@openaire_eu

The OpenAIRE Graph:

Supporting Research Intelligence through Open Data

Andrea Mannocci
Institute of Information Science and Technologies, CNR (CNR-ISTI)

Consiglio Nazionale delle Ricerche

ISTITUTO DI SCIENZA E TECNOLOGIE DELL'INFORMAZIONE "A. FAEDO"

infoscience

OpenAIRE

The OpenAIRE Graph team

OpenAIRE

From Business Intelligence to Research Intelligence

	📈 Business Intelligence (BI)	🔬 Research Intelligence (RI)
🌐 **Context**	Business operations (private sector)	Research and innovation ecosystems (**public** + private)
🗄️ **Data**	ERP • CRM • transactional databases	Publications • datasets • software projects • funding • organisations
📊 **Analytics**	Sales metrics Customer analytics Financial dashboards Market research	Monitoring dashboards Bibliometrics, Scientometrics Collaboration networks Funding flows Research impact
🎯 **Goal**	Improve **business performance and efficiency**	Support **metascience, research on research, science policy, research strategy, and evaluation**

OpenAIRE

If research is public, why is Research Intelligence closed?

- The world invests in research as a **public good**
- Yet the intelligence built on top of that research is still too often controlled by **closed systems**
- **Well, closed systems are closed!**
  - They decide what and how should be measured
  - They decide what can be made visible
  - Their data cannot be fully audited
  - Their methods cannot be challenged
  - Their priorities are commercial rather than public

How can the OpenAIRE Graph help?

OpenAIRE Graph: the basics

A Scholarly Knowledge Graph

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

Timely and comprehensive coverage of research outputs

Monthly updates

Precision, depth & processing

Rigorous cleaning, deduplication, and enrichment for optimal accuracy

Full-text mining of links: Research results to projects, author affiliations, and classifications

AI methods for the identification of FoS, SDGs

Robustness & openness

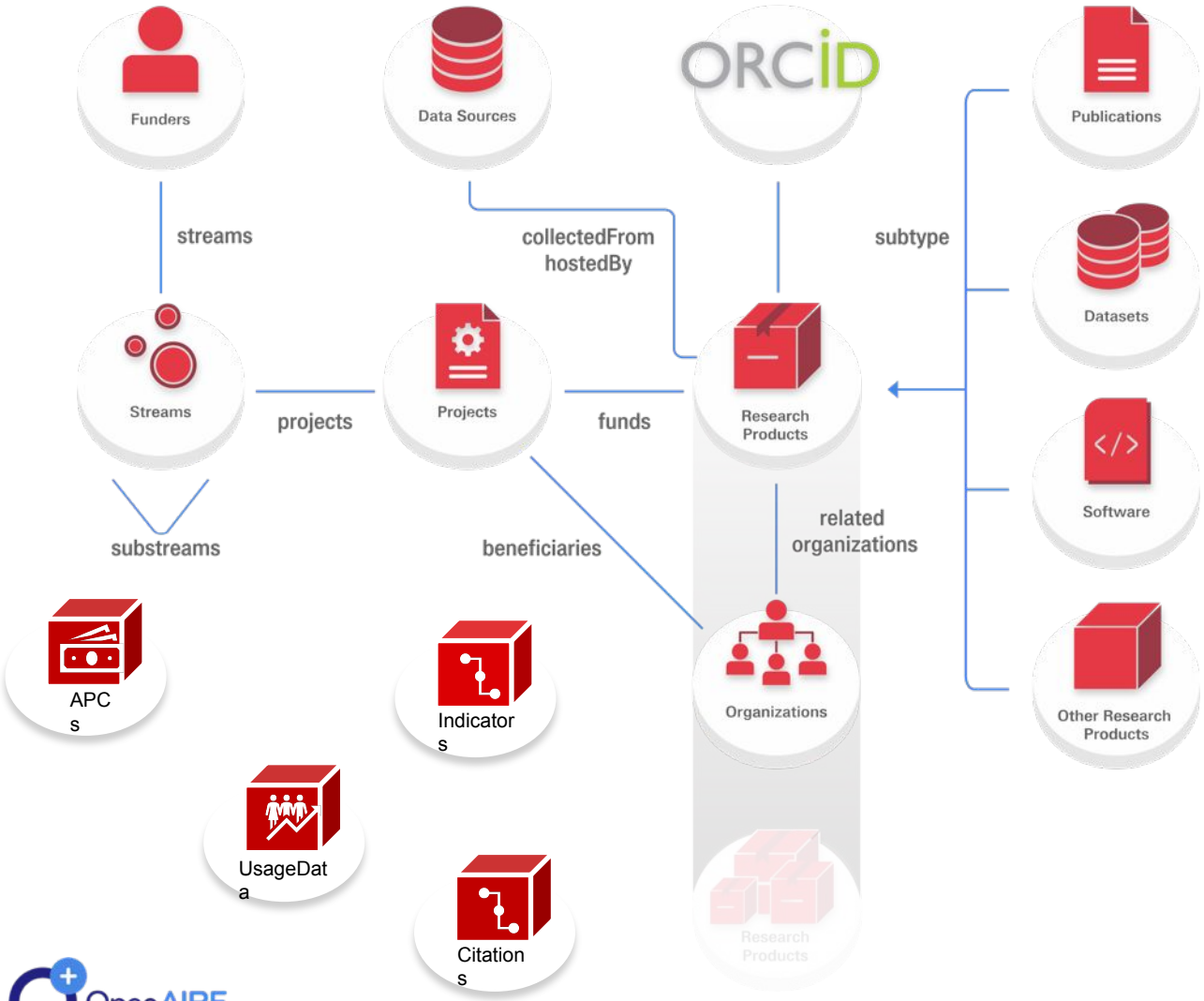
Professional infrastructure: maintenance, load balancing, backups, overseen by the OpenAIRE technology centre (ICM)

Open Data, Open Source & transparent methodologies

OpenAIRE Graph

OpenAIRE

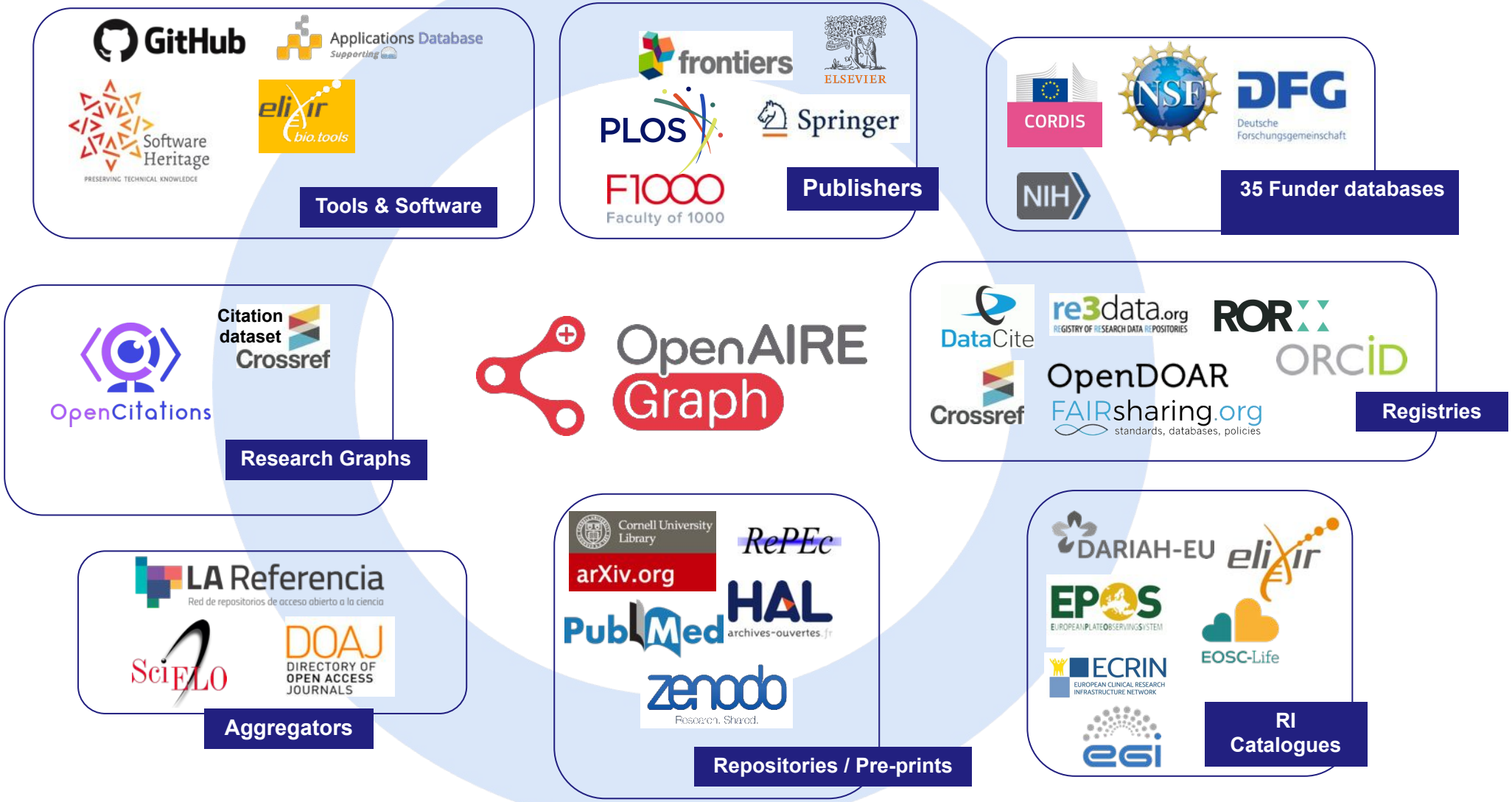
OpenAIRE Graph



- Based on **Open Science principles:** open software, open data, open APIs
- **Global** & **longitudinal** coverage
- Enriched by **AI-processes -** citation, reproducibility, FOS, SDGs, domain specific terms/ontologies

Any PID

OpenAIRE Graph Data Sources



OpenAIRE

Graph entities

Latest Statistics

Scholarly works

211,871,379
Publications

92,817,388
Data items

804,753
Software

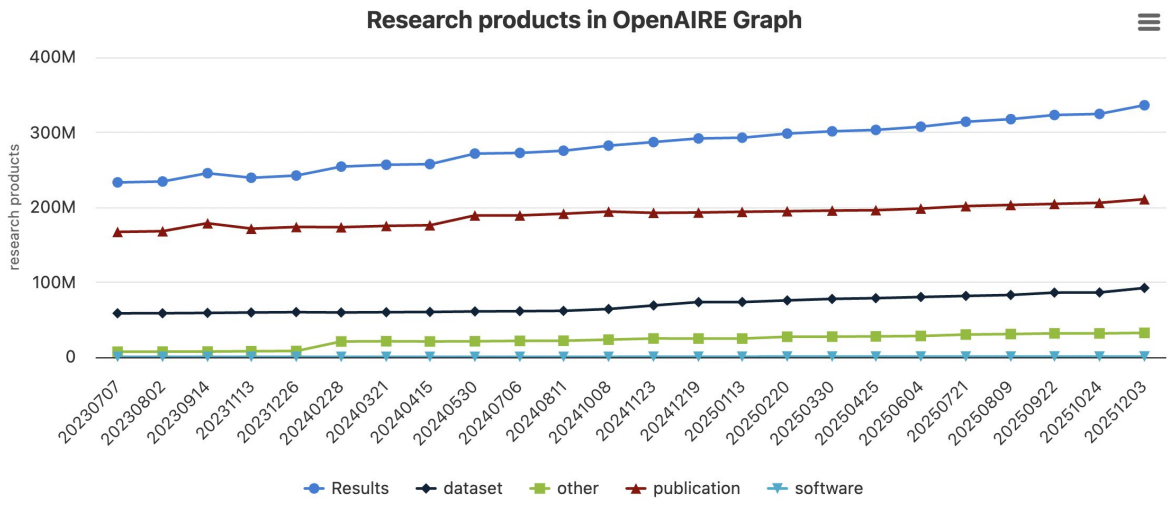
33,307,160
Other

158,874
Data sources

231
Funders

3,743,429
Grants

Growth over time



338.8 M Research Products

PUBLICATIONS. DATA. SOFTWARE. +

Enriched. Deduplicated. Connected.

