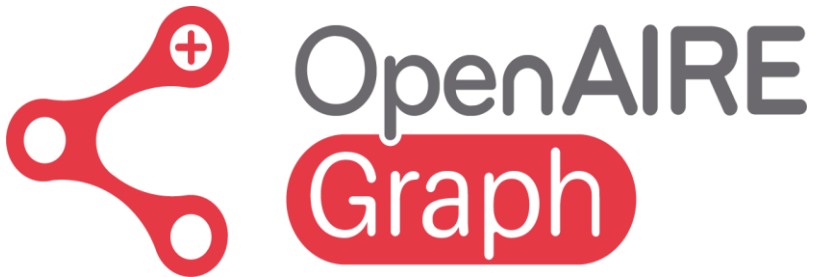


Science. Set free.

The OpenAIRE Graph:
*supporting discovery and
monitoring of Open Science
research with open data*

Paolo Manghi
OpenAIRE AMKE

The OpenAIRE Graph: the basics



Scholarly Knowledge Graph

A collection of metadata describing **entities** of the research lifecycle and **relationships** among them

Open

Complete

De-duplicated

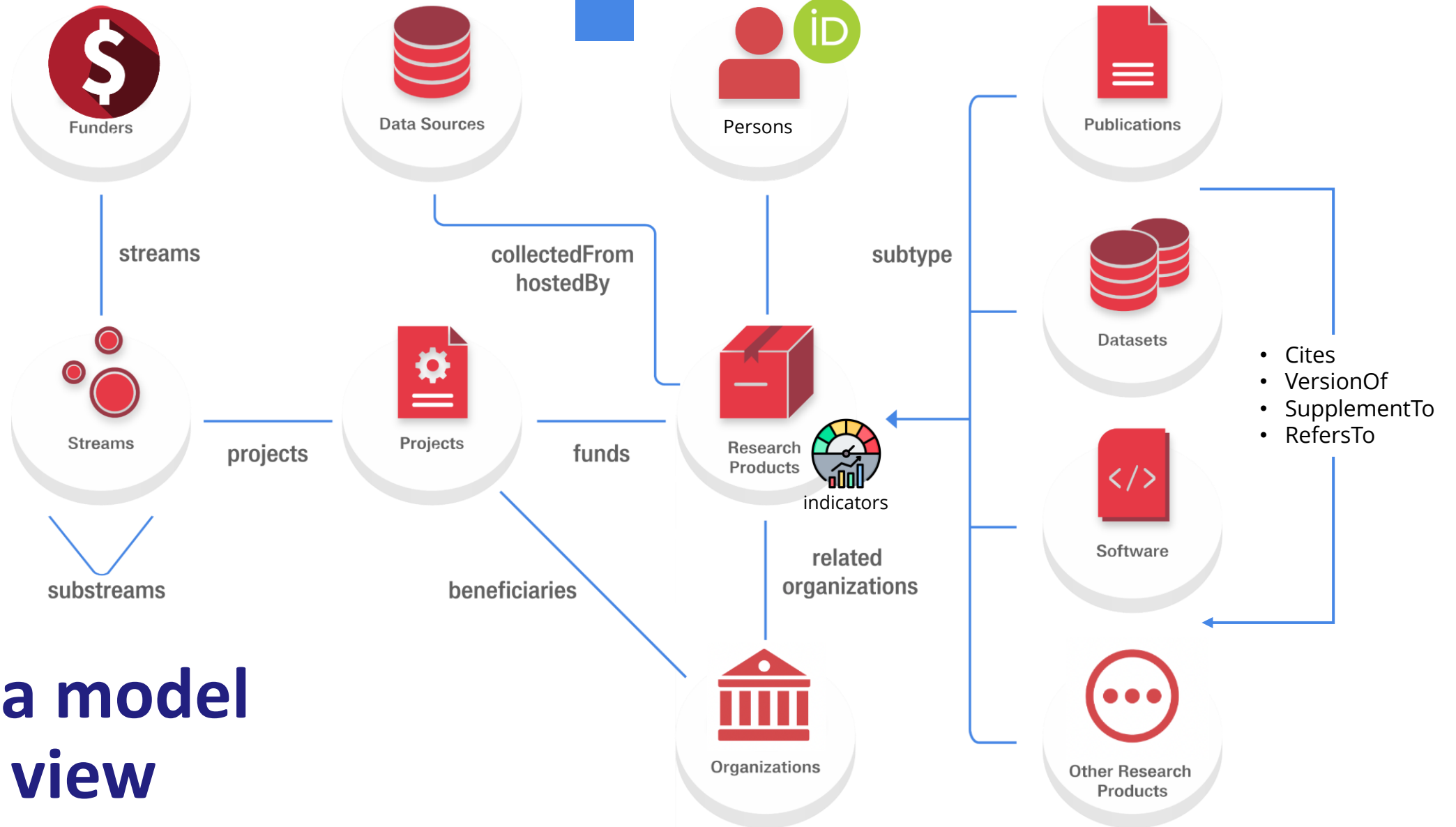
Transparent

Participatory

Decentralized

Trusted





The data model moon's view

Some data model features



Embeds indicators such as citations (from OpenCitations Meta, MAG, DataCite, and OpenAIRE), APCs (from OpenAPC), COUNTER metrics (from OpenAIRE UsageCounts), and other indicators (e.g. popularity and influence from BIP!)

