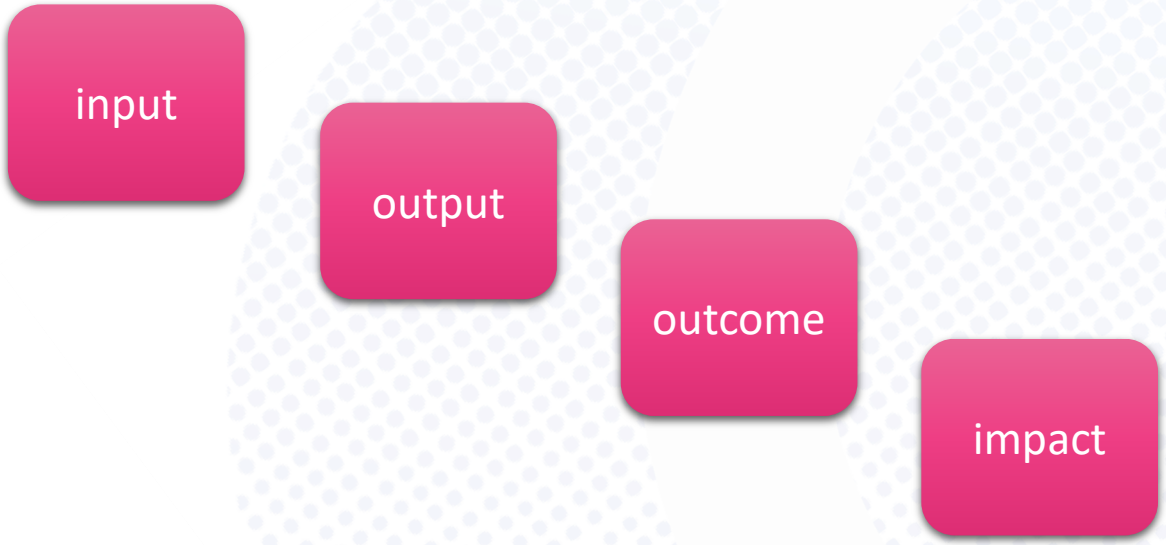
# The OpenAIRE Research Graph in IntelComp – Why?