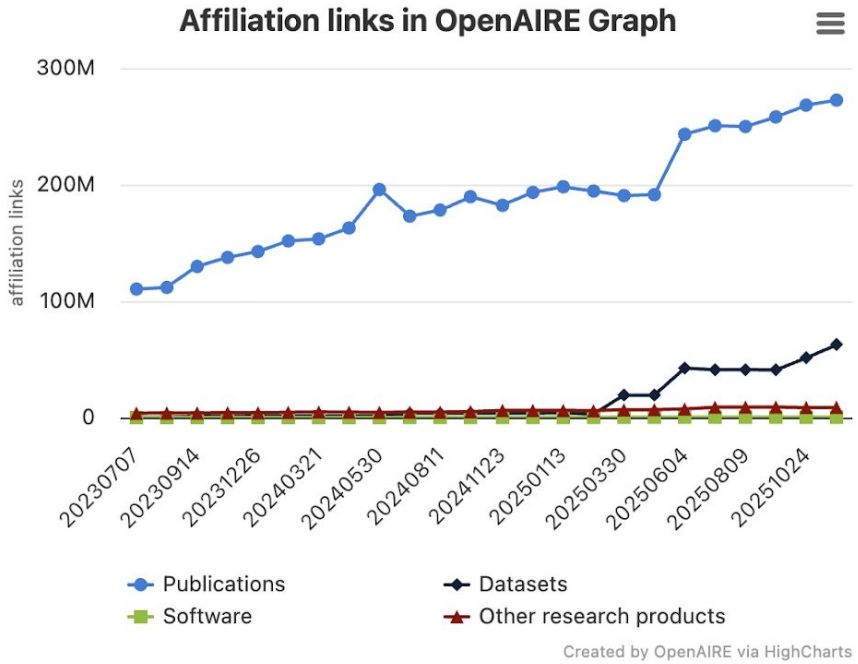
Dive Deeper →

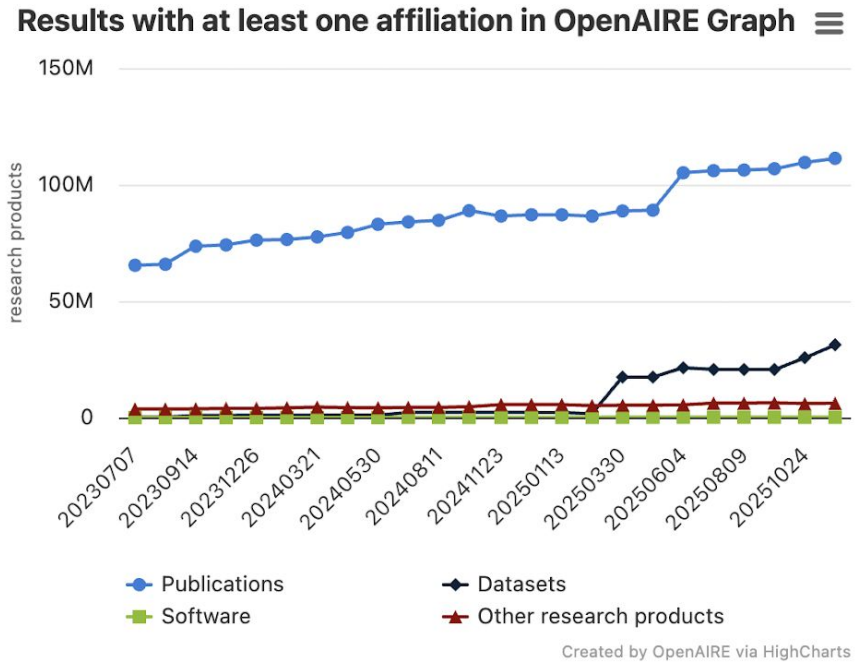
Relations

+3bi citation relations

ScholeXplorer
The Data Literature Interlinking Service

pub-pub
pub-data
pub-software
data-data
....





OpenAIRE

OpenAIRE Graph

Open Data and Analytics

- **Open REST APIs**
  - Graph search (https://graph.openaire.eu/docs/apis/graph-api)
  - Graph relations (https://graph.openaire.eu/docs/apis/scholexplorer/api)
- **Data Access Portfolio (more on this later)**
  - Datasets on Zenodo.org (https://graph.openaire.eu/docs/category/downloads)
  - Beginner's kit (https://graph.openaire.eu/docs/downloads/beginners-kit)
  - Google BigQuery (https://graph.openaire.eu/docs/cloud-access)
- **LLM interfaces**
  - Data + AI agents (work in progress)

For more details: graph.openaire.eu/docs

OpenAIRE

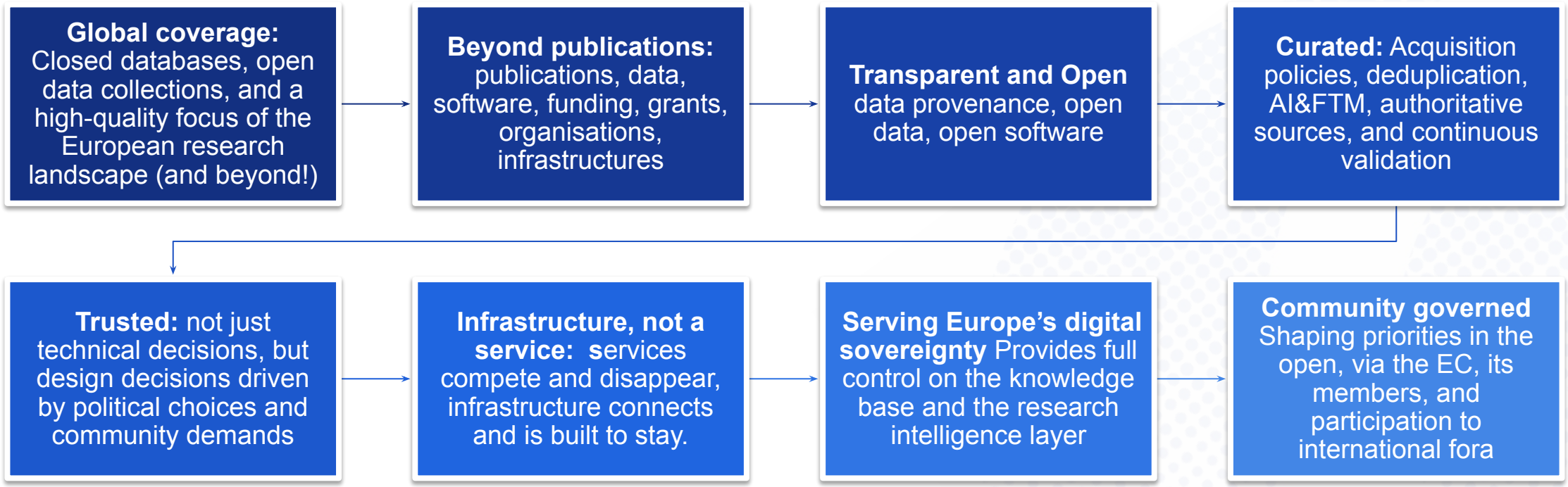
Why transitioning to the OpenAIRE Graph?

- **Open Science** needs **open data** and **open intelligence**
- **Public investment** requires **public control**
- **Public infrastructure** is how **values become durable**

Invest in open infrastructures and help shape the research intelligence layer of European (and global) research

OpenAIRE

The OpenAIRE Graph and Research Intelligence



OpenAIRE

Going beyond publications

“we are more than our publication track record”

New classes of eligible publishing venues

- Institutional repositories
- Catch-all repositories
- National & Thematic aggregators
- CRIS systems
- Registries of authors/organisations

Jisc

OpenDOAR

6,000

FAIRsharing.org
standards, databases, policies

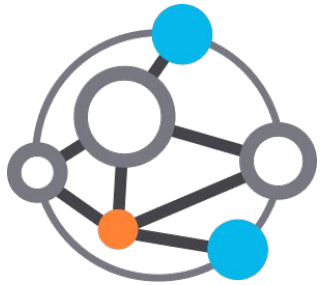
2,000

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

3,500

Research product typologies and bibliotyping

Traditional typologies: peer-reviewed publications + preprints + open access tags + pub-pub citations



Coarse-grained bibliotyping
A **graph entry** can be a paper, a deliverable, a report, with no distinction

Open Science typologies: traditional typologies + dissemination products + research data + research software + data and software citations

COARA

OpenAIRE Graph

Fine-grained bibliotyping
(A **deliverable** is not an **article**)

COAR Confederation of Open Access Repositories

Metadata curation and quality

”Quantity matters, but without quality it’s mostly noise"

OpenAIRE Graph provision chain

Onboarded data sources

OpenAIRE Interoperability Guidelines

Repositories · Publishers
OA Journals · Registries
Aggregators · CRIS

Notable examples: zenodo · EPIsciences

PROVIDE
Metadata Validator
OpenAIRE BROKER

Instrumental data sources

From OpenAIRE partners:

OPEN ORGS · OpenCitations · openAPC
Funder databases
SciNoBo · BIPI DB · OpenAIRE UsageCounts · DOAJ
ORCID
DataCite · arXiv · PubMed · Crossref

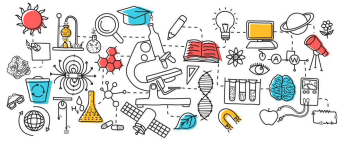
~2.2k direct data sources (155K indirectly via aggregators)

70Mi PDFs (80K per day)

full - texts

publication – project/funder links
publication – data/software links
publication – publication citations
publication – organization affiliations
product – product similarity
FOS, SDGs
….

Fields of Science



SUSTAINABLE DEVELOPMENT GOALS

Aggregation → Enrichment by mining → Deduplication → Enrichment by inference → Finalisation

OpenAIRE Graph

Public Graph Dataset & API

User feedback

OpenAIRE added-value services

Discover: EXPLORE · CONNECT · ScholeXplorer
Assess: OPENSCIENCE OBSERVATORY · MONITOR
Manage Data: argos

Third -party services

EUROPEAN OPEN SCIENCE CLOUD
CORDIS · SyGMa
ELSEVIER · eprints
AUTHORSERVICES Supporting Taylor & Francis authors

OpenAIRE

Digital Sovereignty via community ownership

OpenAIRE

Sovereignty depends on control of its research intelligence layer

- **Without a public, governed knowledge base, the global scientific community outsources strategic decisions** about science, innovation, and assessment
- **Autonomy requires control** on how knowledge is represented and interpreted
- **Avoid dependency** from commercial solutions or **opaque** representations
- Keep **ownership** on strategic choices and interpretability in our hands

OpenAIRE

Who is OpenAIRE

- Non-profit organisation, established **Oct 2018 (AMKE, Greece-based)**
- **European Scholarly Communication e-Infrastructure**
- **Federated.** Bringing together **human capital** and advanced **ICT services**
  - A **network of experts** from major national organisations (National Open Access Desks) in operation since 2009
  - A **catalogue** of scholarly communication services

51 members

36 countries

Global Collaborations: Latin America, Canada, Korea, China,…

OpenAIRE

Open Scholarly Communication Infrastructure

Co-shaping & implementing Open Science. In Europe and beyond.

National ecosystems, institutions, funders shape priorities
Proactive engagement of communities and fast responses

Participation and collaboration builds cumulative trust!

Network
Connecting people

Services
Implementing

Training
Building Capacities

OpenAIRE

Interoperability and community consultation

Examples

- Scholix.org
- Scholarly Knowledge Graphs Interoperability Framework
- DMP data models

- Open Infrastructures

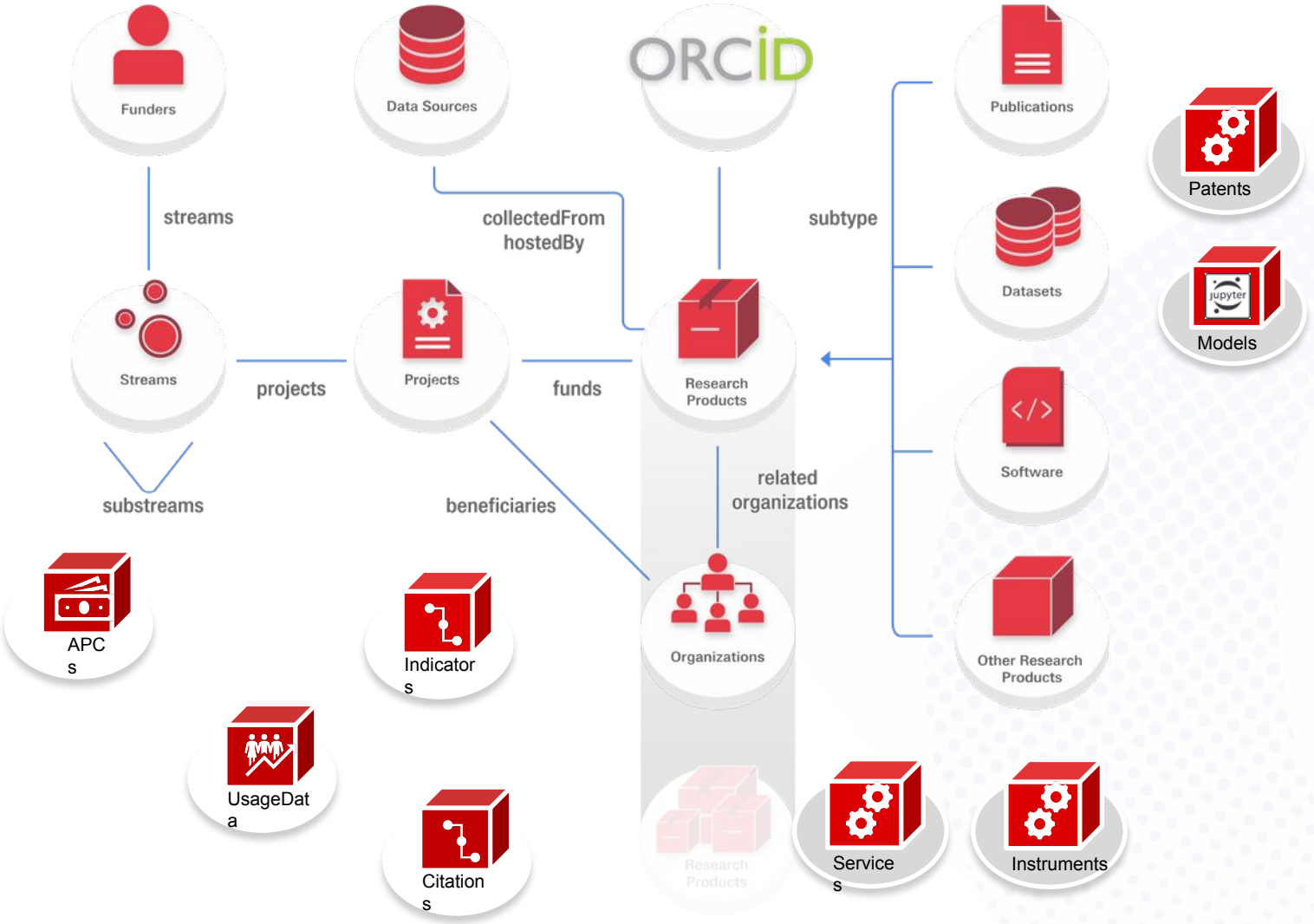
- Catalogues Interoperability frameworks