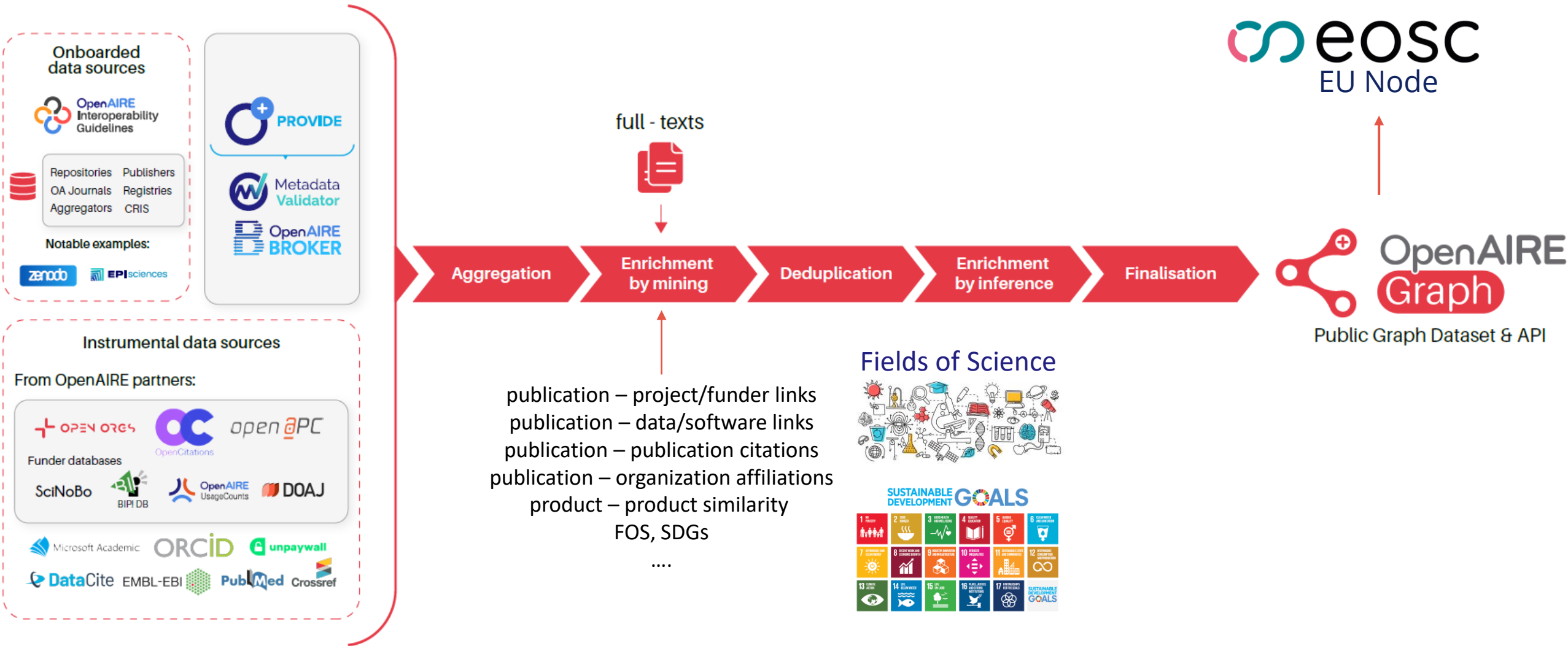


Persistent identifiers include today DOI, ArXiv, PubMed, Handles for *publications*; DOI, accession numbers, handles for *research data*; DOI, handles, and software heritage IDs for *research software* (the data model can flexibly include any PID schema!)



