graph.openaire.eu

# OpenAIRE Services and the Research Graph

## *Open and Sustainable monitoring in the Era of Open Science*

**Paolo Manghi**

*OpenAIRE AMKE, CTO*

*Institute of Information Science and Technologies, National Research Council, Italy*

OpenAIRE

European Commission

@openaire_eu

## Tracking Open Science

Reproduciblity and transparency require tracking of all outcomes of science and related ”context”

## Monitoring Open Science

Monitoring quality, impact, and ”open scienceness” of science should be a transparent, reproducible process for all, inclusive of research “context”

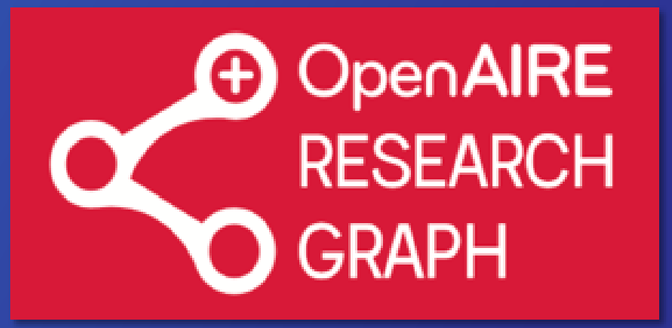
## Discovery Open Science

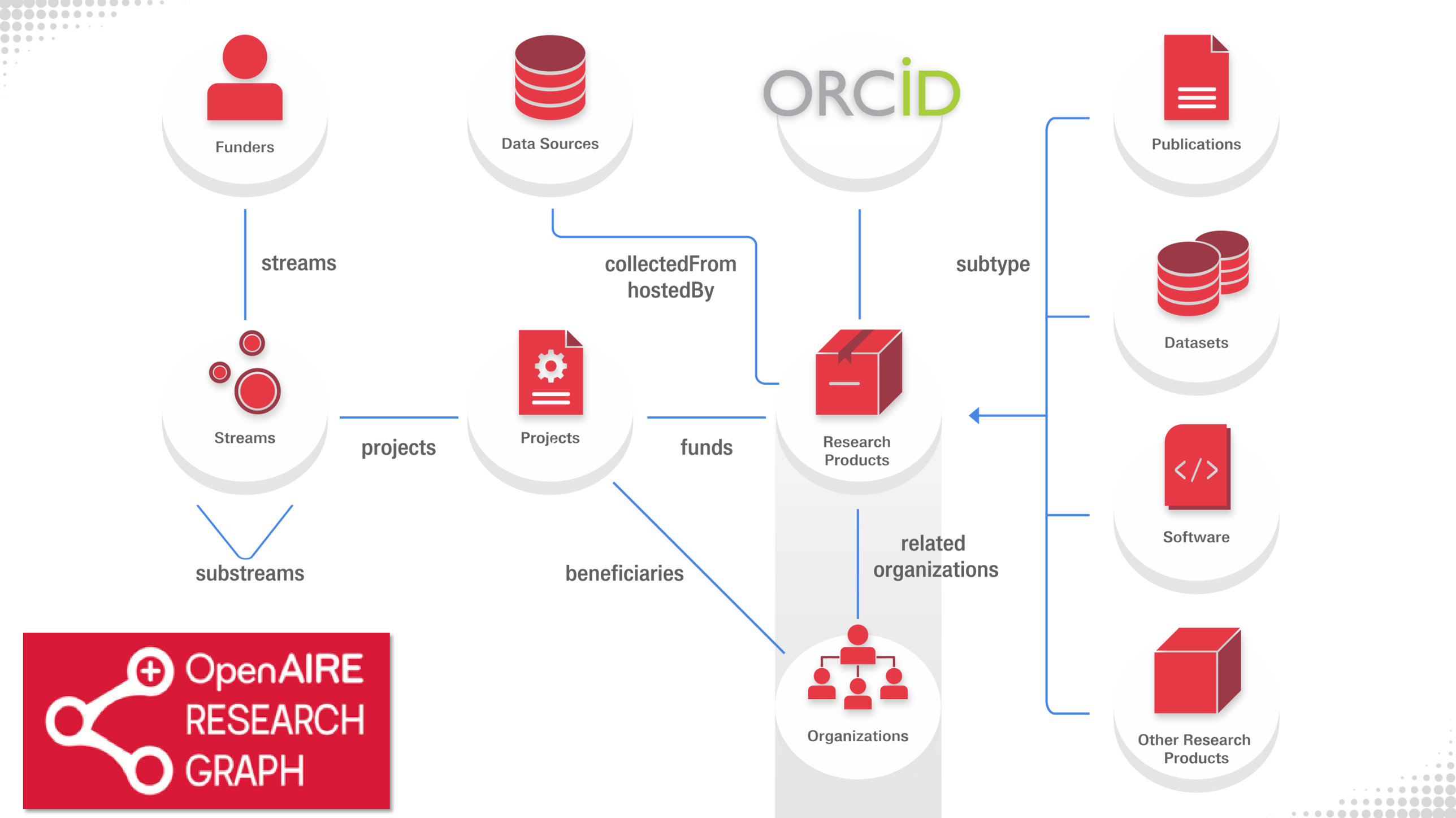
Discovery of reproducibe science outcomes must find new ways, driven by “scientific intentions” that go beyond the “find articles related to a research topic”.