- Platform integration

- OpenAIRE guidelines for institutional repositories
- OpenAIRE guidelines for CRIS systems

- EC funding tagging
- Open Access mandates
- Open Data pilot (DMPs)

Graph data model updates



Digital Sovereignty principles

- The community steers the future and shapes the demands
- Bottom-up process with top-down facilitators
- May be slower, but brings control & transparency, and builds trust & infrastructure
- Grounds on institutional and community repositories and open access publishers
- Facilitates global alignment of publishing practices and research assessment methodologies

Infrastructure VS. Service

SKGs as pivotal in the scholarly comunication life cycle

- Improve publishing workflows
- Improve reporting workflows
- Steer Open Science publishing policies

Publishing

- Publishing venues: repositories, publisher archives, CRIS systems

Discovery

Tracking

- Aggregation of publishing sources
- Cleaning: crosswalks, data curators
- Deduplication of records
- Enrichment: AI, FTM methods, user-feedback

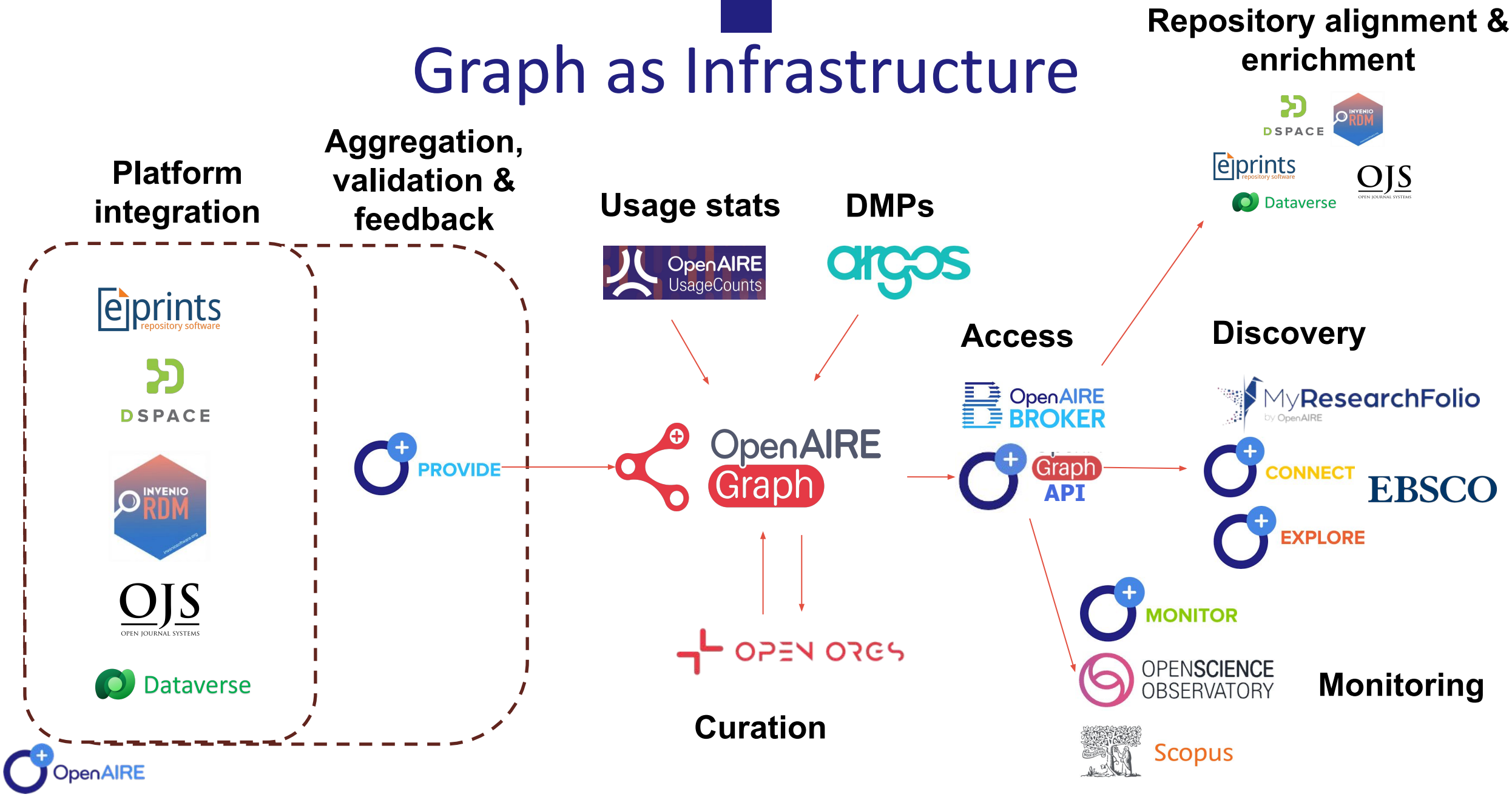
Feedback

Monitoring

- Bibliometrics
- Research assessment indicators
- Open Science trends indicators
- Monitoring Dashboards

OpenAIRE

Graph as Infrastructure



Co-shaping the scholarly communication eco-system

Traditional provision chain

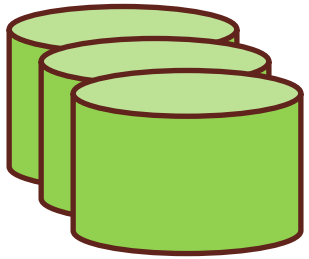
Heterogeneity

Graph chain
- Cleaning, Deduplication, Enrichment

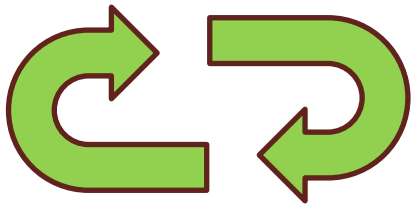
End of the stream
- No community customization (ROR.org, ORCID)
- Limited feedback loop

Heterogeneity

Graph chain
- Cleaning, Deduplication, Enrichment



Repositories, publisher archives, registries, funder databases



OpenAIRE Graph

End of the stream
- Community customization: orgs and sub–orgs not in ROR.org, local author IDs (under testing)
- Improved feedback loop

Community engagement
- OpenAIRE PROVIDE aaS
- OpenAIRE OpenOrgs
- OpenAIRE PROVIDE Broker
- Publishing practices: anomaly detection

OpenAIRE PROVIDE aaS: delegation of aggregation

Communities (e.g., countries, Ris, Science Clusters, networks, EOSC Nodes) can

- Take over **the aggregation of publishing venues** metadata into the OpenAIRE Graph
- Populate a **sub-graph with integrated support of discovery gateways and monitoring dashboards** (OpenAIRE CONNECT, OpenAIRE MONITOR)
- Steering the alignment of publishing practices across a network of **publishing venues**

OpenAIRE OpenOrgs

- **Aggregation and Bridging of** PIDs of Organization Registries
  - Global registries: e.g., ROR.org
  - Funder registries: e.g., European Commission CORDIS
  - National registries: e.g., CRIS systems, University Registries
- Combination of **automated deduplication and humans in the loop**
  - More than 100 data curators in Europe
- **Creation of sub-orgs** (where local registries do not allow)
- **Data is CC-0** and published in Zenodo.org and via OpenAIRE Graph APIs

OpenAIRE PROVIDE Broker

- **Identifies metadata enrichment “topics”** for records in repositories
  - “URLs to open access versions”
  - “links to projects”
  - “links to datasets”
  - “ORCID IDs”
- Repository managers can **subscribe to topics and be notified** of record-level enrichments
- Accessible via **OpenAIRE PROVIDE** for a window of six-months
- **DSpace** platform offers UIs to validate notifications from OpenAIRE Broker

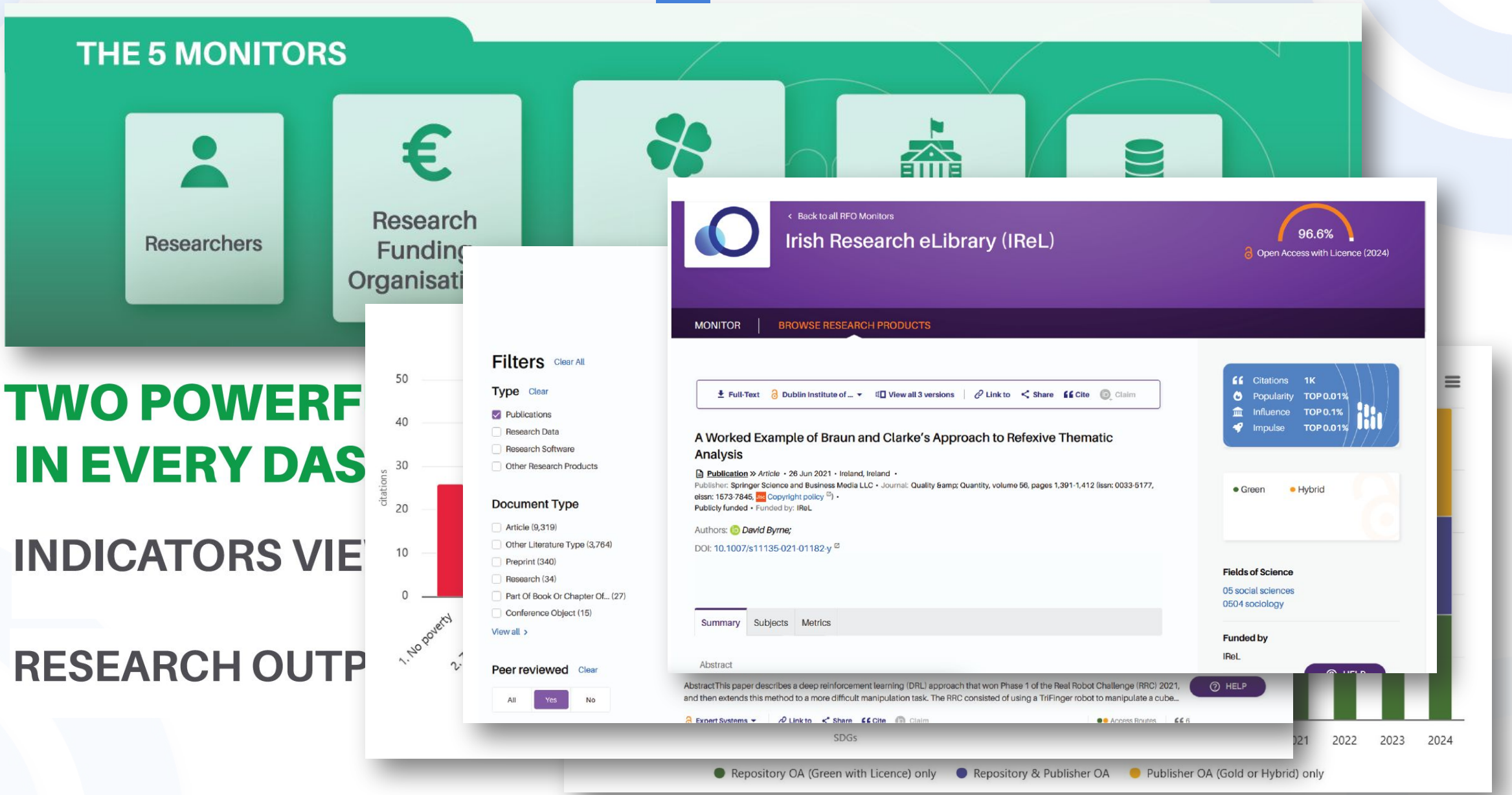
Publishing practices: anomaly detection

- Identifying **anomalies** in the metadata (e.g., inconsistencies) and poor publishing best-practices (e.g., poor metadata)
  - Publication associated to a project that started after the publishing date
  - Dataset abstract is a cut and paste of the publication abstract
  - Software does not come with relationships to publications or links to a software repository
  - Is this a dataset or supplementary material?
- Notifying **anomalies** to publishing venues and researchers
- Anomalies can be used to
  - Rule out unsatisfactory research products
  - Identify violation of mandates
  - Inspire refinements to Open Science mandates
  - Inspire ICT improvements in publishing venues or platforms
  - Boost interest in researchers at improving their metadata records
  - …