Stable identifiers for graph records

Provision chain



Data sources contributing to the Graph

Spotlights on data acquisition
2.5k+ direct data sources
(140K+ total)

GitHub Applications Database
Software Heritage PRESERVING TECHNICAL KNOWLEDGE
elixir bio.tools
... and more tools and software sources

frontiers
PLOS Springer
F1000 Faculty of 1000
... and more publishers

CORDIS NSF DFG
Deutsche Forschungsgemeinschaft
... and more European and international funders

unpaywall
OpenCitations Microsoft
... and more research s



OpenAIRE
Graph

DataCite re3data.org ROR ORCID
Crossref OpenDOAR FAIRsharing.org
standards, databases, policies ... and more registries

LA Referencia Red de repositorios de acceso abierto a la ciencia
SciELO DOAJ DIRECTORY OF OPEN ACCESS JOURNALS
... and more aggregators

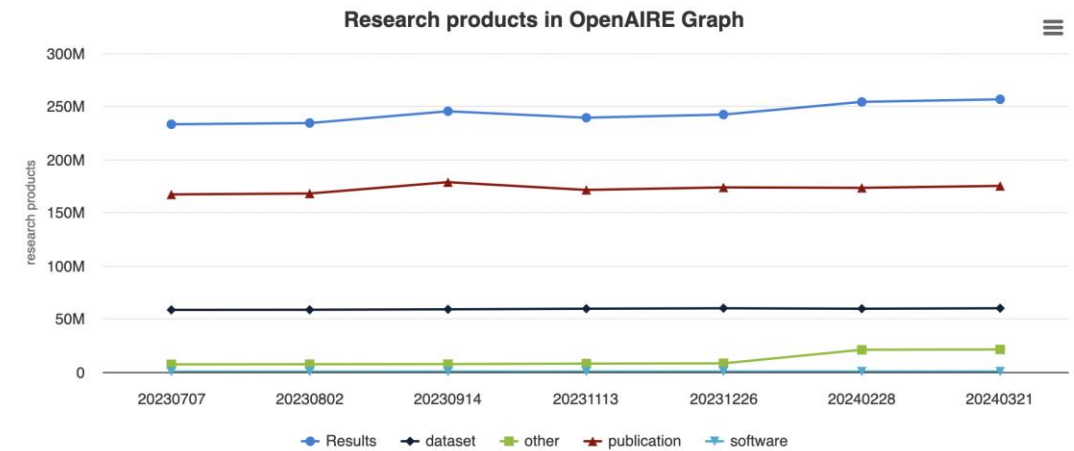
Cornell University Library arXiv.org RePEc
PubMed HAL archives-ouvertes.fr
zenodo Research. Shared.
... and more thematic and institutional repositories

DARIAH-EU elixir
EPOS EUROPEAN PLATE OBSERVING SYSTEM
ECRIN EUROPEAN CLINICAL RESEARCH INFRASTRUCTURE NETWORK
EOSC-Life
egi
... and more einfra and RI sources

OpenAIRE Graph in numbers (25th of Oct., 2024)

- 194Mi publications
- 405k software
- 64Mi research data
- 23Mi other research products
- 186 funders (30 with projects)
- 3.4Mi grants
- 318K organizations

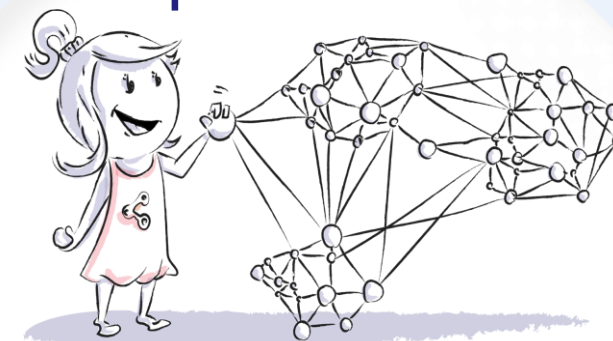
Growth over time



OpenAIRE
Graph

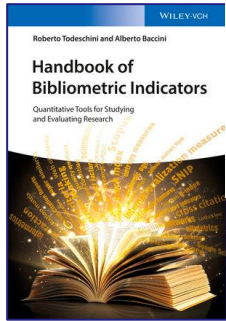
OpenAIRE Graph in numbers

- **Links from pubs to products**
 - 2.1Bi pubs-pubs links (citations)
 - 138Mi pubs-data links
 - 437K pubs-software links
- **Links to grants**
 - 5.4Mi pubs-grant links
 - 862K data-grant links
 - 16.3K software-grant links
- **Affiliations**
 - 172Mi pubs-organization links (affiliations)
 - 3.7Mi data-organization links (affiliations)
 - 166K software-organization links (affiliations)
- 432Mi pubs-FoS (77Mi pubs with FoS)



Consumers

Researchers



"Transparent peer review will serve as a valuable tool for readers to see how quality control is done at *Quantitative Science Studies*."