## Agenda Setting – policy maker

- R&I landscape for a **topic x country/EU**

## Evaluation - funder

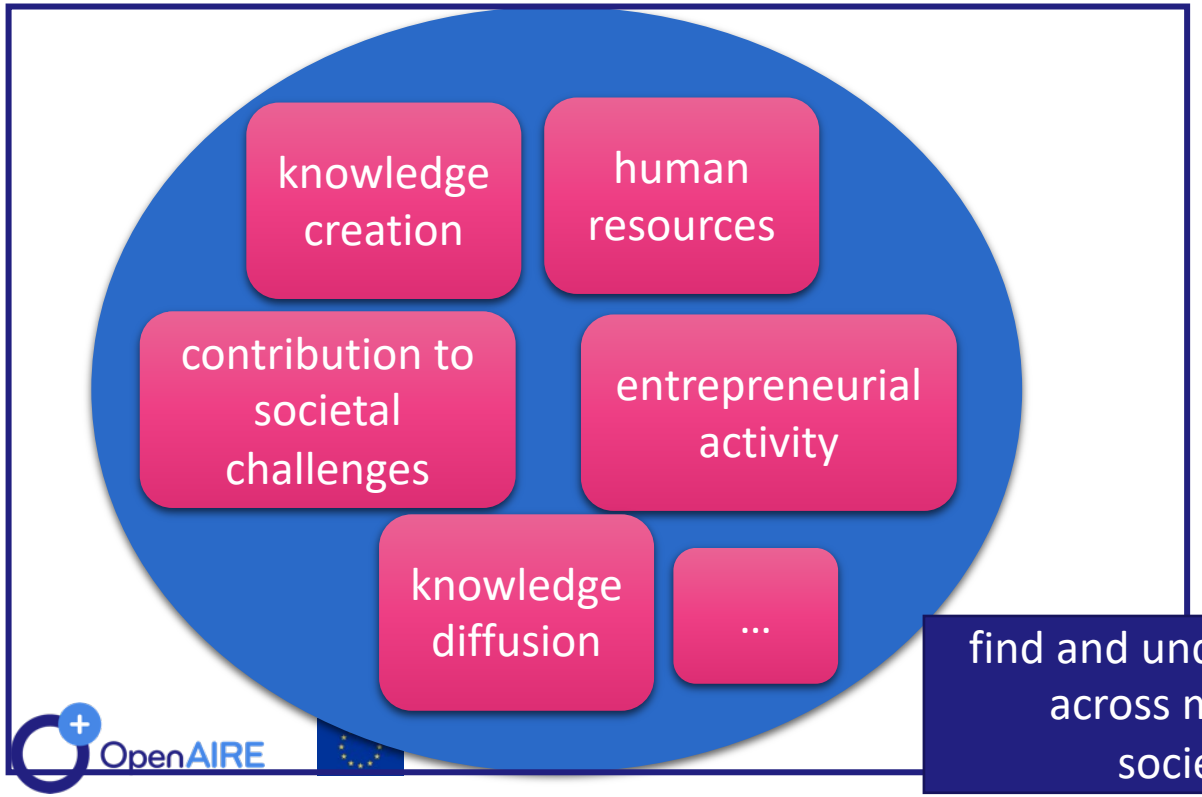
- Impact of **project or programme**



# The OpenAIRE Research Graph in IntelComp – Why?

## Agenda Setting – policy maker

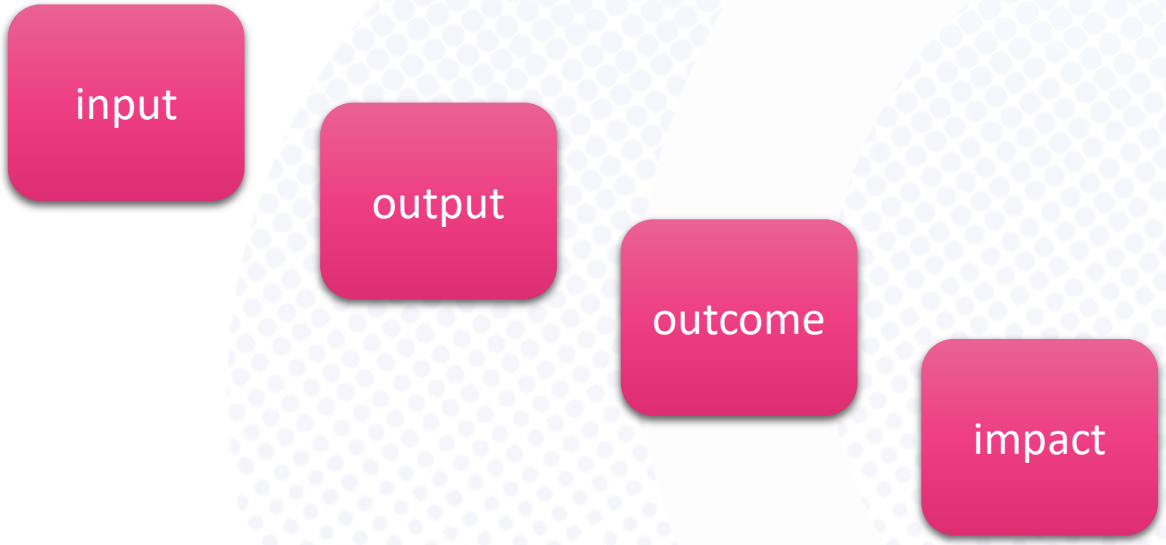
- R&I landscape for a **topic x country/EU**




find and understand topic dynamics across multiple areas of the society/data sources