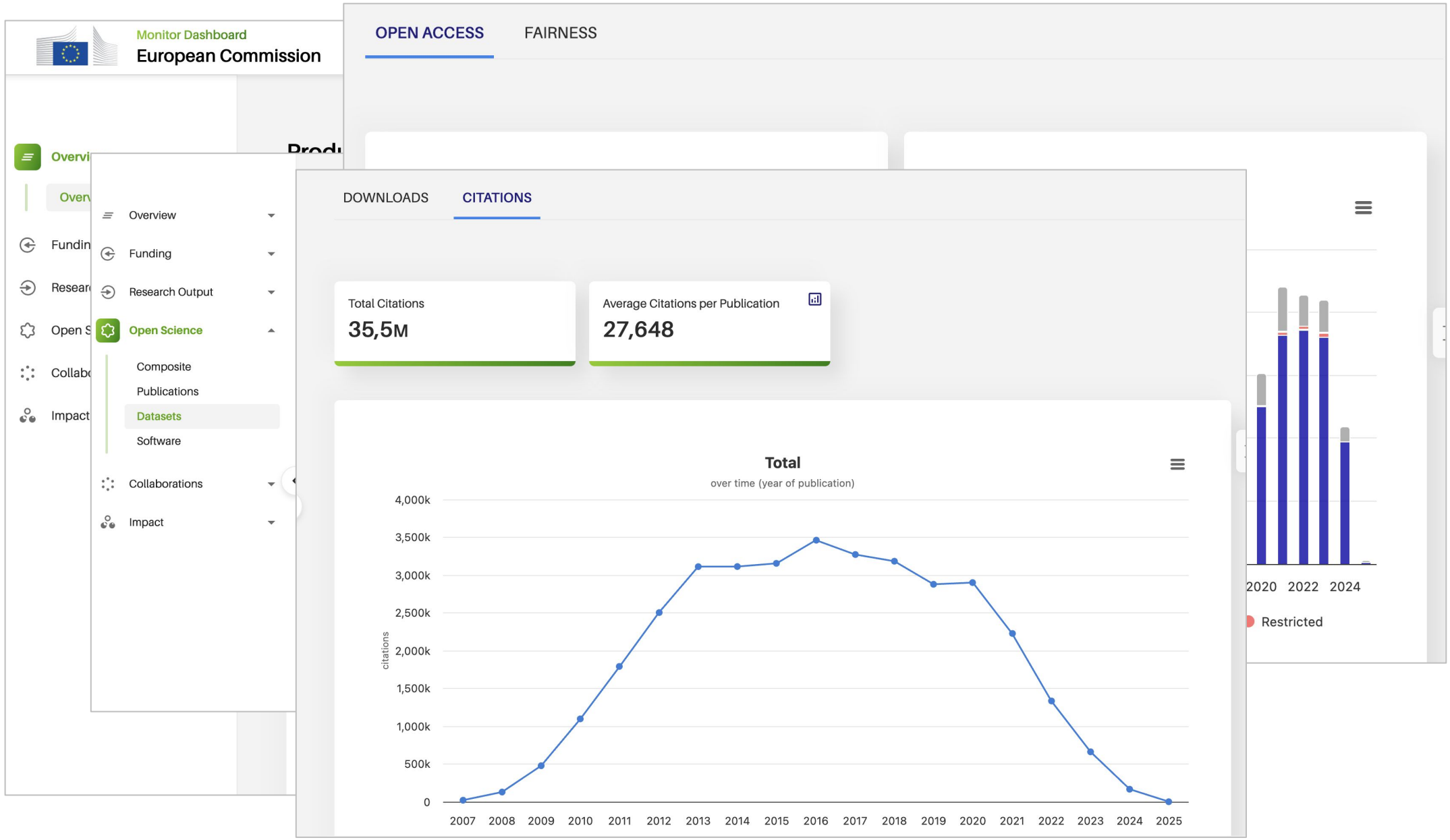
Real applications powered by the OpenAIRE Graph: Monitoring dashboards

The Irish OA MONITOR

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



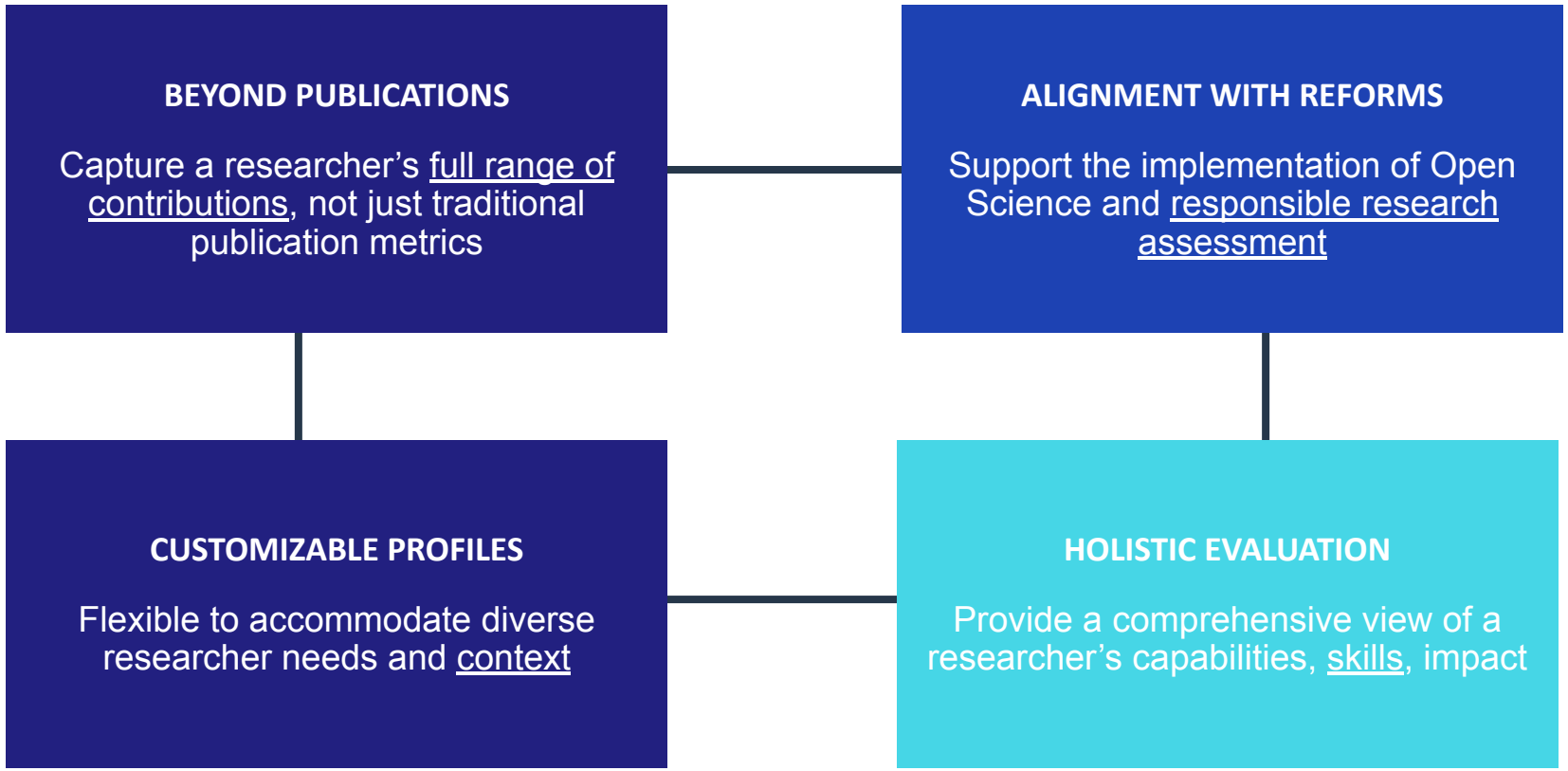
The European Commission Open Science MONITOR


Monitor Dashboard
European Commission
OPEN ACCESS
FAIRNESS
Overview
Funding
Research Output
Open Science
Composite
Publications
Datasets
Software
Collaborations
Impact
DOWNLOADS
CITATIONS
Total Citations
35,5M
Average Citations per Publication
27,648
Total
over time (year of publication)
citations
4,000k
3,500k
3,000k
2,500k
2,000k
1,500k
1,000k
500k
0
2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025
2020 2022 2024
Restricted

Real applications powered by the OpenAIRE Graph: MyResearchFolio *(coming soon)*

A Comprehensive Framework for Researcher Profile

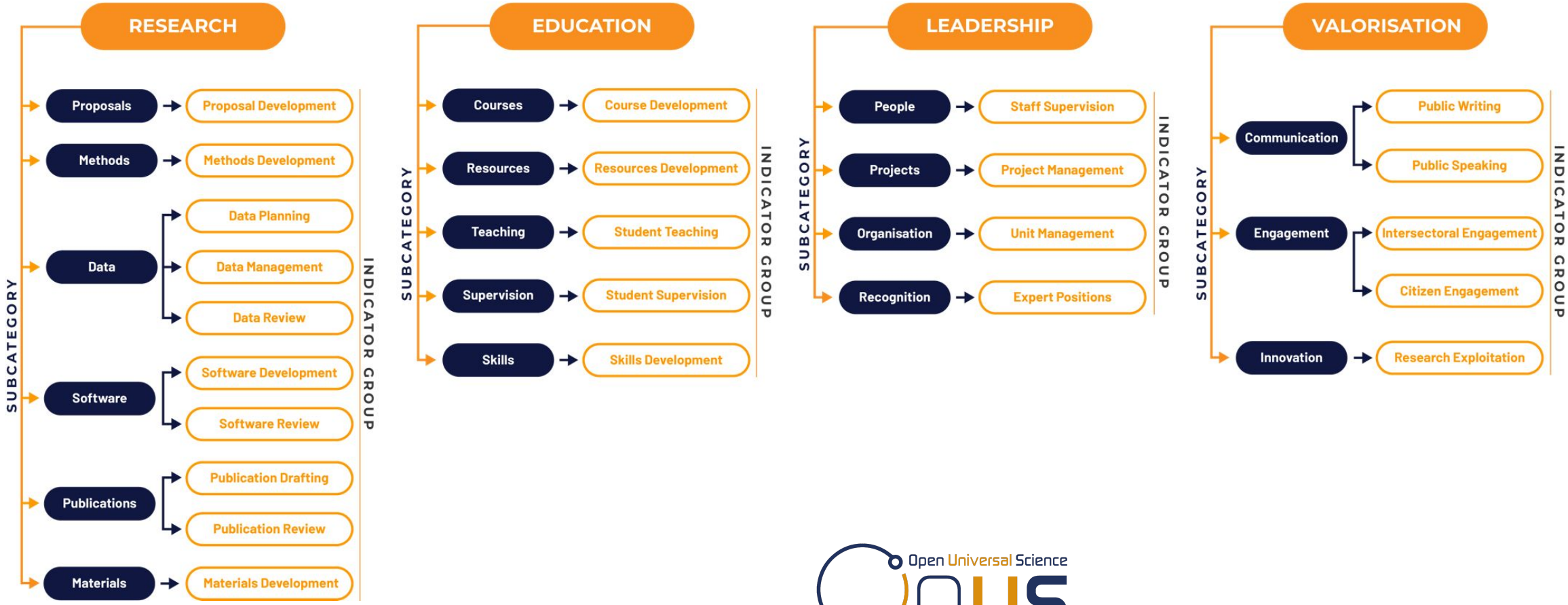
Highlight and promote researchers' achievements throughout their careers, evaluating both **quantitatively and qualitatively** their contributions to science



OpenAIRE

The foundations (1/3)

Policy Blueprints: A framework for assessment



Open Universal Science OpUS

OpenAIRE

The foundations (2/3)

Narrative CVs: A Tool for Change

Describe *contributions*, not just *counts*
Captures leadership, engagement, collaboration, and openness