– Ludo Waltman, Editor-in-Chief

Service Providers/Companies



Scopus

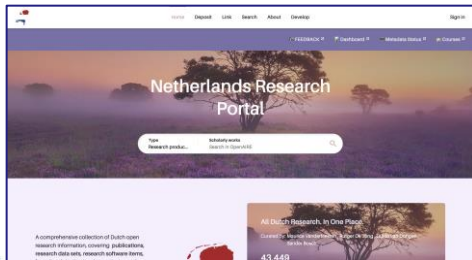


Reporting to funders,
Discovery of pubs, data, software
Assessment, monitoring (FAIRness,
Open Science, Open Access)

Organizations



THE WORLD BANK



Funders

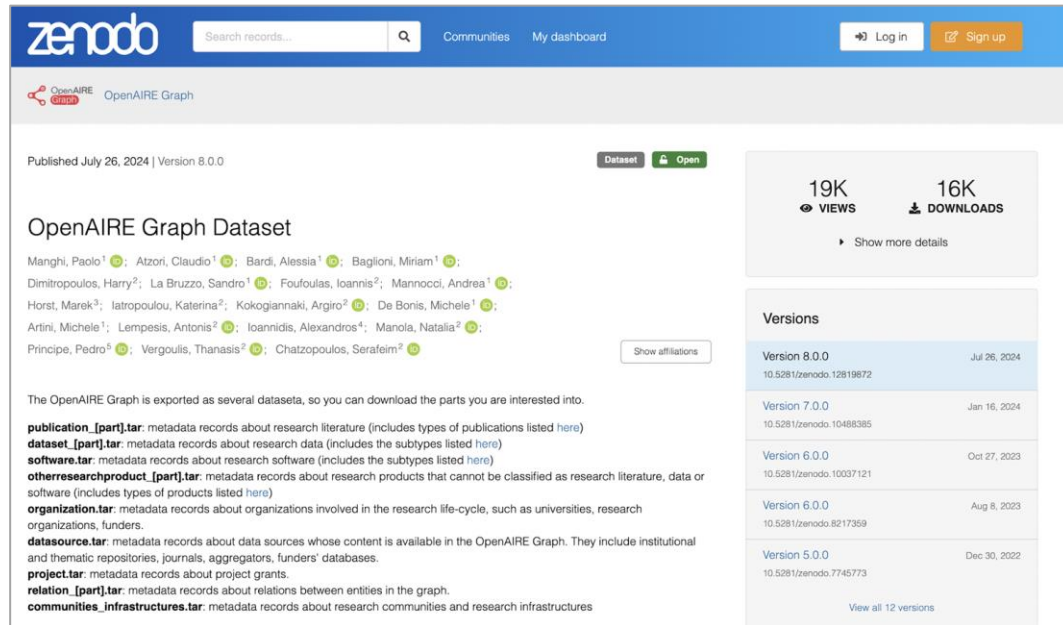


European Commission



Access to the Graph

- Dataset in Zenodo
- API on documentation web site (graph.openaire.eu/docs)



zenodo Search records... Communities My dashboard Log in Sign up

OpenAIRE Graph OpenAIRE Graph

Published July 26, 2024 | Version 8.0.0 Dataset Open

OpenAIRE Graph Dataset

Manghi, Paolo¹; Atzori, Claudio¹; Bardi, Alessia¹; Baglioni, Miriam¹; Dimitropoulos, Harry²; La Bruzzo, Sandro¹; Foufoulas, Ioannis²; Mannocci, Andrea¹; Horst, Marek³; Iatropoulou, Katerina²; Kokogiannaki, Argiro²; De Bonis, Michele¹; Artini, Michele¹; Lempesis, Antonis²; Ioannidis, Alexandros⁴; Manola, Natalia²; Principe, Pedro⁵; Vergoulis, Thanasis²; Chatzopoulos, Serafeim²

Show affiliations

The OpenAIRE Graph is exported as several datasets, so you can download the parts you are interested into.

- publication_[part].tar**: metadata records about research literature (includes types of publications listed [here](#))
- dataset_[part].tar**: metadata records about research data (includes the subtypes listed [here](#))
- software.tar**: metadata records about research software (includes the subtypes listed [here](#))
- otherresearchproduct_[part].tar**: metadata records about research products that cannot be classified as research literature, data or software (includes types of products listed [here](#))
- organization.tar**: metadata records about organizations involved in the research life-cycle, such as universities, research organizations, funders.
- datasource.tar**: metadata records about data sources whose content is available in the OpenAIRE Graph. They include institutional and thematic repositories, journals, aggregators, funders' databases.
- project.tar**: metadata records about project grants.
- relation_[part].tar**: metadata records about relations between entities in the graph.
- communities_infrastructures.tar**: metadata records about research communities and research infrastructures

19K VIEWS 16K DOWNLOADS Show more details

Versions

Version	Published
Version 8.0.0 10.5281/zenodo.12819872	Jul 26, 2024
Version 7.0.0 10.5281/zenodo.10488385	Jan 16, 2024
Version 6.0.0 10.5281/zenodo.10037121	Oct 27, 2023
Version 6.0.0 10.5281/zenodo.8217359	Aug 8, 2023
Version 5.0.0 10.5281/zenodo.7745773	Dec 30, 2022

View all 12 versions

The OpenAIRE APIs

Our APIs can be accessed anonymously and are free of charge. Higher rate limits are available for signed in users and registered services. For more information please check [here](#).