## Evaluation - funder

- Impact of **project or programme**



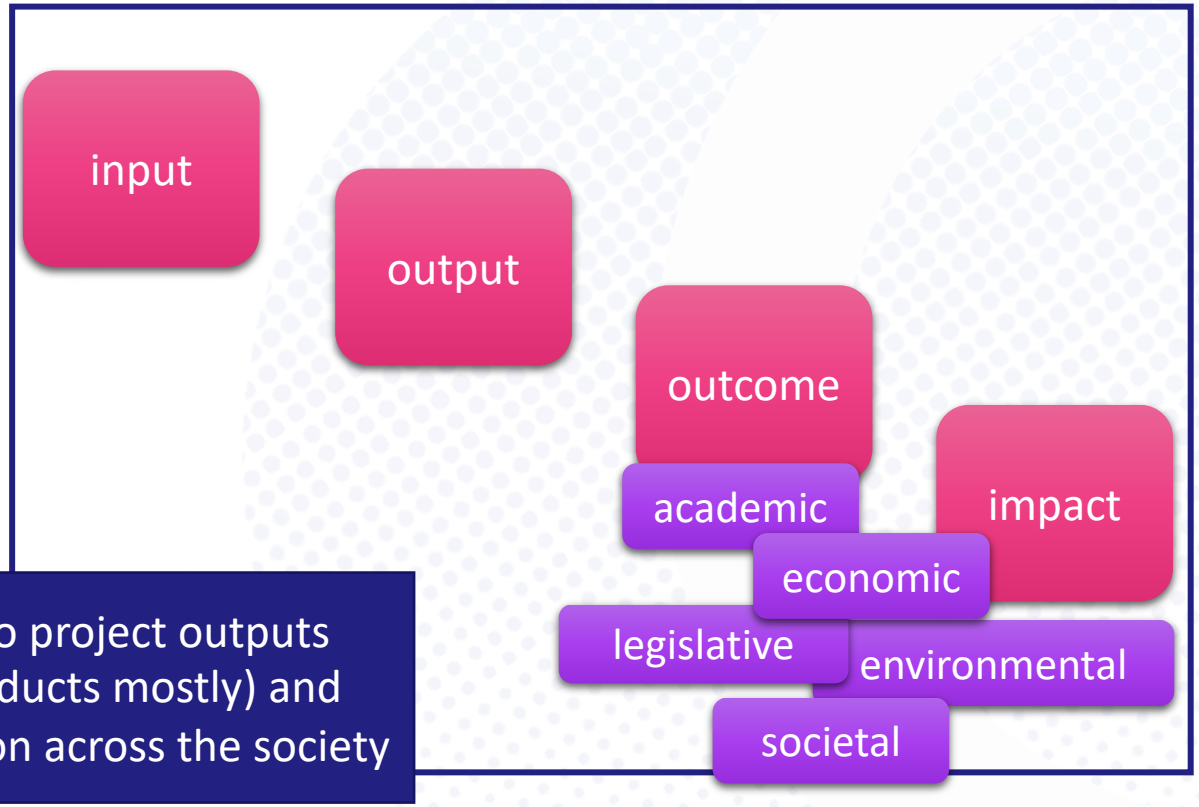
# The OpenAIRE Research Graph in IntelComp – Why?

## Agenda Setting – policy maker

- R&I landscape for a **topic x country/EU**

## Evaluation - funder

- Impact of **project or programme**




link project to project outputs (research products mostly) and track propagation across the society

OpenAIRE

# The OpenAIRE Research Graph in IntelComp

## WHY a Research Graph for Scientific Products?

- **KEY**
  - as a major data source for agenda setting (what's going on in science?)
  - as step # 1 in impact assessment (what did this project directly create?)