THE ROYAL SOCIETY

Personal Details

The four-module narrative section

Personal Statement

Additions

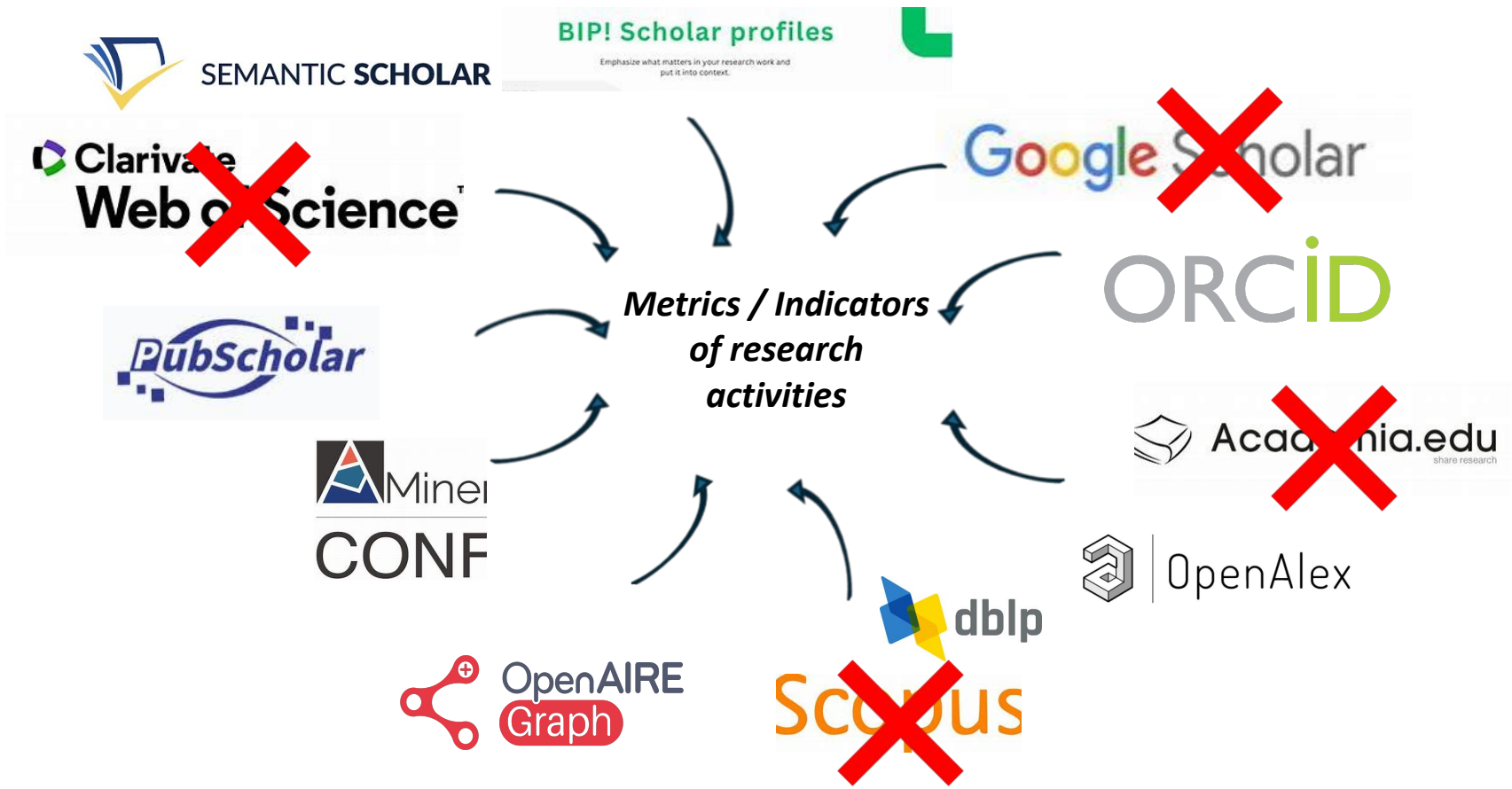
Contribution to

- **Generation of knowledge**
- **Development of individuals**
- **Wider research community**
- **Broader society**

OpenAIRE

The foundations (3/3)

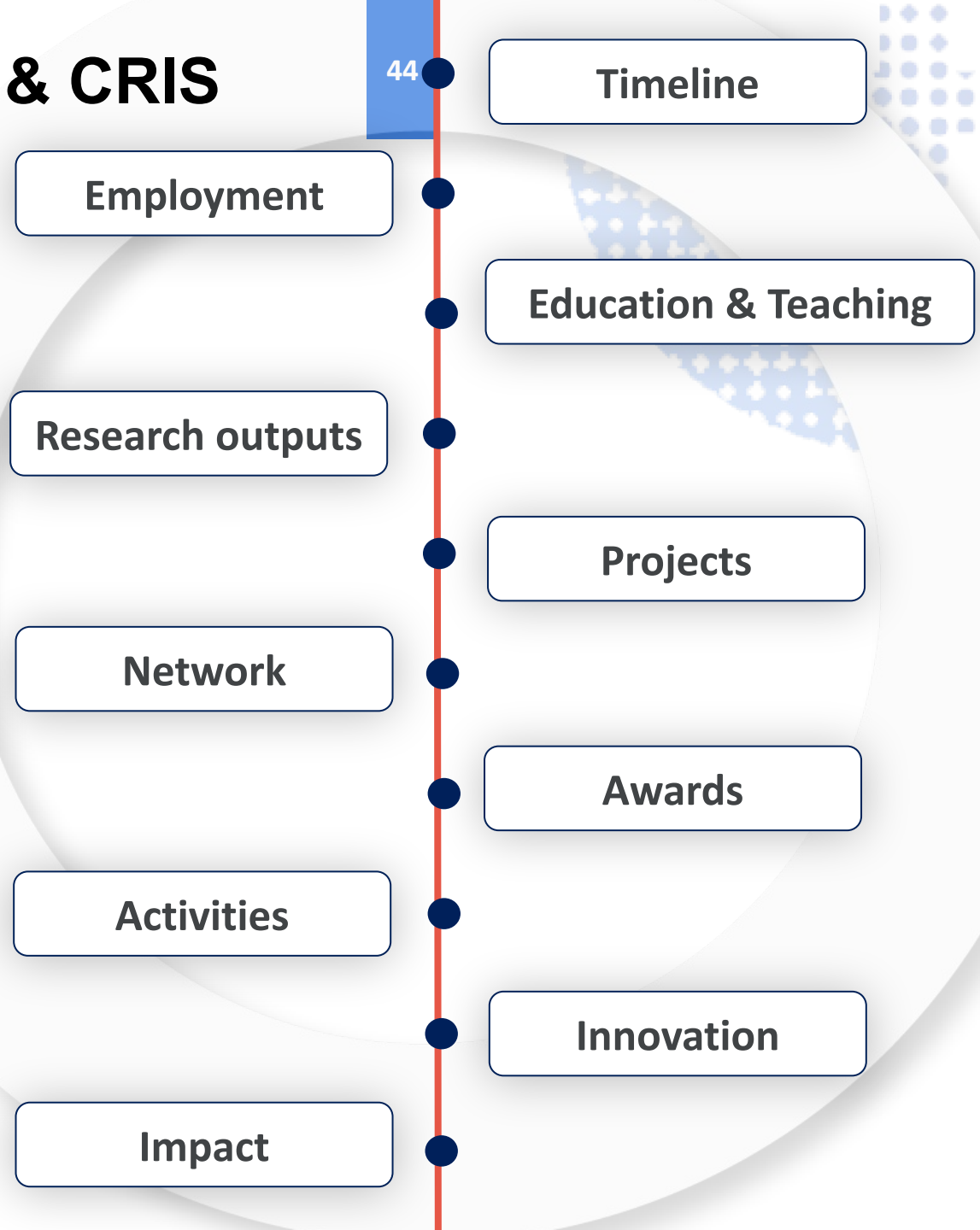
Enabling Change: Open Infrastructures in Action



Overlay on ORCID & CRIS

Blended views

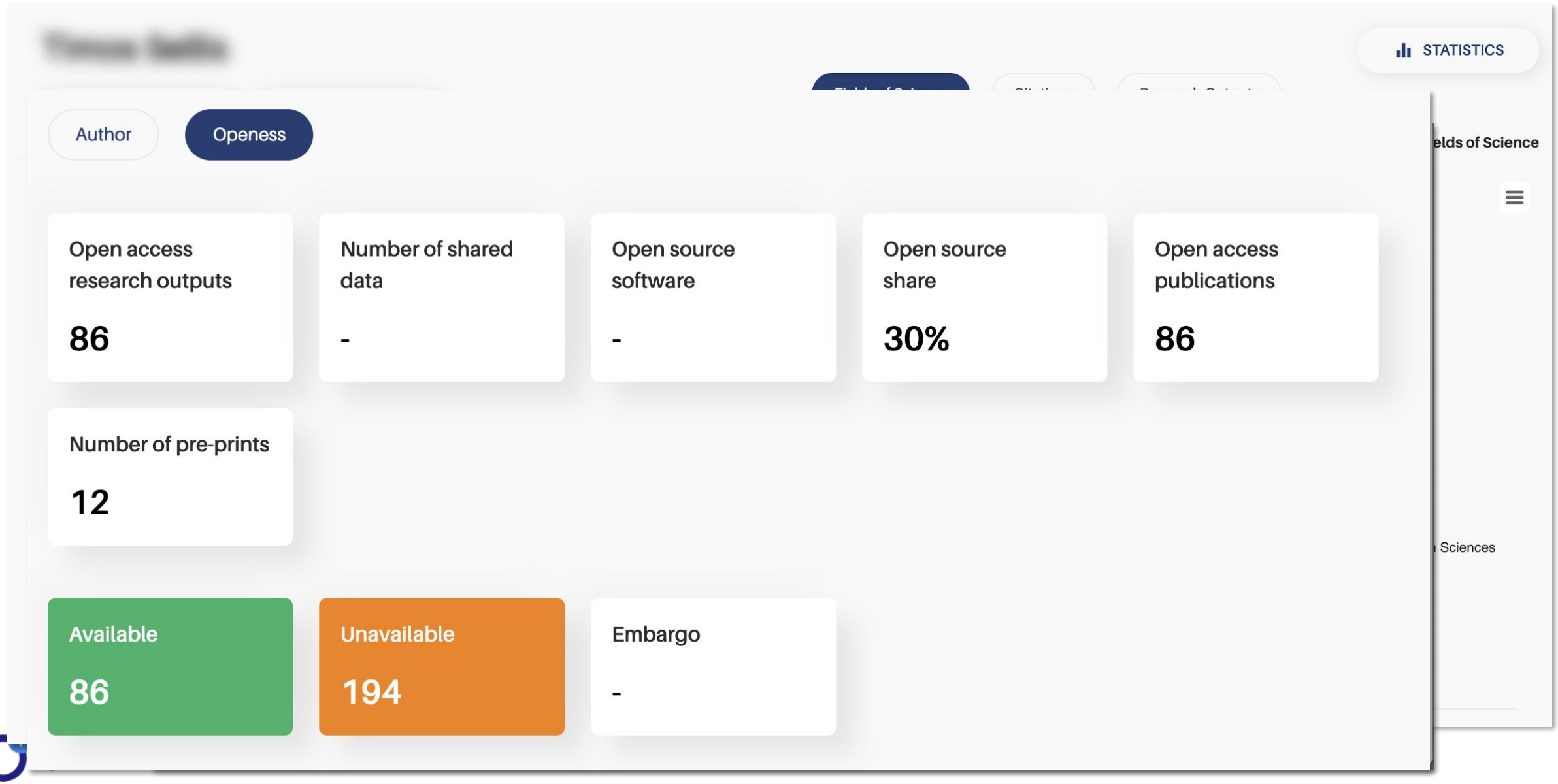
- Key Information
- Narrative CV
- Open Science