Broker

API to enrich metadata for repositories, publishers and aggregators

[Documentation](#)
[Swagger](#)

Selective access

Access via HTTP

[Projects](#)
[Research Products](#)

- [Publications](#)
- [Research data](#)
- [Research Software](#)
- [Other Research Products](#)

Linked Open Data

[Documentation](#)
[OpenAIRE LOD Ontology](#)
[OpenAIRE RDF dump](#)
[SPARQL endpoint](#)

Applications: discovery and research assessment

OpenAIRE EXPLORE and CONNECT

<https://explore.openaire.eu/>

OpenAIRE | EXPLORE

Search Deposit Link Data sources

Sign in

Discover open linked research.

A comprehensive and open dataset of research information, including publications, 18m research data, 304k research software sources, linked to 3m grants and 178k organizations. All linked together through citations and semantics.

Type
All Content

Scholarly works
Search in ...

Try browsing by:

 SUSTAINABLE DEVELOPMENT GOALS (SDGs) →

 FIELDS OF SCIENCE (FOS) →

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



Home Deposit Link Search About Develop

Sign in

Digital Humanities and Cultural Heritage

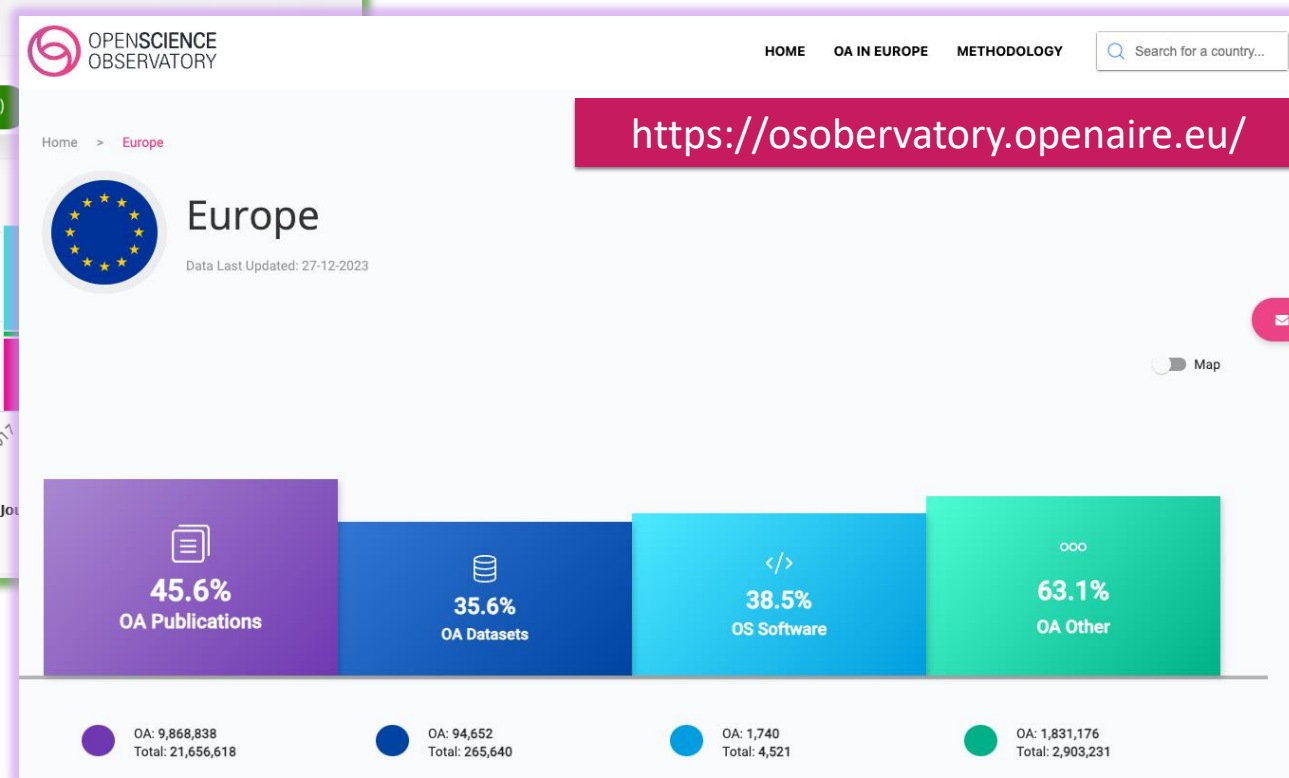
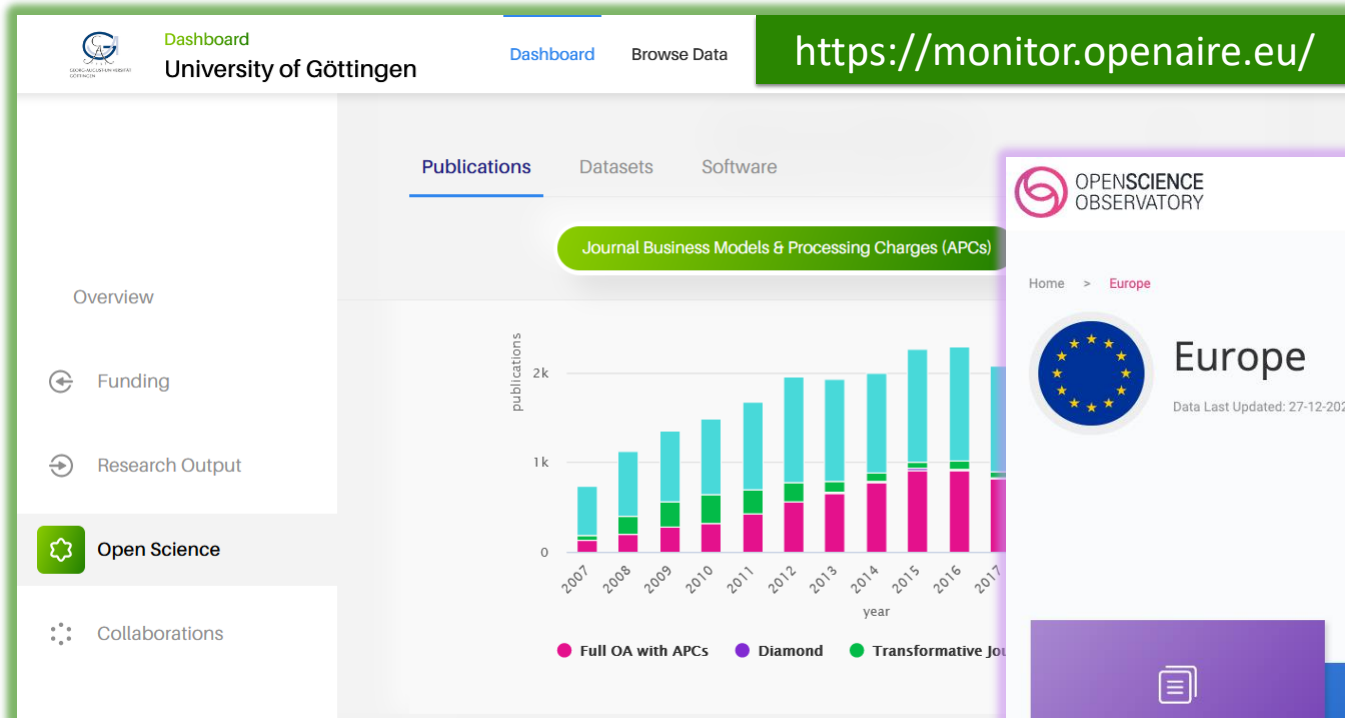
Join