OpenAIRE

European Commission

# Tracking Open Science: the OpenAIRE Research Graph



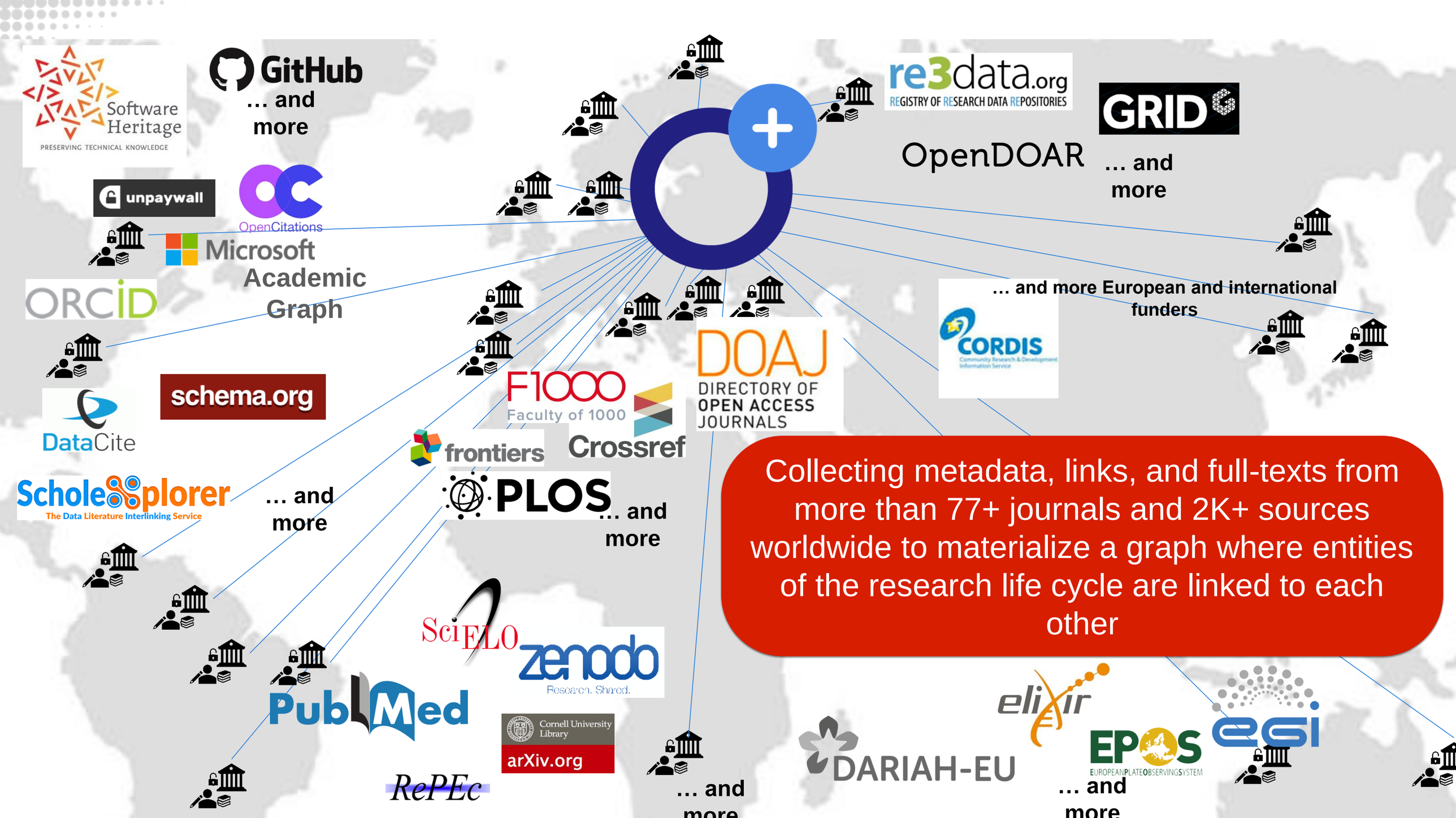

Funders
Data Sources
ORCID
Publications
streams
collectedFrom
hostedBy
subtype
Datasets
Streams
projects
Projects
funds
Research
Products
Software
substreams
beneficiaries
related
organizations
Organizations
Other Research
Products
OpenAIRE
RESEARCH
GRAPH

creative commons 0 ZERO
CC BY

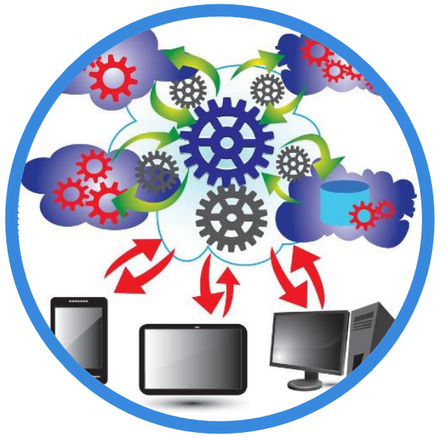
OpenAIRE RESEARCH GRAPH

**Providing an open metadata research graph of interlinked scientific products, with Open Access information, linked to funding information and research communities**

- Open
- Complete
- De-duplicated
- Transparent
- Participatory
- Decentralized
- Trusted


Software Heritage
PRESERVING TECHNICAL KNOWLEDGE
GitHub
… and more
re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES
GRID
OpenDOAR
… and more
unpaywall
OpenCitations
Microsoft Academic Graph
ORCID
… and more European and international funders
CORDIS
Community Research & Development Information Service
schema.org
DataCite
DOAJ
DIRECTORY OF OPEN ACCESS JOURNALS
F1000
Faculty of 1000
frontiers
Crossref
ScholeXplorer
The Data Literature Interlinking Service
… and more
PLOS
… and more
Collecting metadata, links, and full-texts from more than 77+ journals and 2K+ sources worldwide to materialize a graph where entities of the research life cycle are linked to each other
SciELO
zenodo
Research. Shared.
PubMed
Cornell University Library
arXiv.org
RePEc
… and more
elixir
DARIAH-EU
EPOS
EUROPEANPLATEOBSERVINGSYSTEM
egi
… and more

# Data wrangling challenges



## HETEROGENEITY

Standards, crosswalks, community and cross-community view

GUIDELINES.openaire.eu

## QUALITY & ACCURACY

Trust and stability of the data: researchers and operators in the loop

PROVIDE (validation, brokering), user feedback, PIDs

## COMPLETENESS

Practices, sharing of technology and data (Knowledge Graphs)

Open Science publishing workflows, RDA OSG interoperability

OpenAIRE

# OpenAIRE Research Graph: the supply chain

**Figures (Jan 2019-today):**
- Direct data sources +2400
- Products +360Mi
- Harvested/mined links 236Mi +
- Full-texts 14.Mi+

Full-text mining 13 algorithms: dataset, project, research-initiatives, patent, software, pdb, affiliations, communities, concepts, covid-19, citations, similarity, classification

Propagation of properties via relationships

Data sources

Aggregation

OpenAIRE RESEARCH GRAPH

RAW

Validator

OpenAIRE PROVIDE

The Dataverse Project · eprints repository software · DSPACE · DataCite FIND, ACCESS, AND REUSE DATA · euroCRIS

Deduplication

Different records representing the same entity (results or organization) are merged in one

Metadata records corresponding to equivalent objects are merged. Pre-print, post-print, published versions are considered equivalent for stats & monitoring purposes