graspos
open research assessment dataspace

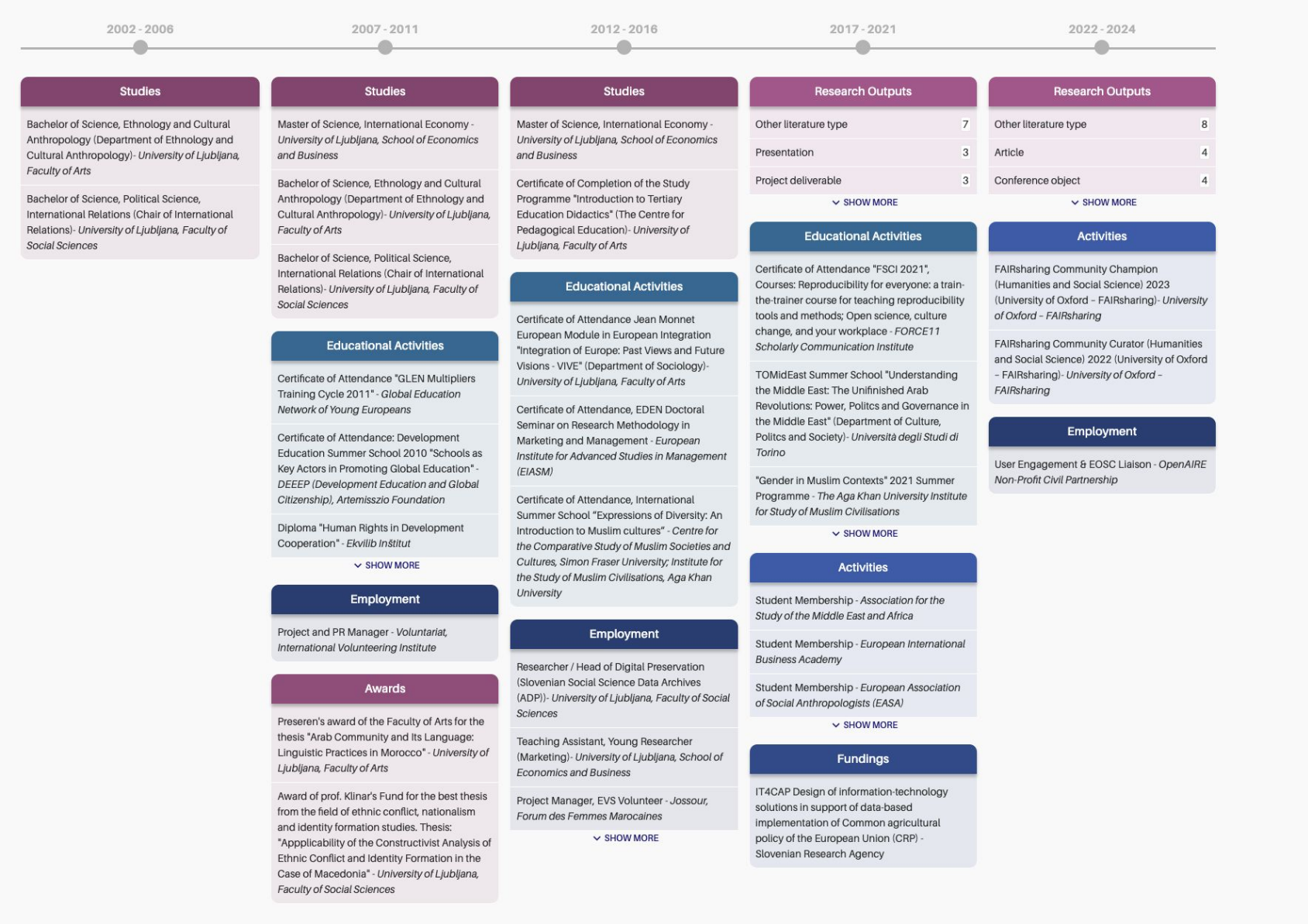
OpenAIRE

OpenAIRE MyResearchFolio



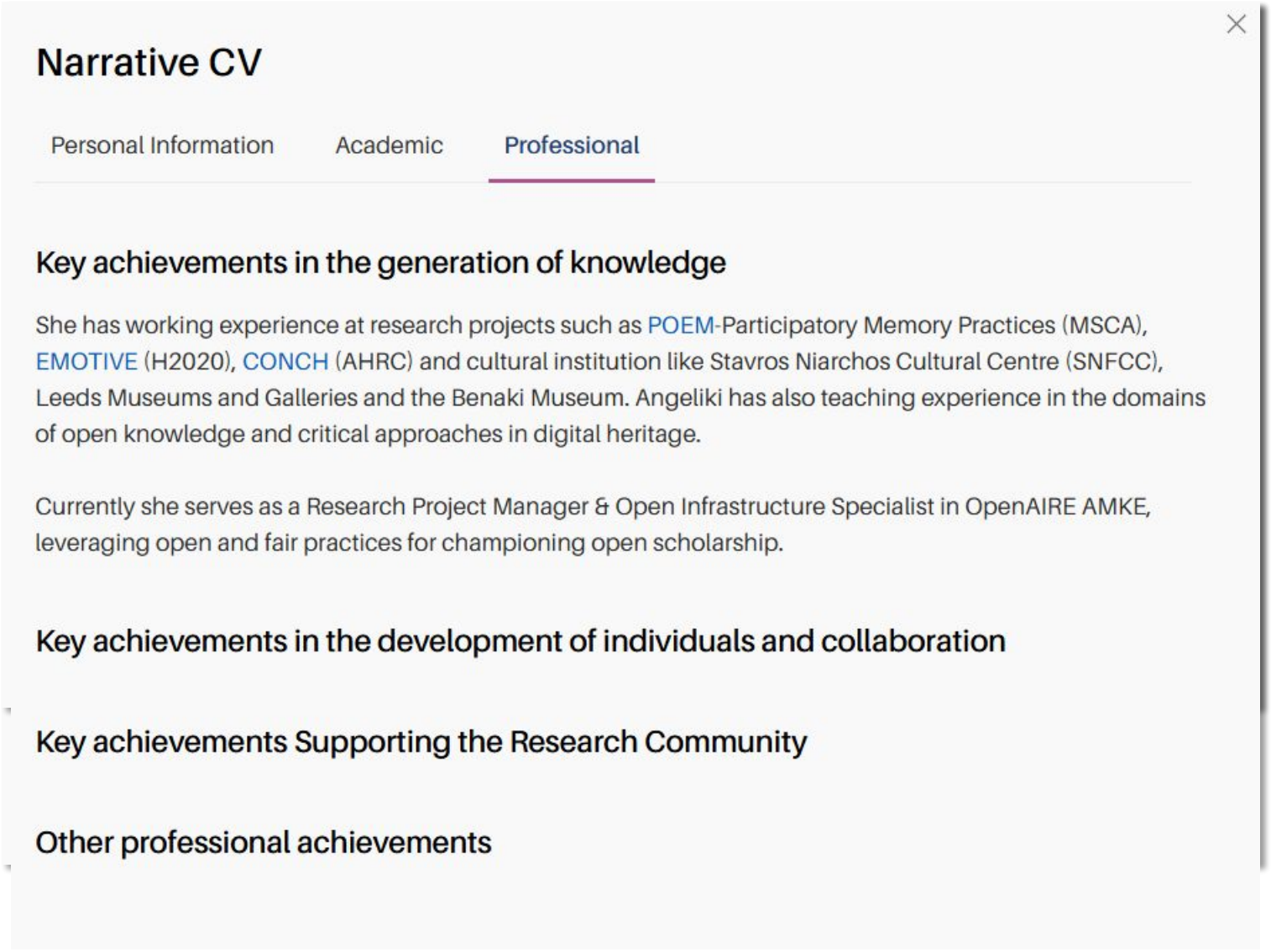
PRESENTATION

Experimenting with different modes



ASSESSMENT

Experimenting with different types of Narrative CVs

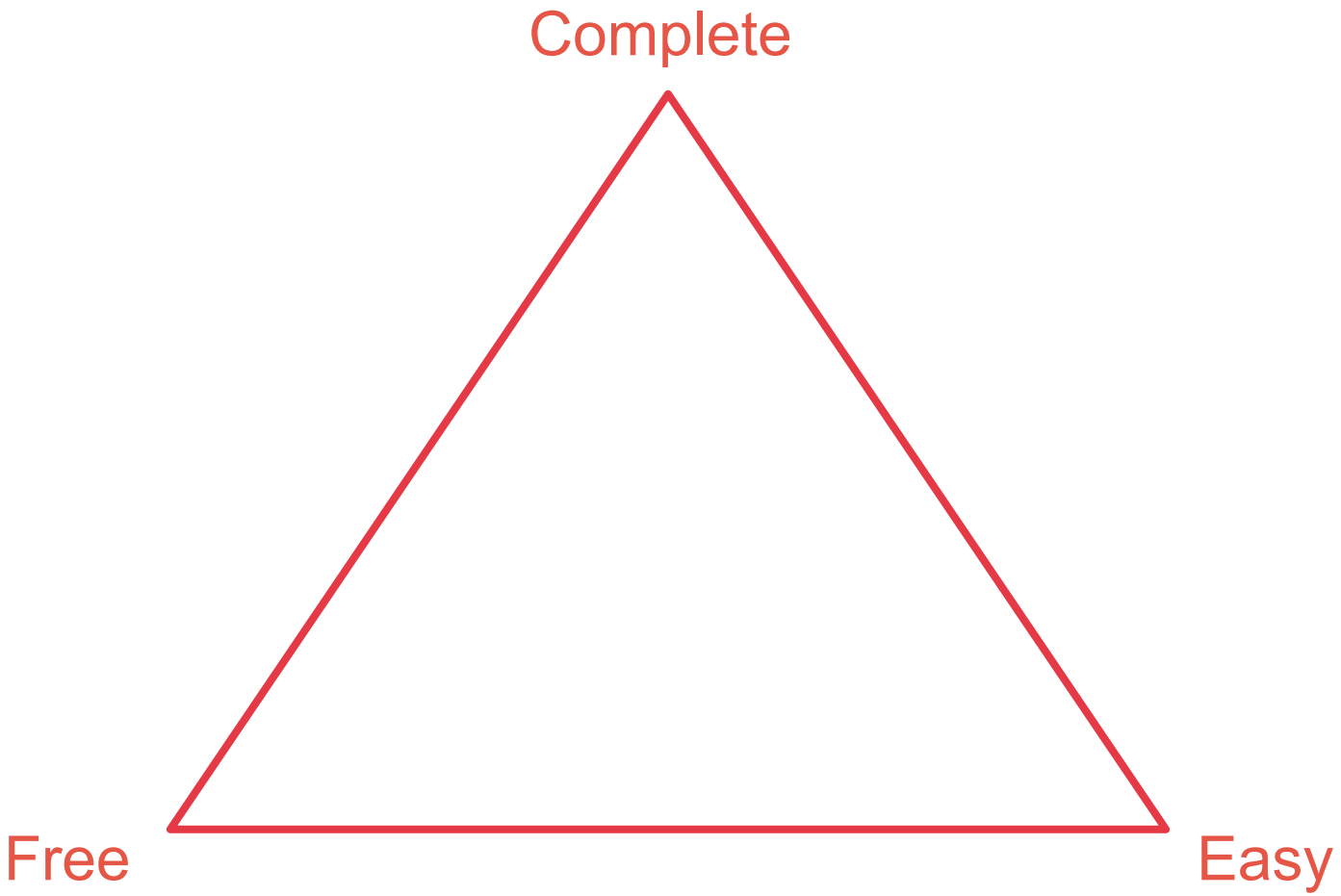


OpenAIRE Graph: Data Access Portfolio

Slides from
Mannocci A., Fernandes Mazoni A. “Accessing the OpenAIRE Graph via Google BigQuery“
Tutorial presented at ISSI 2025, Yerevan, Armenia

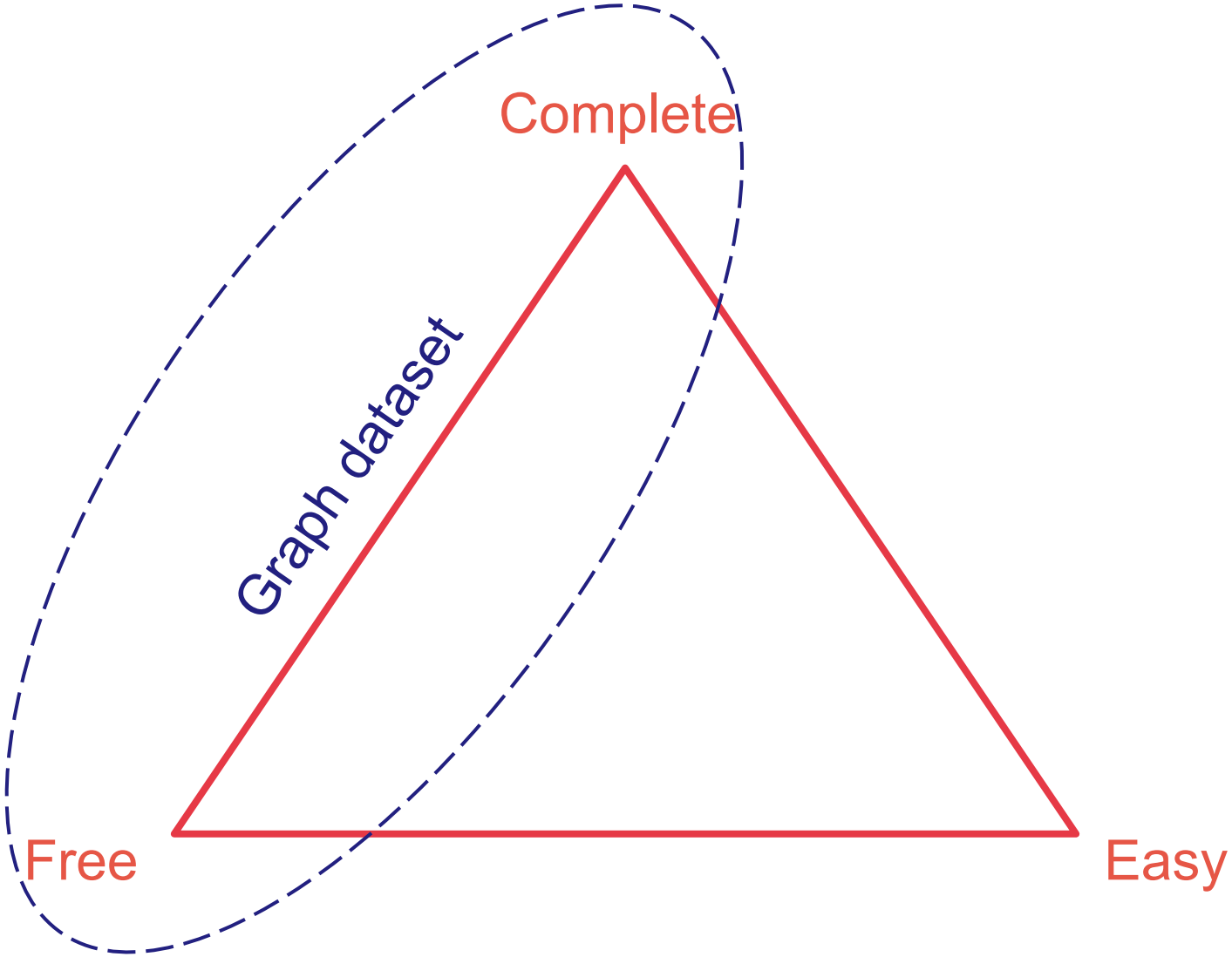
OpenAIRE

Data access trilemma (pick two)