## WHY the OpenAIRE Research Graph?

- **Coverage, Readiness, Timeliness**
  - scientific research outputs + links to each other and projects
  - rich metadata (organizations, data sources, citations, APCs, etc.)
  - fully operational big data infrastructure
- **All about open science and open data**
  - inclusive, transparent, replicable
- **Fully embedded in EOSC infrastructure**

OpenAIRE ResearchGraph

# The OpenAIRE Research Graph in IntelComp – How?

## Major data source

- **in NLP pipelines & analytics workflows**
  - extract information to enrich research outputs
    - Topic modelling, FOS & SDG classification
    - innovation extraction, claim/evidence, etc
- **linkages across R&I products, actors** (networks)

➔R&I indicators, broken down in a granular level

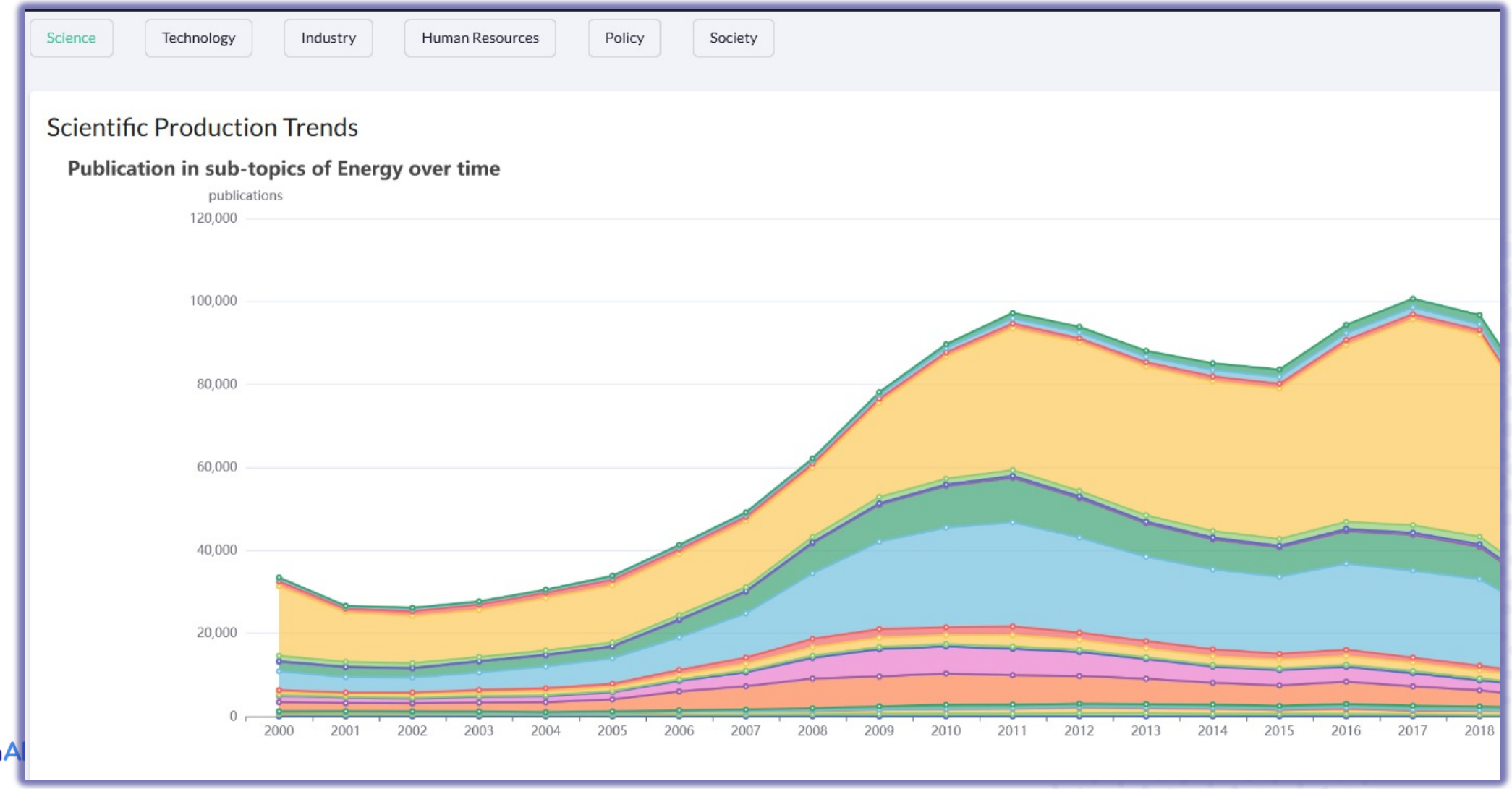
# The OpenAIRE Research Graph in IntelComp – What?