Enrichment

OpenAIRE RESEARCH GRAPH

Stats analysis

OpenAIRE MONITOR

indexing

OpenAIRE EXPLORE

OpenAIRE CONNECT

OpenAIRE DEVELOP

APIs

**Figures (2020):**
1,7Mi/y API hits
650K/y user visits

EC Partipant Portal

**Figures (Apr 2021)**:
- Data sources 94K+
- Publications 123Mi
- Datasets 14Mi +
- Software 200K
- Others 8Mi

Funders 23
Projects 2.8Mi
Organizations 276K

OpenAIRE

European Commission

# OpenAIRE Research Graph: PIDs

## Authors

ORCID

Microsoft Academic

## Organizations

GRID

ROR

CORDIS Community Research & Development Information Service

and other funders

## Products

DataCite FIND, ACCESS, AND REUSE DATA

EMBL-EBI

arXiv.org

Identifiers.org

Crossref

## Funders/Projects

FCT Fundação para a Ciência e a Tecnologia

CORDIS Community Research & Development Information Service

NWO

Science Foundation Ireland sfi For what's next

ACADEMY OF FINLAND

and other funders

## Data sources

re3data.org REGISTRY OF RESEARCH DATA REPOSITORIES

OpenDOAR

DOAJ DIRECTORY OF OPEN ACCESS JOURNALS

\+ journal lists from publishers

FAIRSFAIR Fostering Fair Data Practices in Europe

## Services (fortcoming…)

EUROPEAN OPEN SCIENCE CLOUD

…Clusters and RIs service catalogues

OpenAIRE

European Commission

# OpenAIRE Research Graph: PDFs for full-text mining

Publication full-texts

14.Mi+

frontiers

acm

SPRINGER NATURE

Europe PubMed Central

DOAJ DIRECTORY OF OPEN ACCESS JOURNALS

Cornell University Library
arXiv.org

zenodo

ELSEVIER

**Institutional repos**

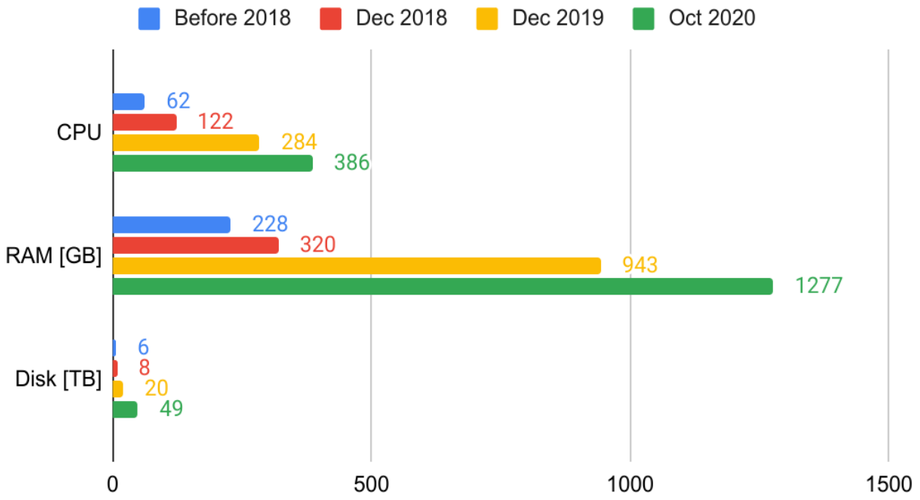
OpenAIRE

European Commission

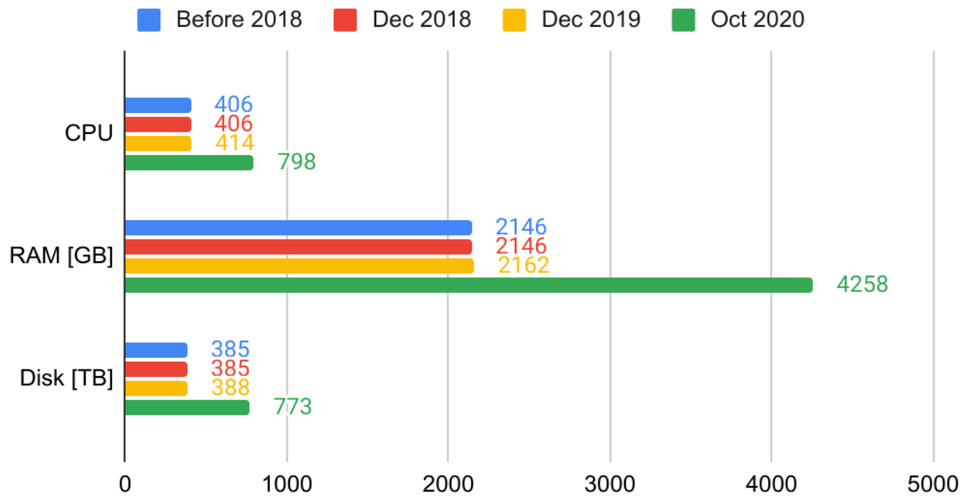
# System infrastructure

| Public System | Mining System | Data provision System | Testing System |
|---|---|---|---|
| 44srv | 38srv | 23srv | 16srv |
| 386CPU | 798CPU | 154CPU | 102CPU |
| 1277GB | 4.3TB | 430GB | 326GB |
| 49TB | 773TB | 24TB | 10TB |

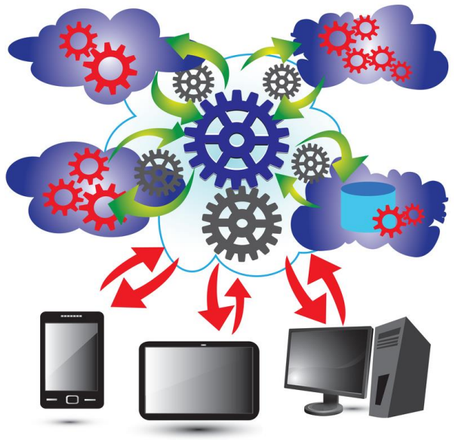
Public subsystem resources




Mining subsystem resources



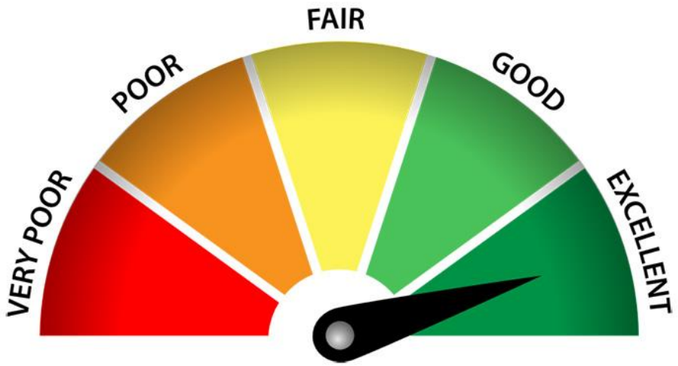
# Tracking Open Science: Challenges



## Heterogenity

Standards, crosswalks, community and cross-community view

GUIDELINES.openaire.eu




## Quality and accuracy

Trust and stability of the data: researchers and operators in the loop

PROVIDE (validation, brokering), user feedback, PIDs

## Completeness

Practices, sharing of technology and data (Knowledge Graphs)