Type
Research products

Scholarly works
Search in OpenAIRE

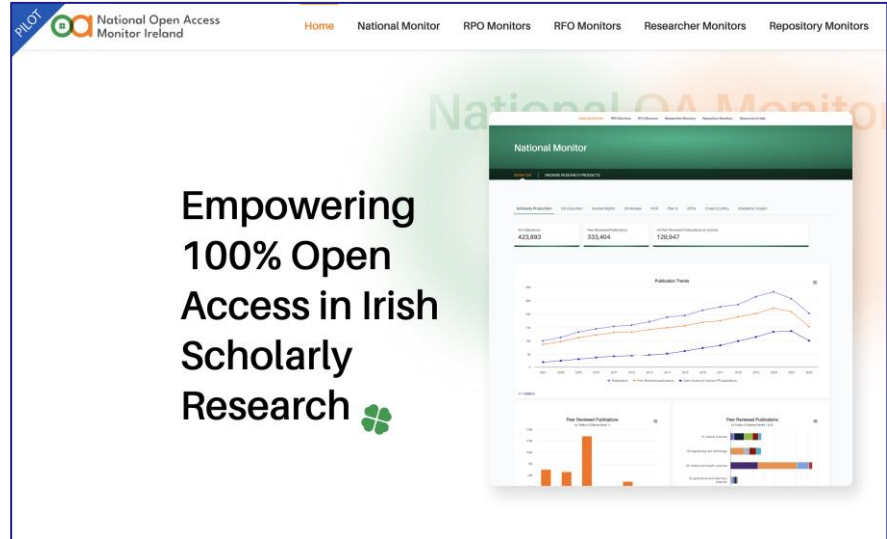


OpenAIRE MONITOR and OSmonitor



The Irish OS monitor tender

<https://oamonitor.ireland.openaire.eu>



Onboarded data sources

OpenAIRE Interoperability Guidelines

Repositories Publishers
OA Journals Registries
Aggregators CRIS

Notable examples:

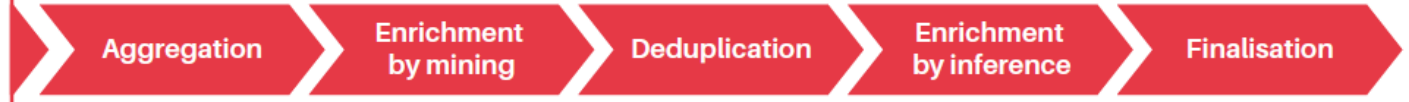
zenodo EPIsciences

PROVIDE

Metadata Validator

OpenAIRE BROKER

full - texts



Public Graph Dataset & API

Instrumental data sources

From OpenAIRE partners:

OPEN ORGS OpenCitations open@PC

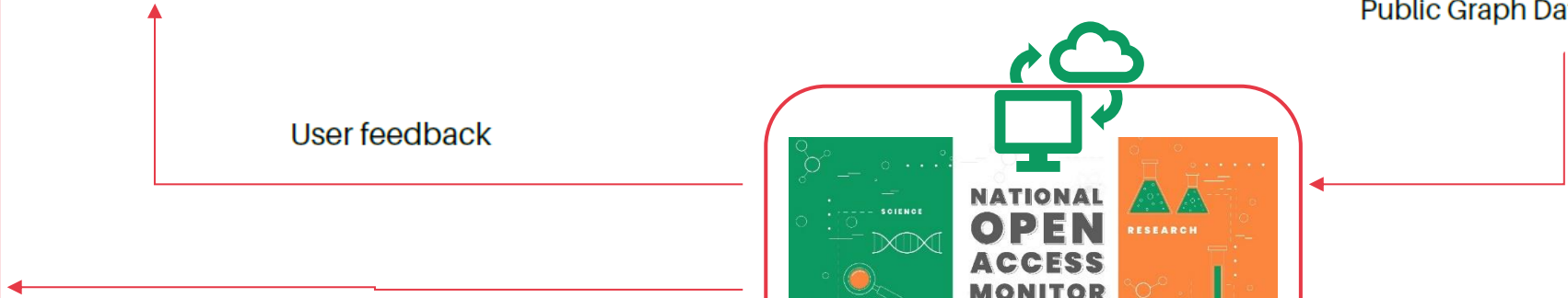
Funder databases

SciNoBo BIPI DB OpenAIRE UsageCounts DOAJ

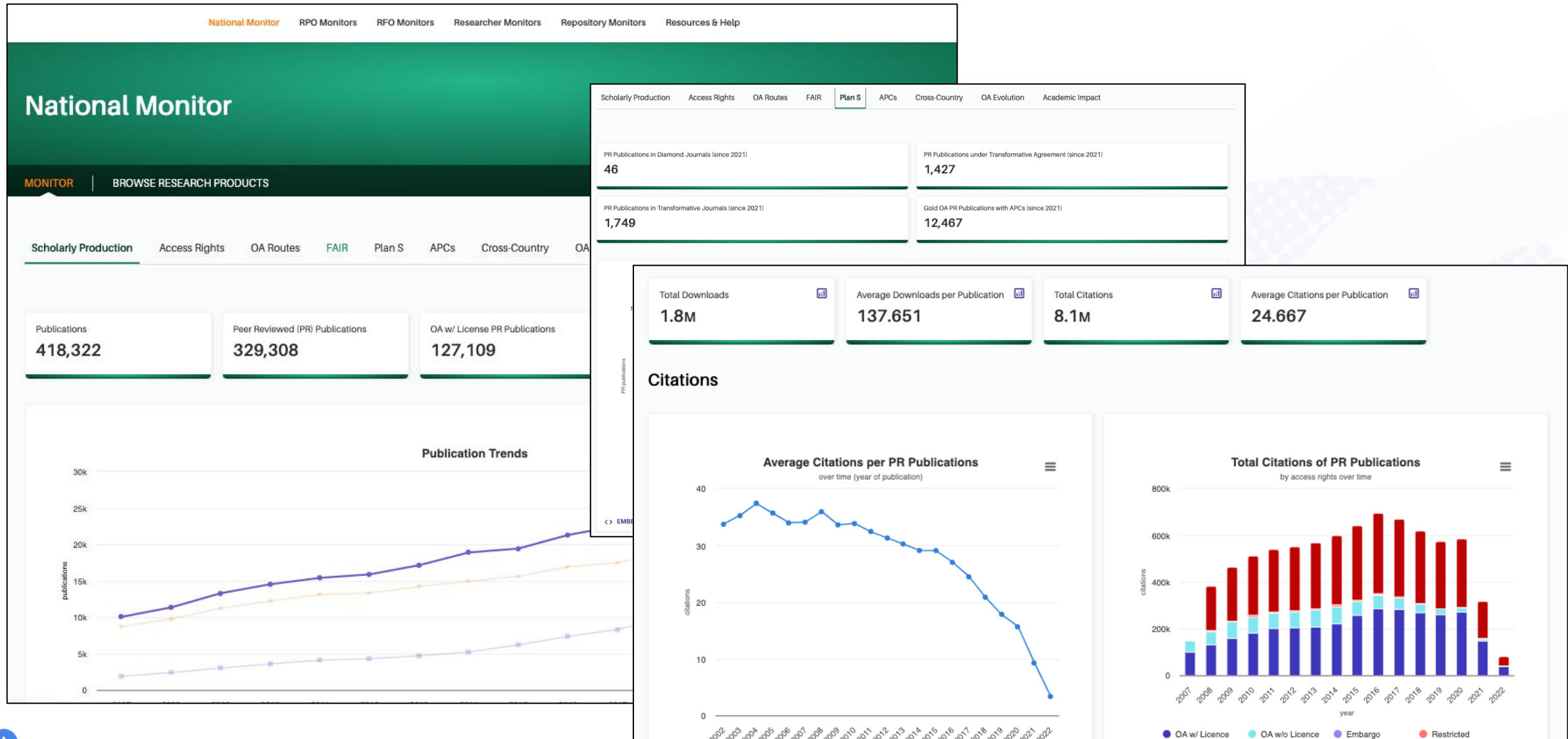
Microsoft Academic ORCID unpaywall

DataCite EMBL-EBI PubMed Crossref

User feedback



The Irish OS monitor tender



For and by the community

The Graph is public good!

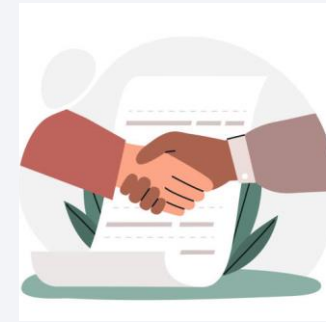
In the hands of the Community



Community governance

Members of OpenAIRE can evaluate results and steer Graph's roadmap and vision by contributing to OpenAIRE Steering Committees

Members can join the technical team and contribute to the development and operation of the Graph



Community alignment

Synergies with repository managers, libraries, service providers at the global level to ensure alignment of data models, metadata, vocabularies at the global level (***EOSC guidelines, OpenAIRE Guidelines***, SKG interoperability framework)

Community involvement



Seeking collaborations to enrich quality and coverage of data

We're interested to integrate solid research results in the domain (e.g., SDGs inference)

We're interested to include high-quality data sources (e.g., OpenCitations integration)



End-users feedback

We care adapting roadmap to match community expectations, e.g. research organizations, funders, companies

We care about identifying and fixing errors

We acquire Graph data enrichment from users



Improving scholarly communication infrastructure and workflows

We provide tools for enriching data sources with metadata information they are missing

We provide tools to facilitate researchers, publishers, service providers in “publishing” research trusted scientific outcomes with high-quality metadata

OpenAIRE Graph: take home message