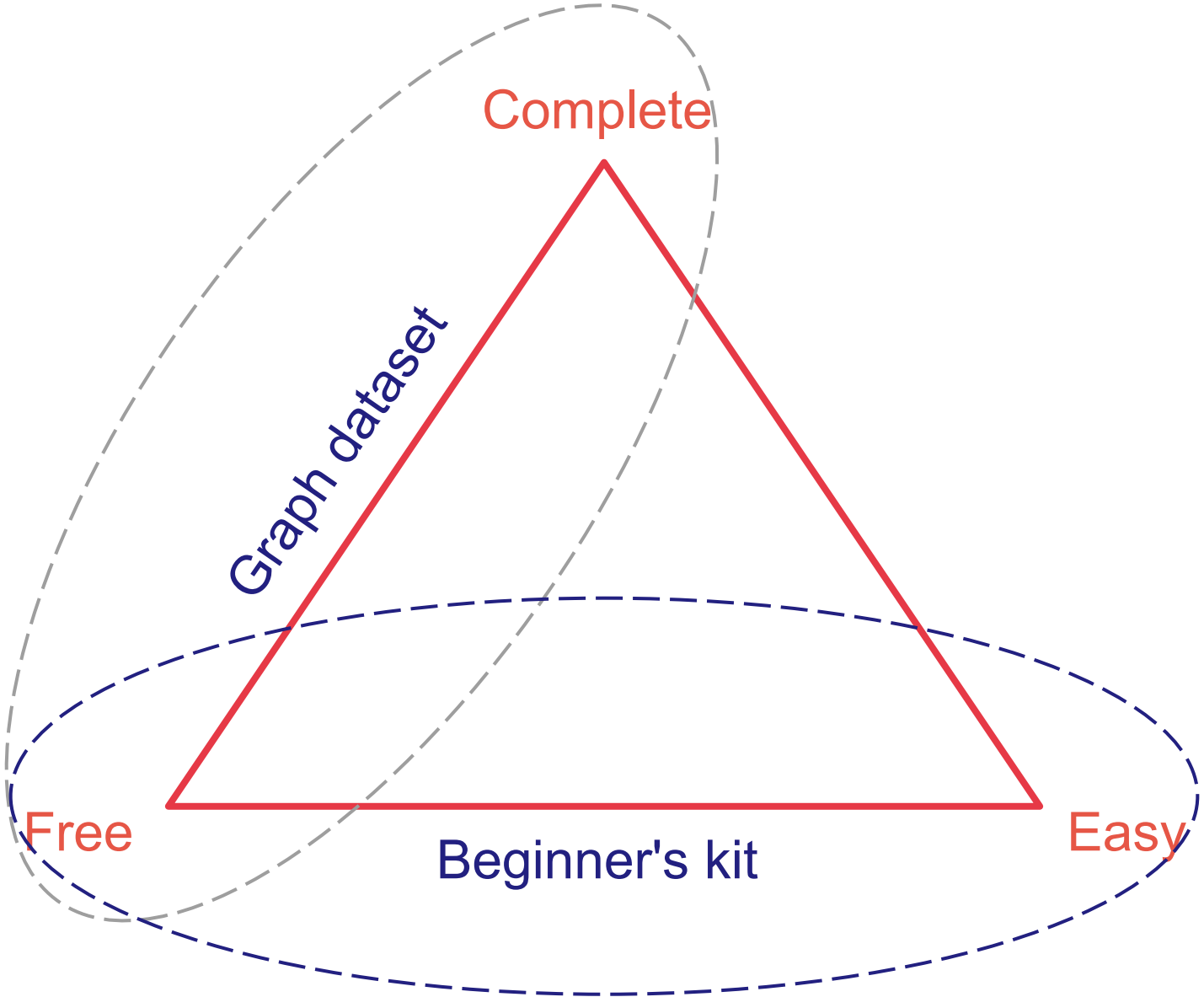


OpenAIRE

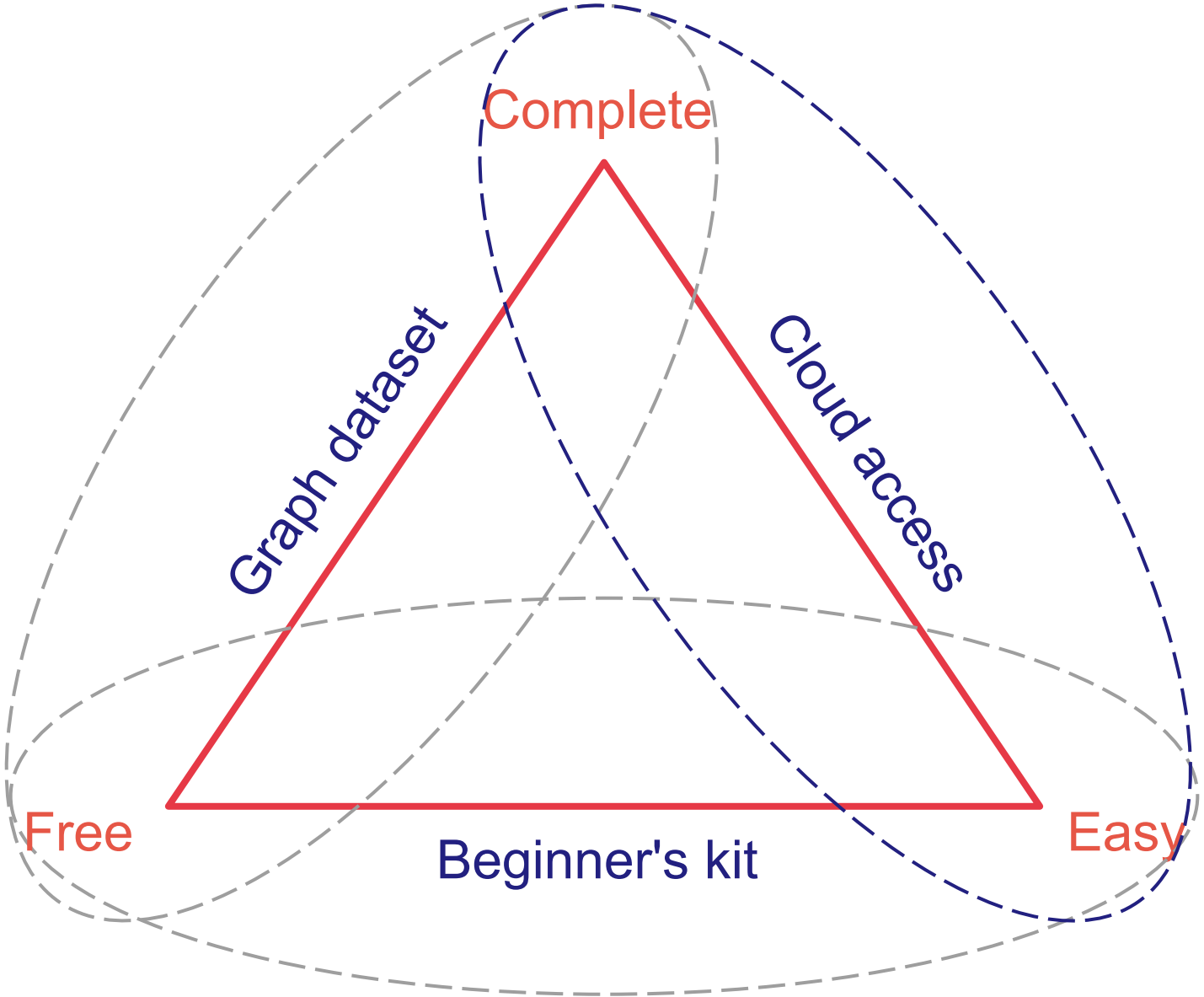
Data access trilemma (pick two)



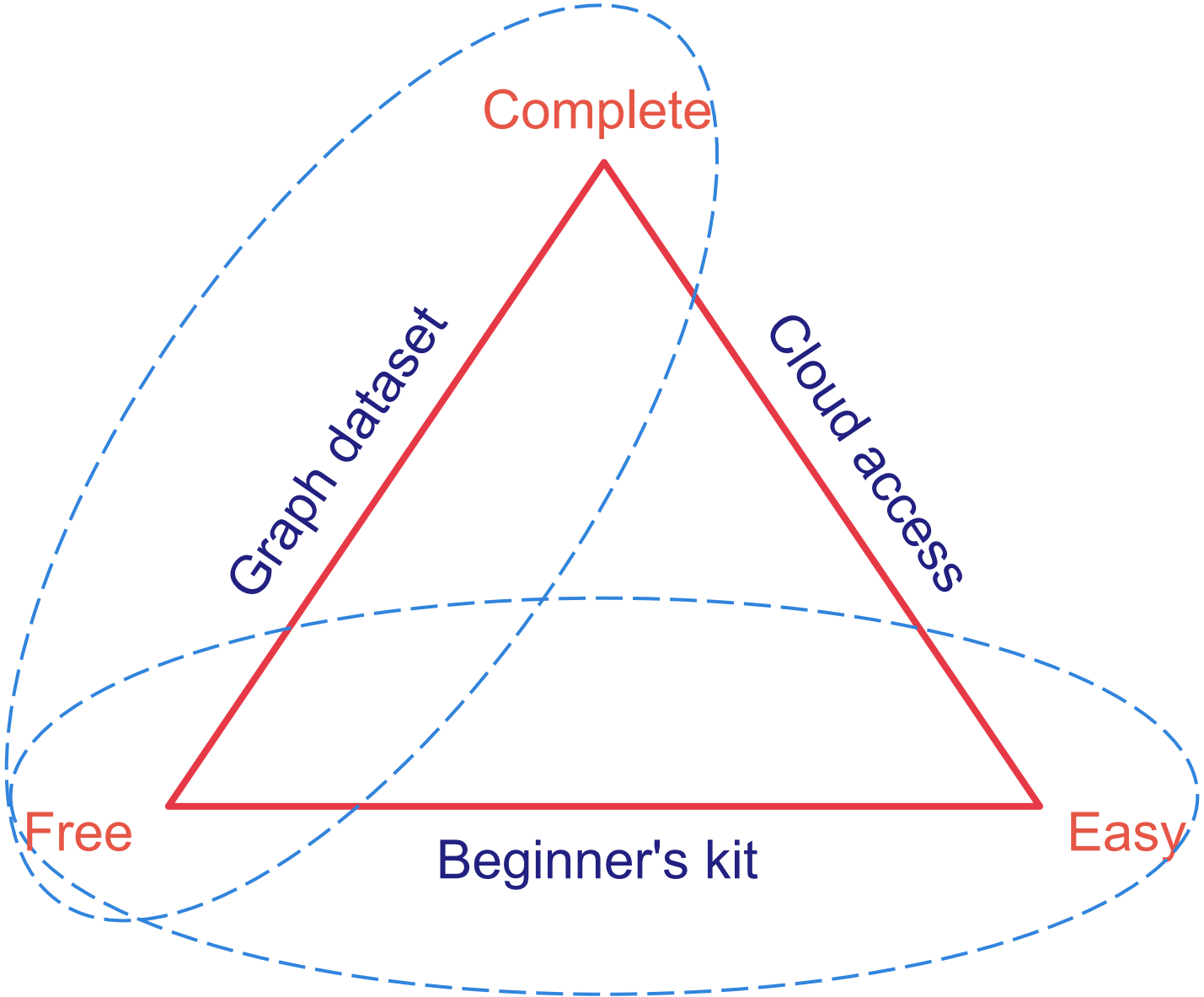
Data access trilemma (pick two)



Data access trilemma (pick two)