## A sneak preview of the **STI Viewer**

- **One of three IntelComp platforms for end-users**
  - STI policy participation portal, Evaluation Workbench
- **Target audience: R&I (policy) analysts**
- Analyze, compare and visualize a comprehensive set of R&I-related KPIs for Agenda Setting, Evaluation

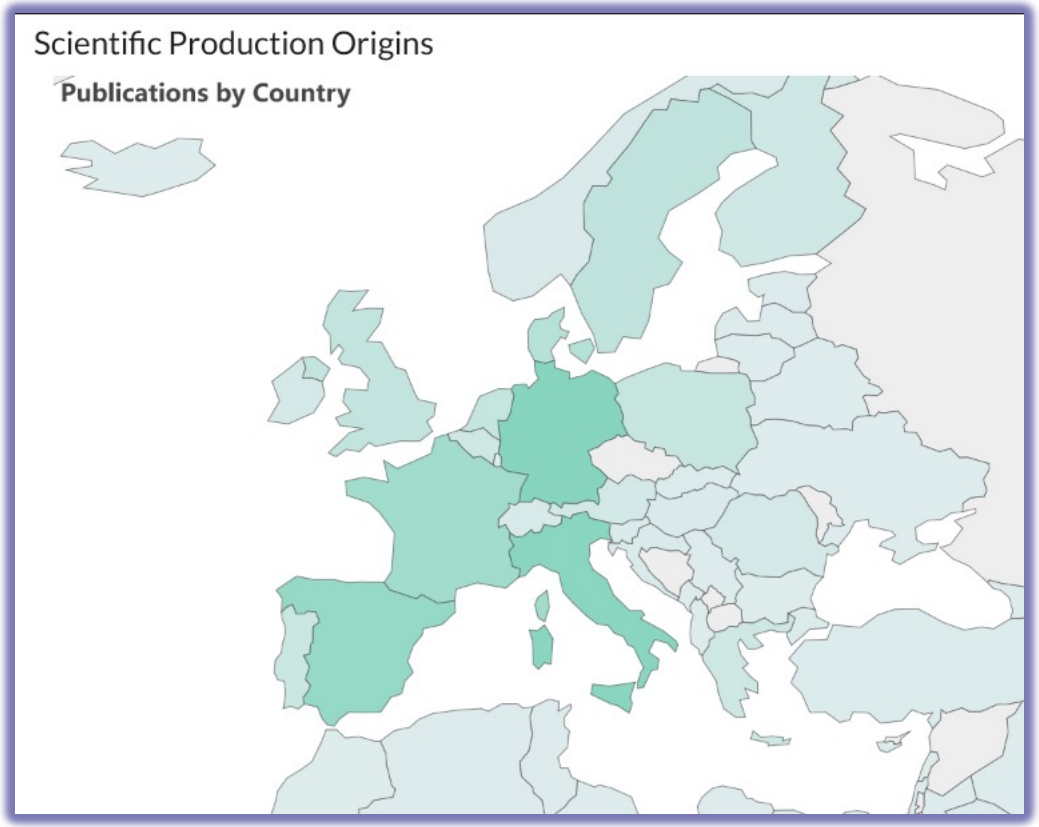
# The OpenAIRE Research Graph in IntelComp