Open Science publishing workflows, RDA OSG interperability, CRISs

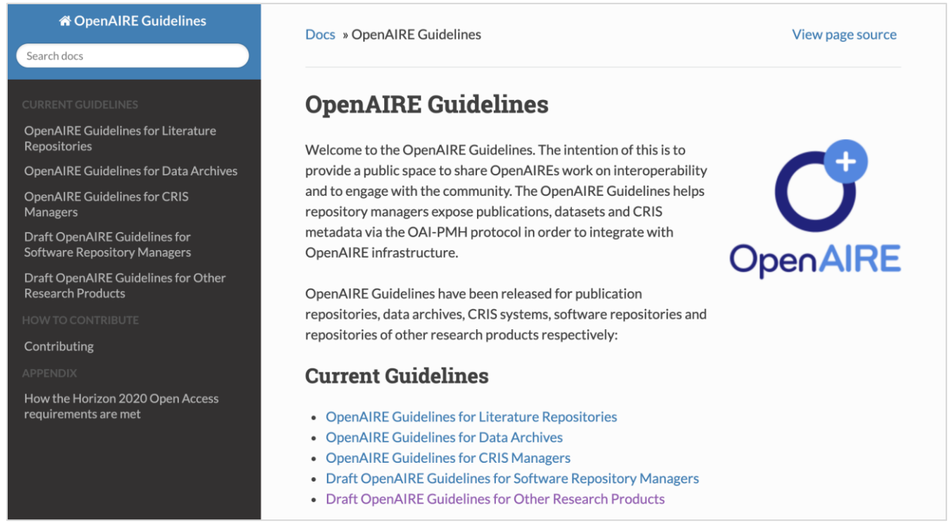
OpenAIRE

European Commission

# OpenAIRE | PROVIDE : data source managers

**VALIDATE**

- **Registration of data sources**
  - Repositories
  - Data archives
  - Software repositories
  - CRIS systems



**REGISTER**

- **Validation against OpenAIRE Guidelines**
- **RDA FAIR Maturity Guidelines**

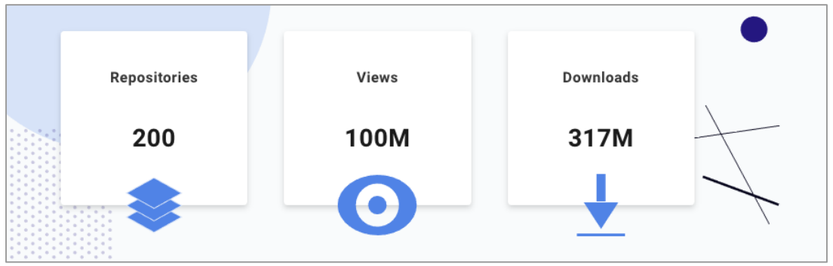
OpenAIRE

European Commission

# OpenAIRE | PROVIDE : added value functions

## Usage statistics for Institutional Repositories

OpenAIRE UsageCounts



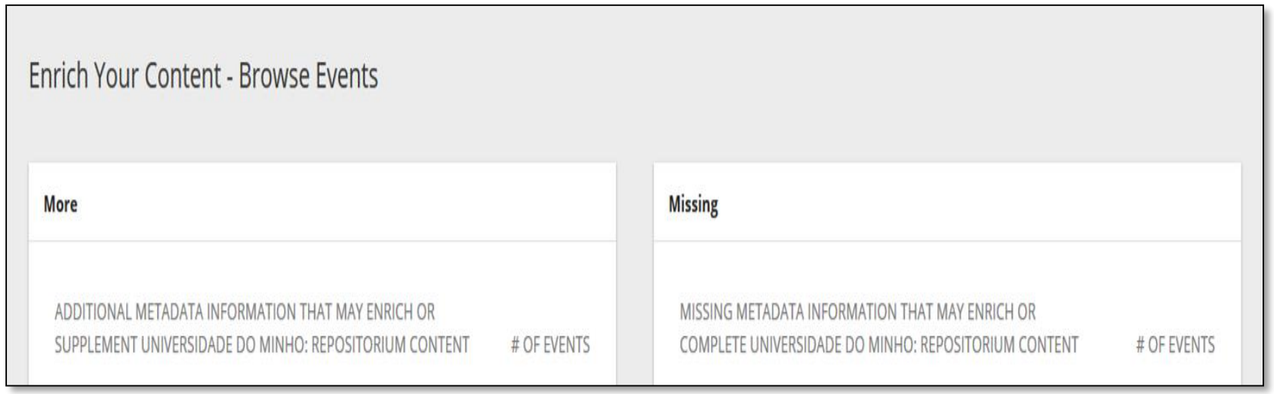

## Brokering of metadata

OA Broker

### Improve your metadata. Get more connections

OA Broker service offers a wealth of information on scholarly communication data. Find out what interests you and subscribe to enrich your records.