It's the largest Open Science-focused graph

Broader **coverage** of publications, research data, and research software, together with links between them



It's open

Accessible via APIs and datasets in Zenodo under **CC-0**



It's transparent

We document the process of generation and track **provenance**



It's a public good

A **public infrastructure** (like highways)



It's participatory

Welcomes data/methods contributions from service providers, organizations, and researchers

For transparency

<https://graph.openaire.eu/docs/>

Documentation

Overview

Data model >

Public API 

Downloads >

Graph production workflow >

Relevant publications

License

Versions & changelog

Helpdesk 

For enquiries

<https://openaire.flarum.cloud>

Forum

The screenshot shows the OpenAIRE Graph Forum homepage. At the top left is the OpenAIRE Graph logo, with 'Home' and 'Guidelines' links. A search bar labeled 'Search Forum' is on the right, along with notification and user profile icons for 'paolomanghi'. Below the header is a navigation bar with a 'Start a Discussion' button and tabs for 'All Discussions', 'Private Discussions', 'Following', 'Categories', and 'Achievements'. The main content area features several topic cards: 'APIs' (Authenticated requests, Unauthenticated requests, ScholeXplorer API, Dspace API, Eprints API, Broker API), 'Data model & Interoperability' (Products, Organisations, Projects, Funders, Communities, Persons, Relationships, Interoperability), 'Data sources' (Repositories, Registries, OA Journals, Publishers, Aggregators, CRIS, Instrumental Data sources), 'Graph datasets' (Full graph, Sub-graphs, Beginner's kit, ScholeXplorer), 'Graph Production Workflow & Provenance' (Aggregation, Enrichment by mining, Deduplication, Enrichment by inference, Finalisation), and 'Resources' (FAQ, Tutorials, etc). A 'Feedback' card at the bottom left invites users to share feedback about the community.

Thank you for your attention!

paolo.manghi@openaire.eu