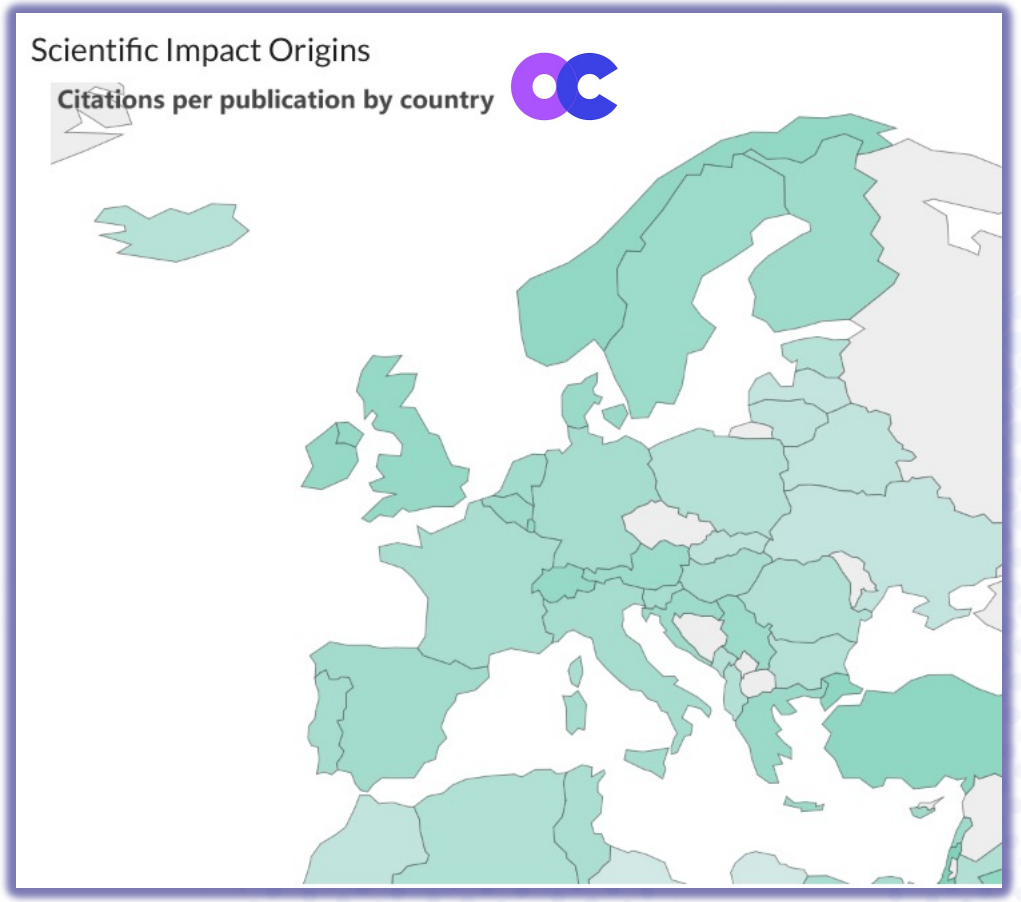
## Agenda Setting in Energy-EU



# The OpenAIRE Research Graph in IntelComp

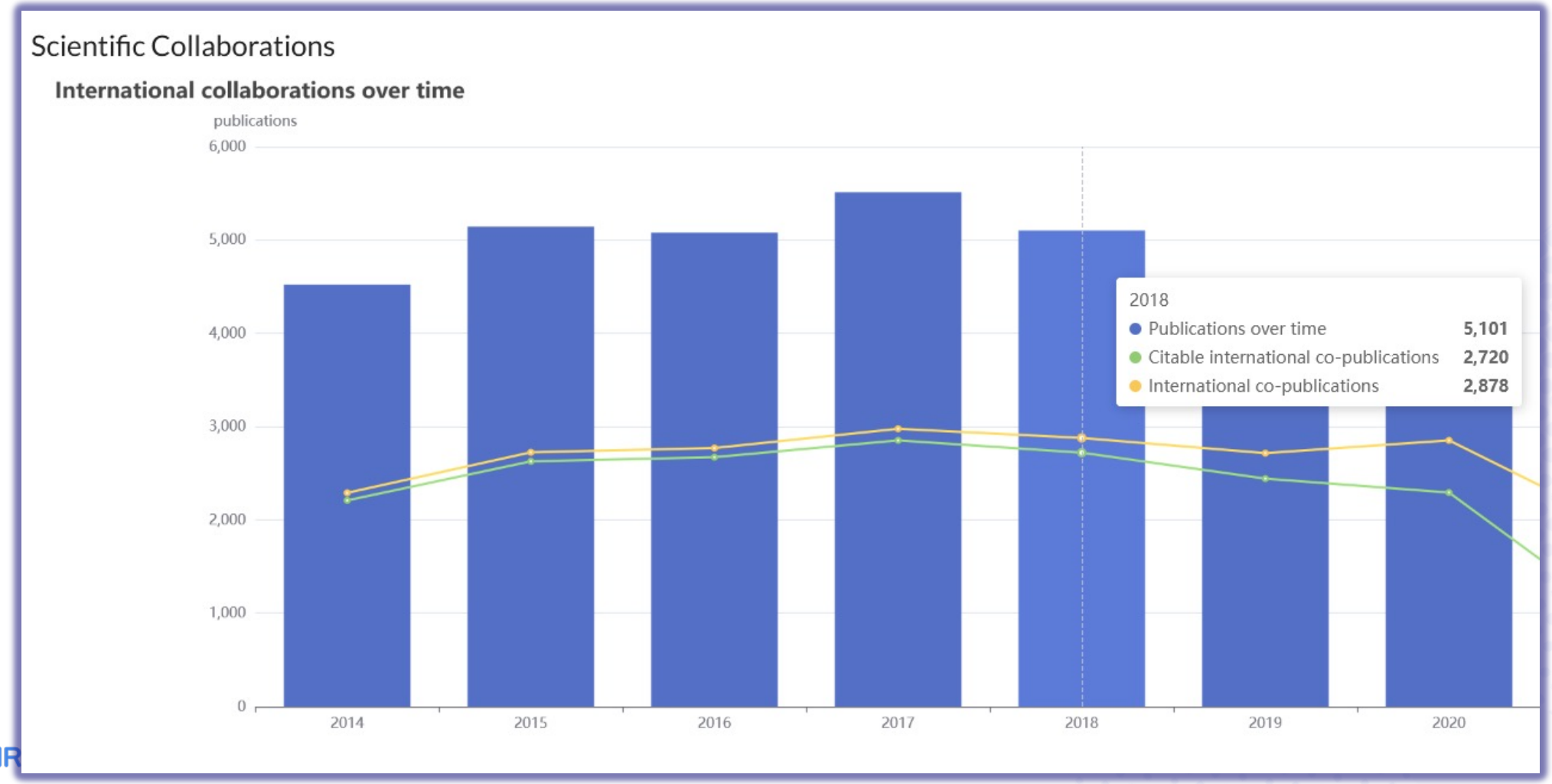
## Agenda Setting in Energy-EU



# The OpenAIRE Research Graph in IntelComp

## Agenda Setting in Energy-EU

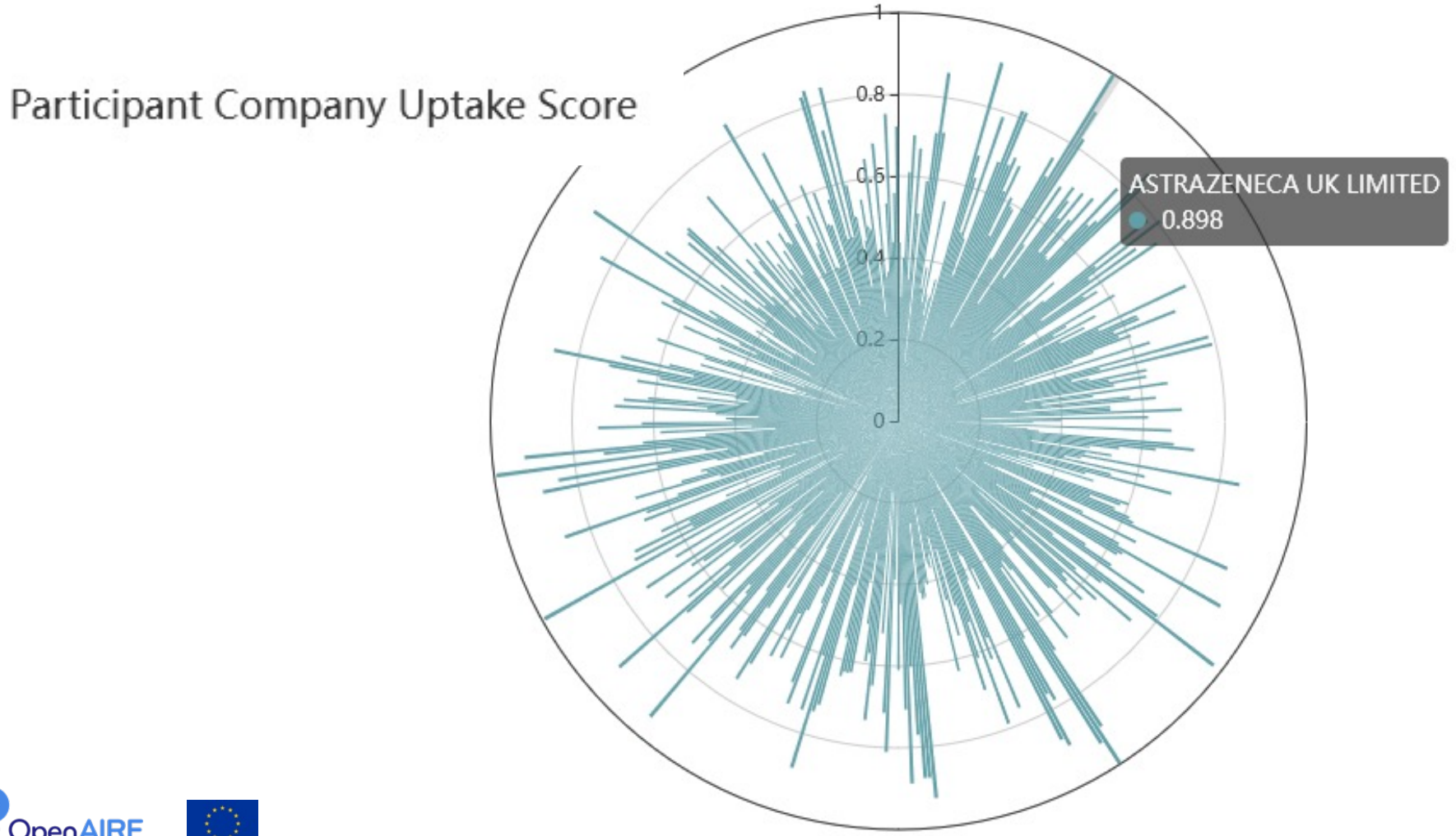


# The OpenAIRE Research Graph in IntelComp

## Agenda Setting in Energy-EU

# The OpenAIRE Research Graph in IntelComp

## Monitoring for European Commission-Health



# TAKE AWAYS

## Evidence-based R&I Policy Making

- Importance of
  - policy intelligence
  - open, transparent & replicable KPIs

- Leveraging the OpenAIRE Research Graph
  - links and metadata
  - to create new ones

  for policy making in R&I
  - timely, comprehensive, granular views

# THANKS

**Email**
claudio.atzori@isti.cnr.it
igrypari@athenarc.gr

**Twitter**
@openaire_eu
@athenaRICinfo
@IntelComp_H2020