ENRICH

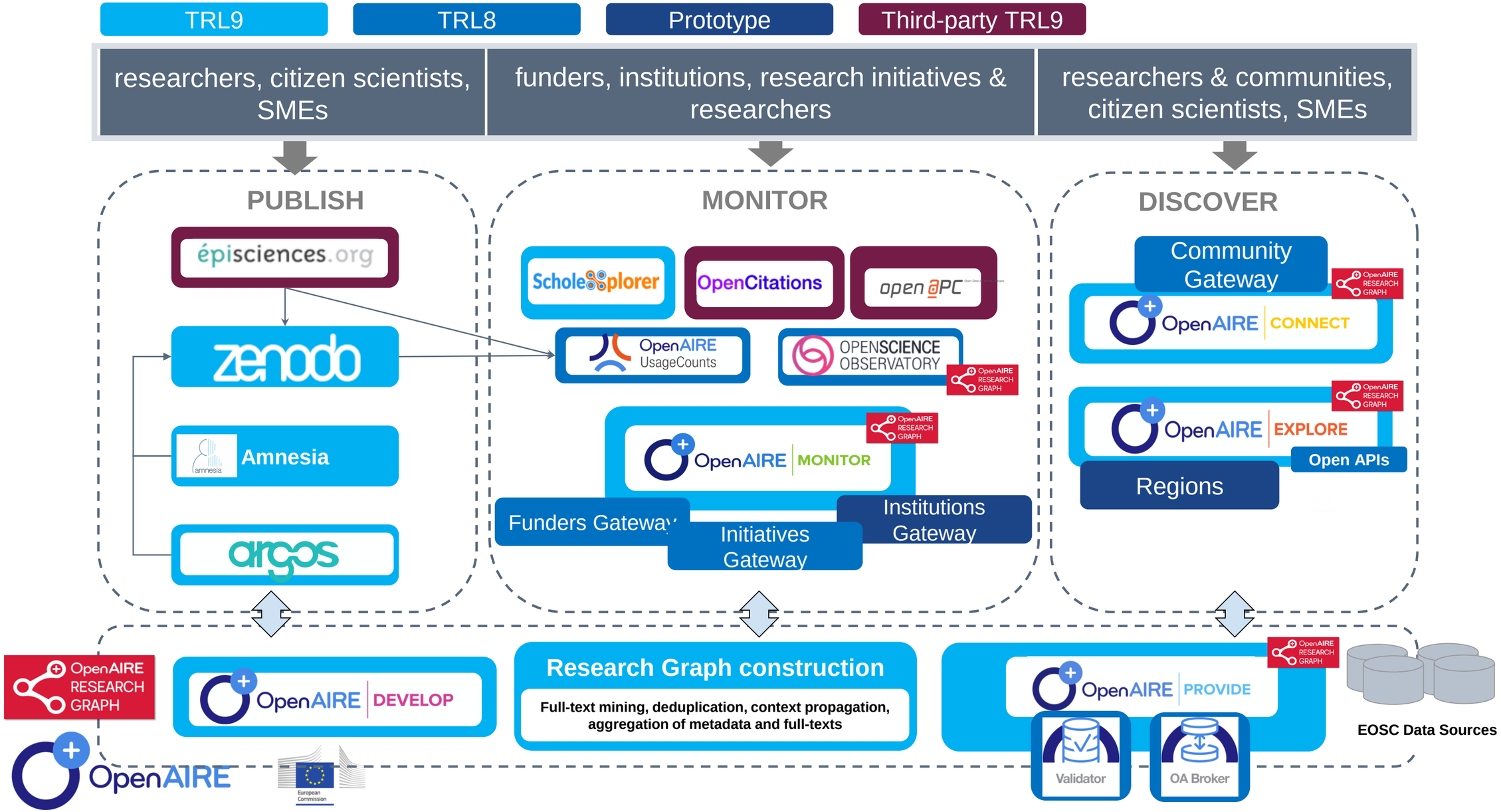



OpenAIRE

European Commission

# Monitoring and Discovery Open Science: Sustainability via provision of services

# OpenAIRE Services



# PUBLISH: researchers, SMEs, citizen scientists

- Overlay journals (HAL, Zenodo,etc.)
- Catch-all repository
- Desktop tool for the anonymization of sensitive data
- Machine-actionable DMPs

OpenAIRE

European Commission

TRL9

TRL8

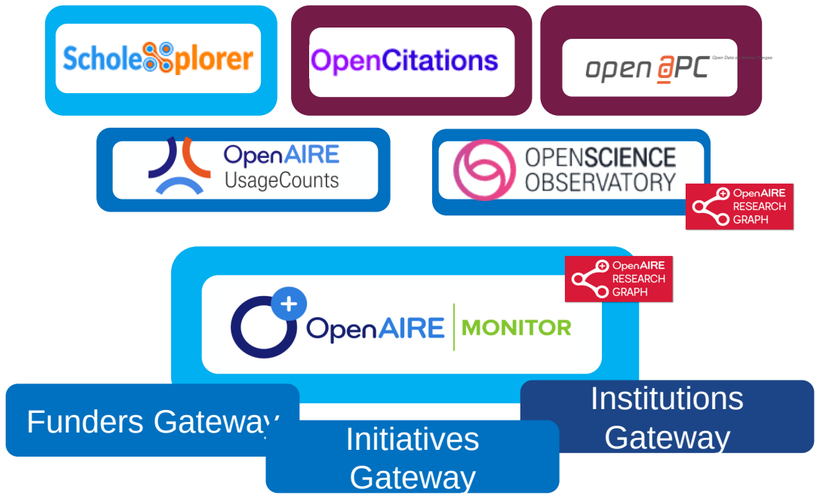
Prototype

Third-party TRL9

# MONITOR

## Data Services

- Data-Article links
- Article-Article links
- APC stats for Europe
- Usage statistics




## Portals

- Research Impact for Funders, Institutions and Research Infrastructures
- Open Science Impact for countries

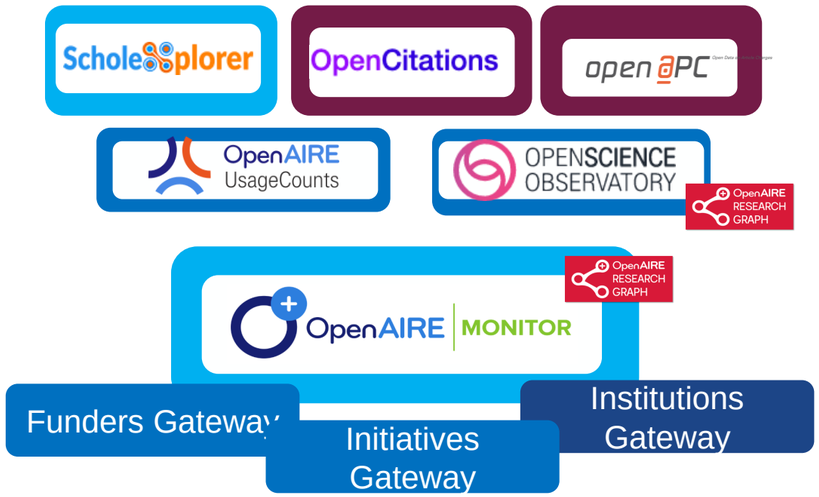
TRL9
TRL8
Prototype
Third-party TRL9

# MONITOR

## Data Services

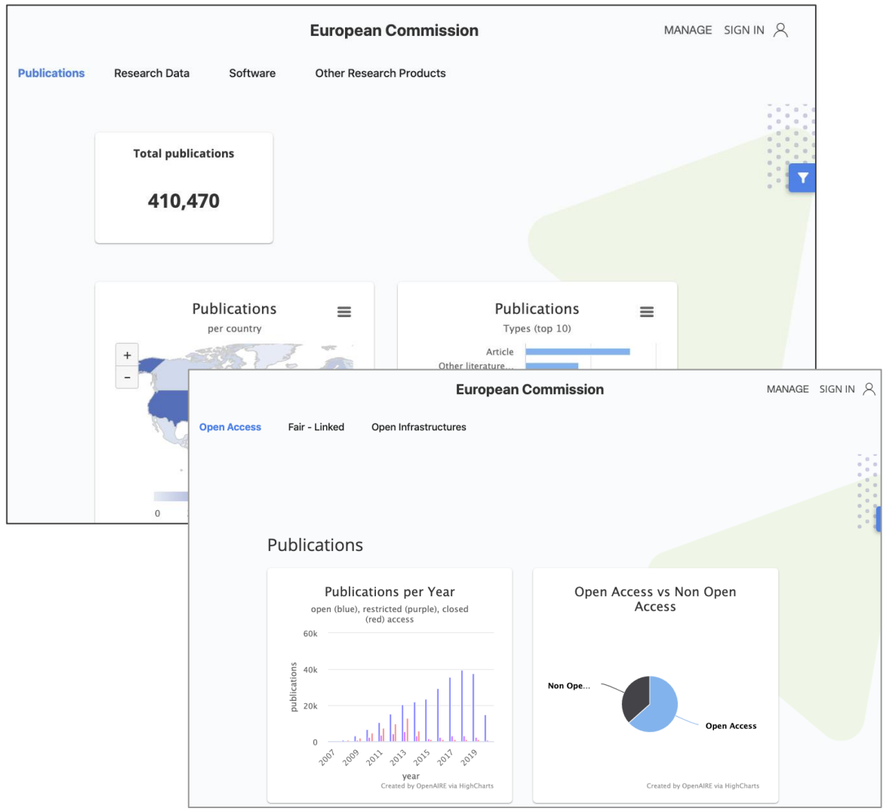
- Data-Article links
- Article-Article links
- APC stats for Europe
- Usage statistics

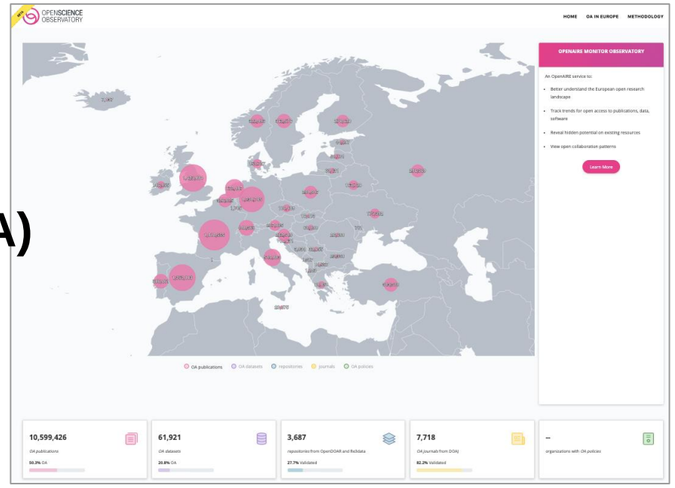
**4 Bi acccesses**




**21 funders (32 in BETA)**

**3 initiatives (8 in BETA)**



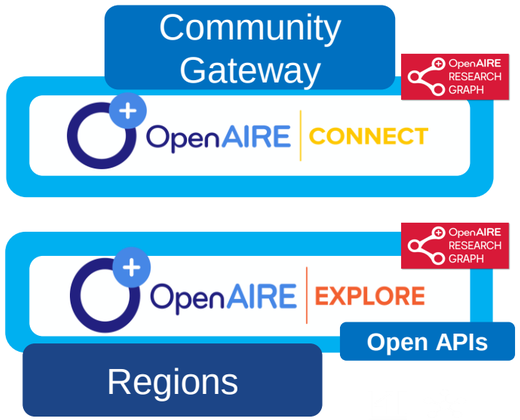

OpenAIRE

European Commission

# DISCOVER

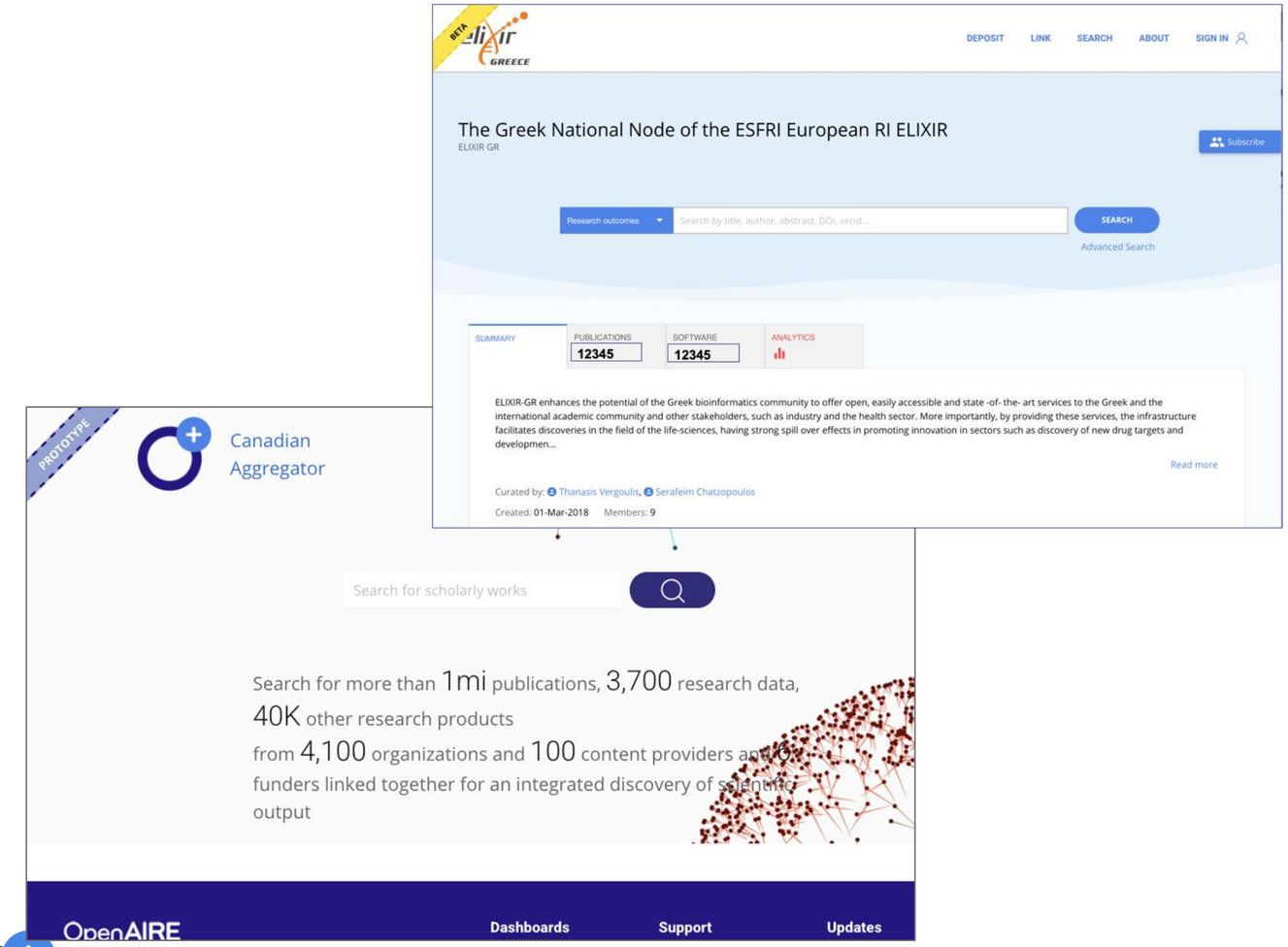
## Portals

- Discovery and stats for Communities
- Discovery and stats for for Countries

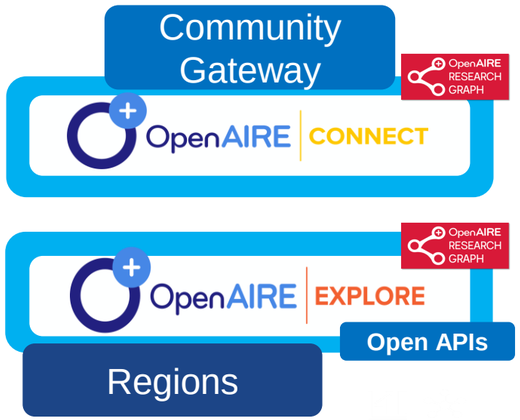



TRL9
TRL8
Prototype
Third-party TRL9

# DISCOVER




## 5 communities (7 in BETA)




OpenAIRE

European Commission

DEVELOP

# APIs and access

## Access points

REST APIs with tokens

Linked Open Data

Dumps in Zenodo

## Numbers

EXPLORE: Average API hits per month 1,7Mi

(54Mi per year)

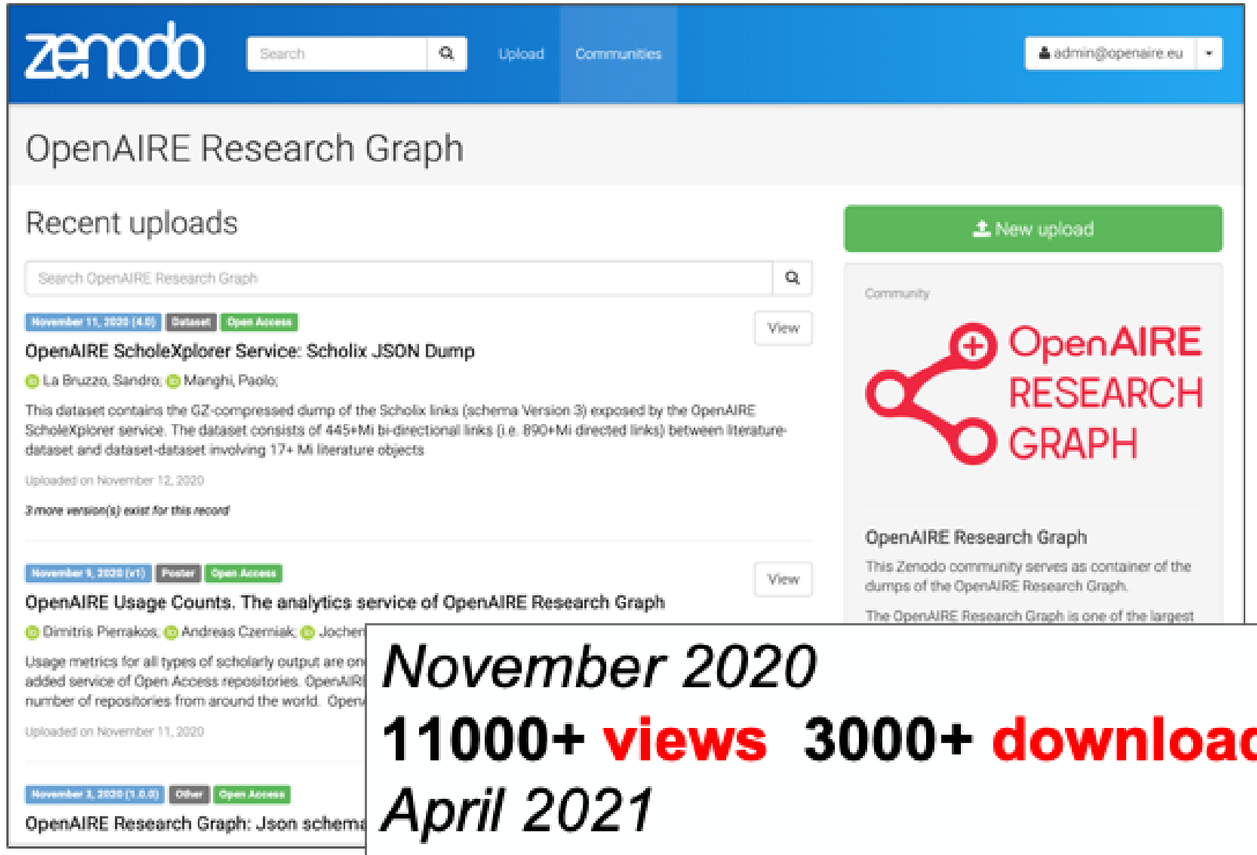
Scholexplorer: 4Bi per year

http://develop.openaire.eu

# OpenAIRE Graph Dumps

DEVELOP

- OpenAIRE Research Graph
- OpenAIRE COVID-19
- OpenAIRE Communities
- OpenAIRE LOD
- DOIBoost
- Scholexplorer
- OpenAIRE RDF PID Graph




*November 2020*

**11000+ views 3000+ downloads**

*April 2021*

**14700+ views 6100+ downloads**

https://zenodo.org/communities/openaire-research-graph

# Consumers

DEVELOP

EUROPEAN OPEN SCIENCE CLOUD

The Graph as EOSC Resource Catalogue

DARIAH-EU · EGI · EPOS · elixir

Research infrastructures

AGINFRA · instruct ERIC · SDSN GREECE · CLARIN

ORCID Early 2021

Connecting Research and Researchers

SciVal

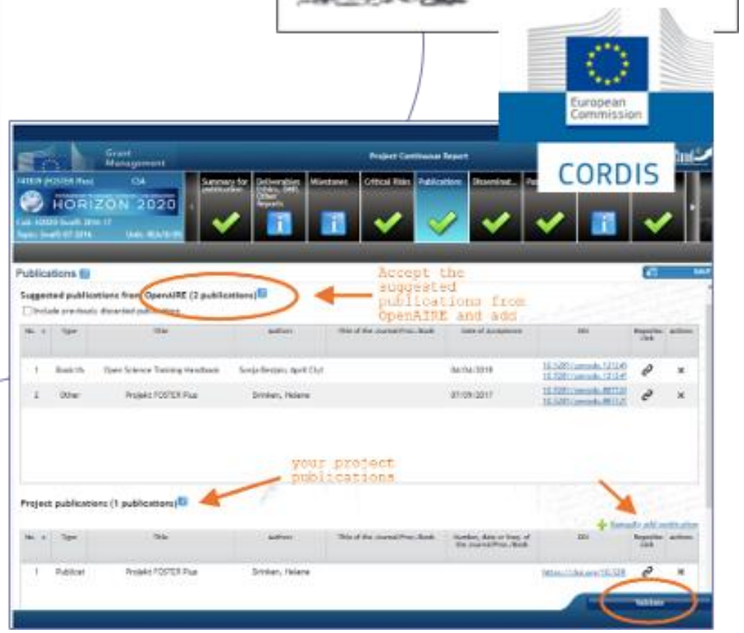
2030 CONNECT

a United Nations online technology platform for the SDGs

OpenAIRE RESEARCH GRAPH

Scopus

Mendeley




DSPACE

eprints repository software

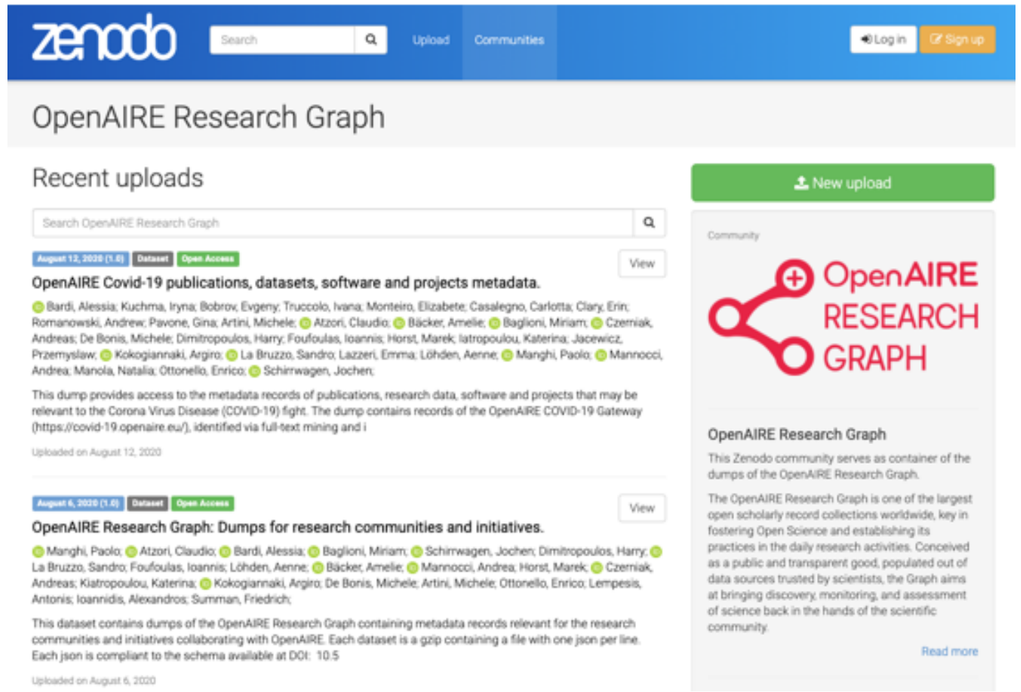
OpenAIRE

European Commission

# OpenAIRE Research Graph: Open APIs and Dumps

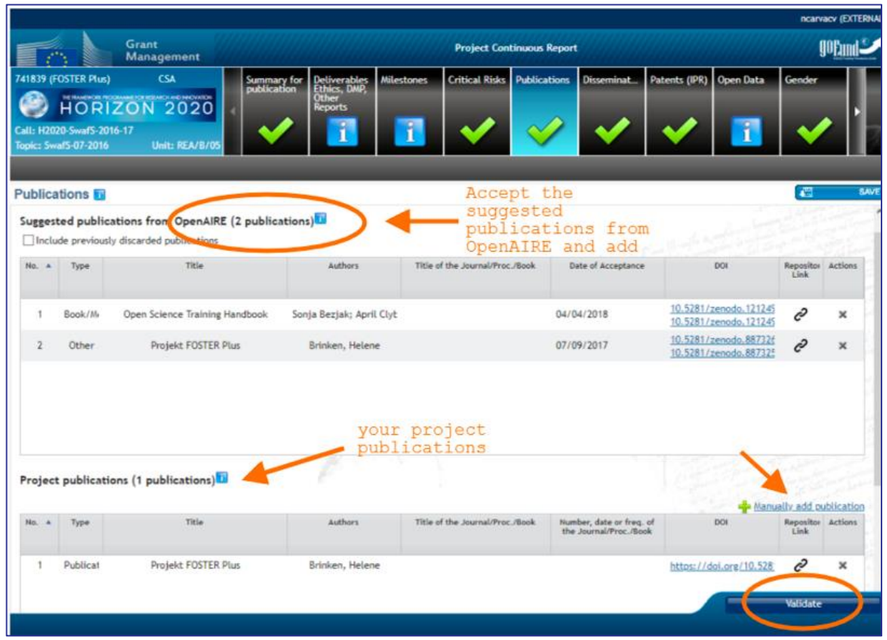
## Supporting third-parties at building added-value services

- http://develop.openaire.eu
- https://zenodo.org/communities/openaire-research-graph

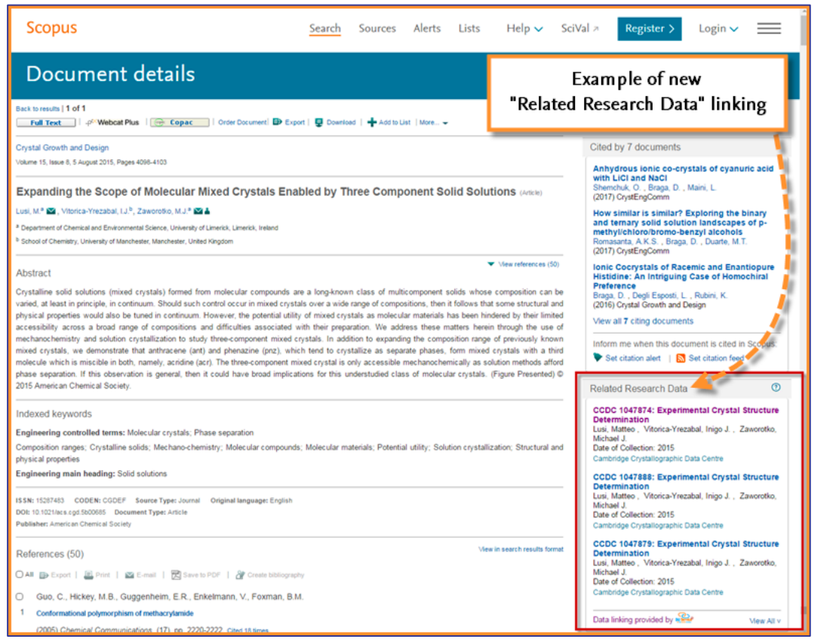


# APIs integration with third-party services

## EC SYGMA




## Scopus




## ORCID

OpenAIRE

Feed

ORCID

OpenAIRE | DEVELOP

OpenAIRE

European Commission

ScholeXplorer

OpenAIRE | PROVIDE

# Call for collaboration

## *Improving quality and completeness the Graph*

Data and infrasructure sharing

OpenAIRE support to access to Graph data (metadata and full-text) and access to full-text mining infrastructure

Embedding improvements as part fo the data provision workflow of the Graph

Devising new Open Science services for the community

## *Improving service provision*

OpenAIRE

European Commission

graph.openaire.eu

# Thank you!

**Paolo Manghi (OpenAIRE/CNR-ISTI)**

paolo.manghi@openaire.eu

graph.openaire.eu