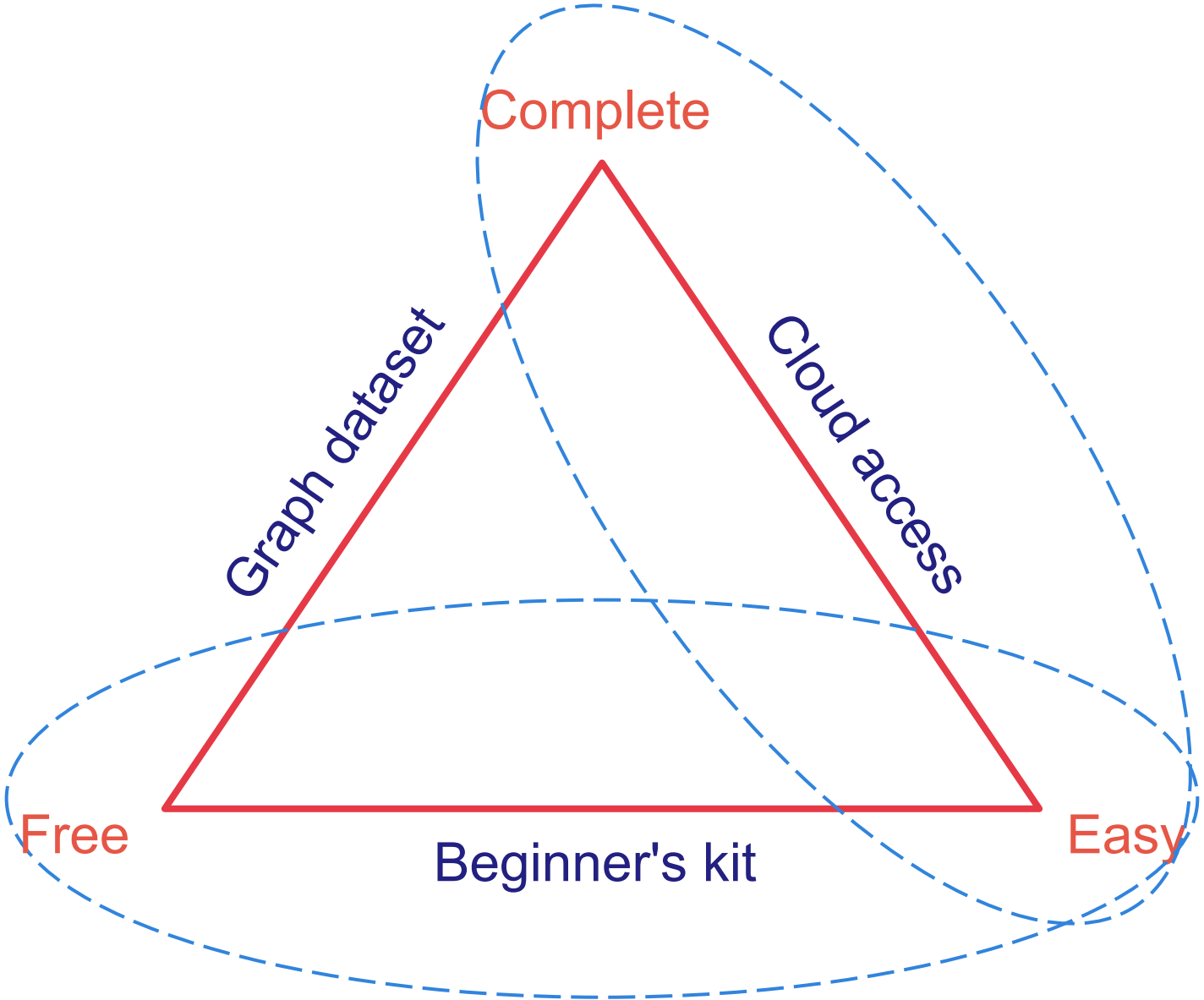
Data access trilemma (pick two)



Run on local resources

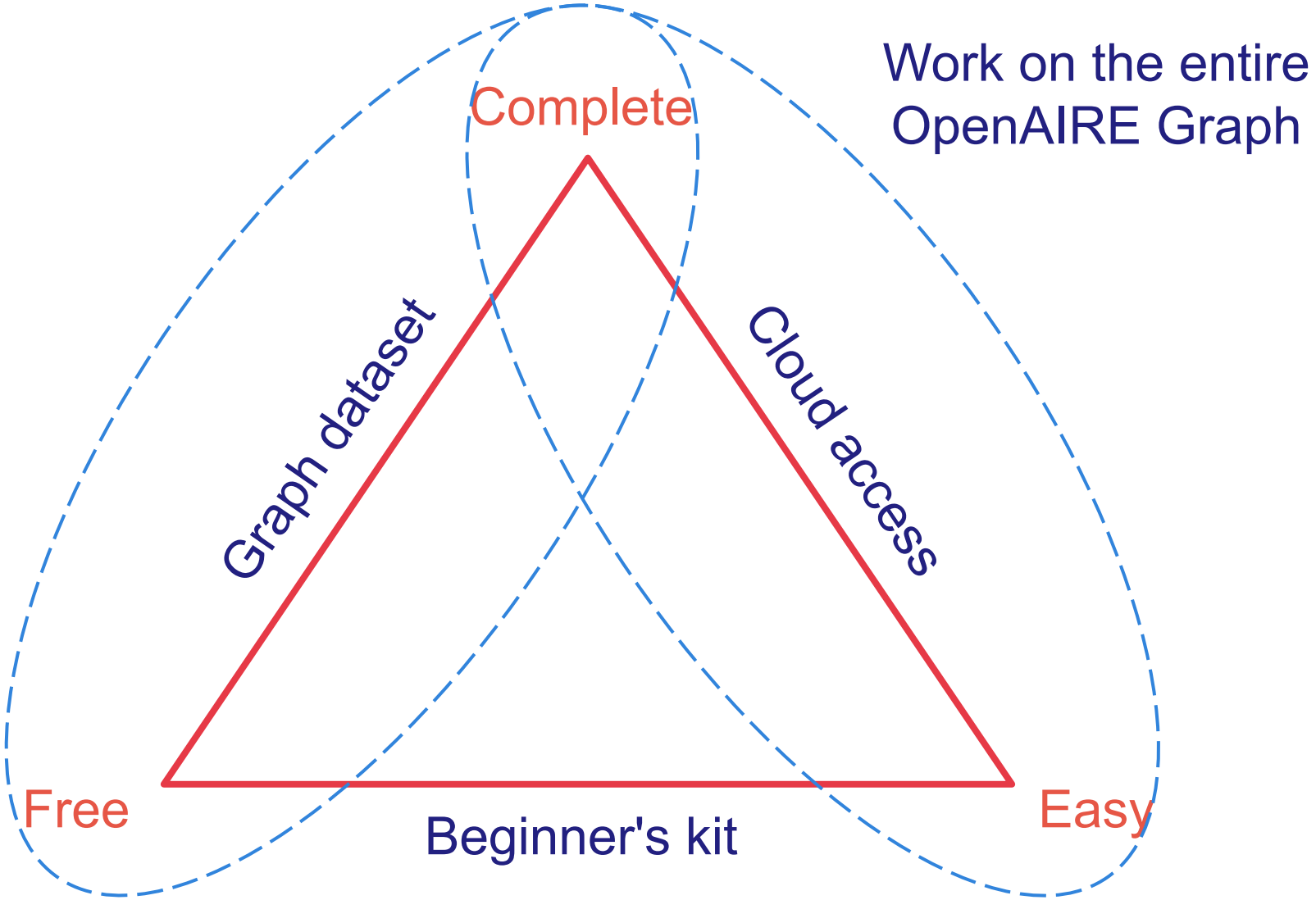
OpenAIRE

Data access trilemma (pick two)



Rely on "SQL"
to access data

Data access trilemma (pick two)



OpenAIRE

Differentiated provisioning cadence

Monthly

- OpenAIRE Explore
- Stats and dashboards
- Public APIs

Semestral* (in sync)

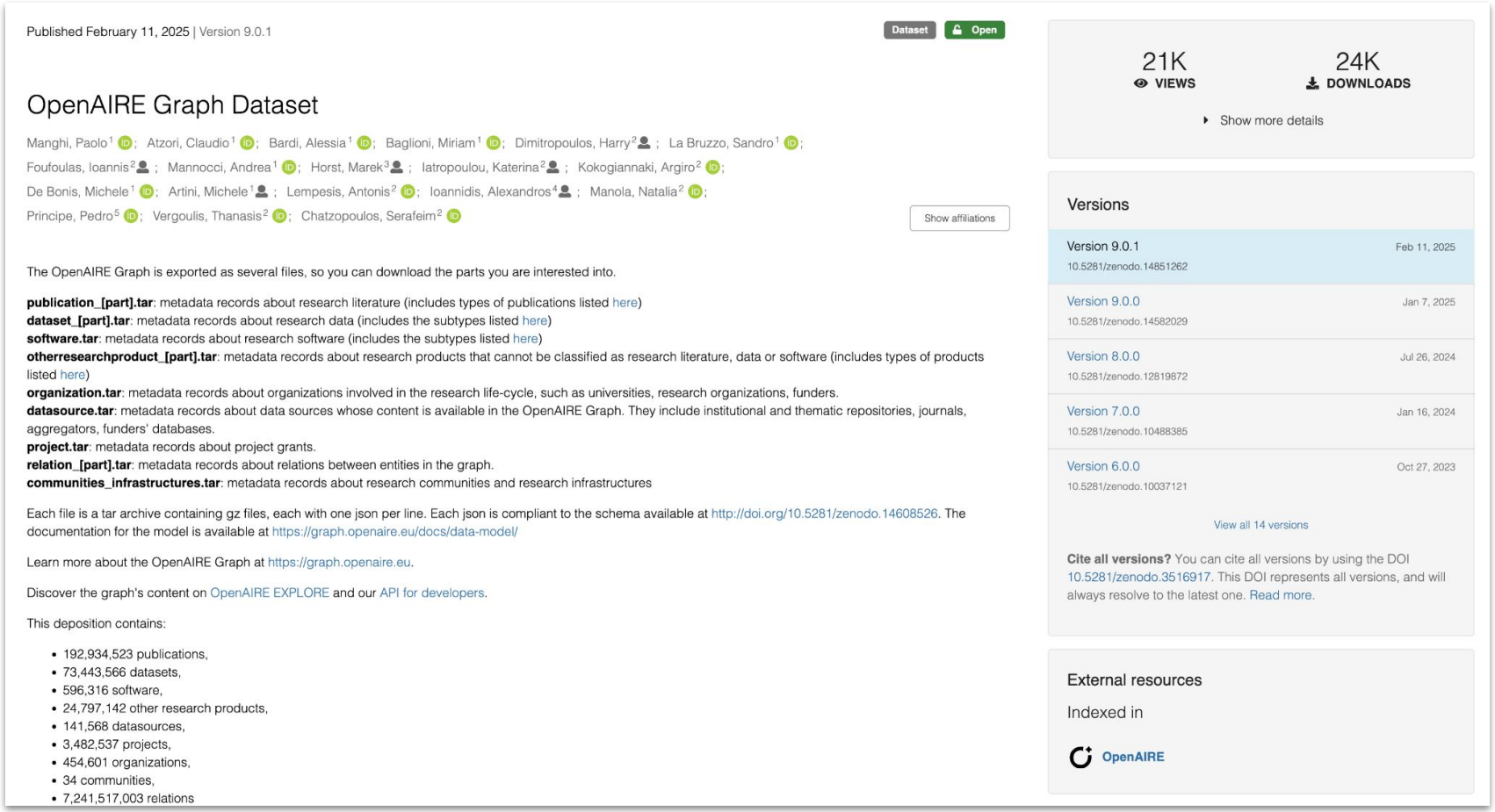
- OpenAIRE Graph dataset
- Beginner's kit
- Cloud access

*unless something urgent has to be released sooner than that

OpenAIRE

OpenAIRE Graph dataset

OpenAIRE Graph dataset



Published February 11, 2025 | Version 9.0.1

Dataset Open

OpenAIRE Graph Dataset

Manghi, Paolo[1]; Atzori, Claudio[1]; Bardi, Alessia[1]; Baglioni, Miriam[1]; Dimitropoulos, Harry[2]; La Bruzzo, Sandro[1]; Foufoulas, Ioannis[2]; Mannocci, Andrea[1]; Horst, Marek[3]; Iatropoulou, Katerina[2]; Kokogiannaki, Argiro[2]; De Bonis, Michele[1]; Artini, Michele[1]; Lempesis, Antonis[2]; Ioannidis, Alexandros[4]; Manola, Natalia[2]; Principe, Pedro[5]; Vergoulis, Thanasis[2]; Chatzopoulos, Serafeim[2]

Show affiliations

The OpenAIRE Graph is exported as several files, 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

Each file is a tar archive containing gz files, each with one json per line. Each json is compliant to the schema available at http://doi.org/10.5281/zenodo.14608526. The documentation for the model is available at https://graph.openaire.eu/docs/data-model/

Learn more about the OpenAIRE Graph at https://graph.openaire.eu.

Discover the graph's content on OpenAIRE EXPLORE and our API for developers.

This deposition contains:

- 192,934,523 publications,
- 73,443,566 datasets,
- 596,316 software,
- 24,797,142 other research products,
- 141,568 datasources,
- 3,482,537 projects,
- 454,601 organizations,
- 34 communities,
- 7,241,517,003 relations

21K VIEWS 24K DOWNLOADS

Show more details

Versions

Version 9.0.1 Feb 11, 2025
10.5281/zenodo.14851262

Version 9.0.0 Jan 7, 2025
10.5281/zenodo.14582029

Version 8.0.0 Jul 26, 2024
10.5281/zenodo.12819872

Version 7.0.0 Jan 16, 2024
10.5281/zenodo.10488385

Version 6.0.0 Oct 27, 2023
10.5281/zenodo.10037121

View all 14 versions

Cite all versions? You can cite all versions by using the DOI 10.5281/zenodo.3516917. This DOI represents all versions, and will always resolve to the latest one. Read more.

External resources

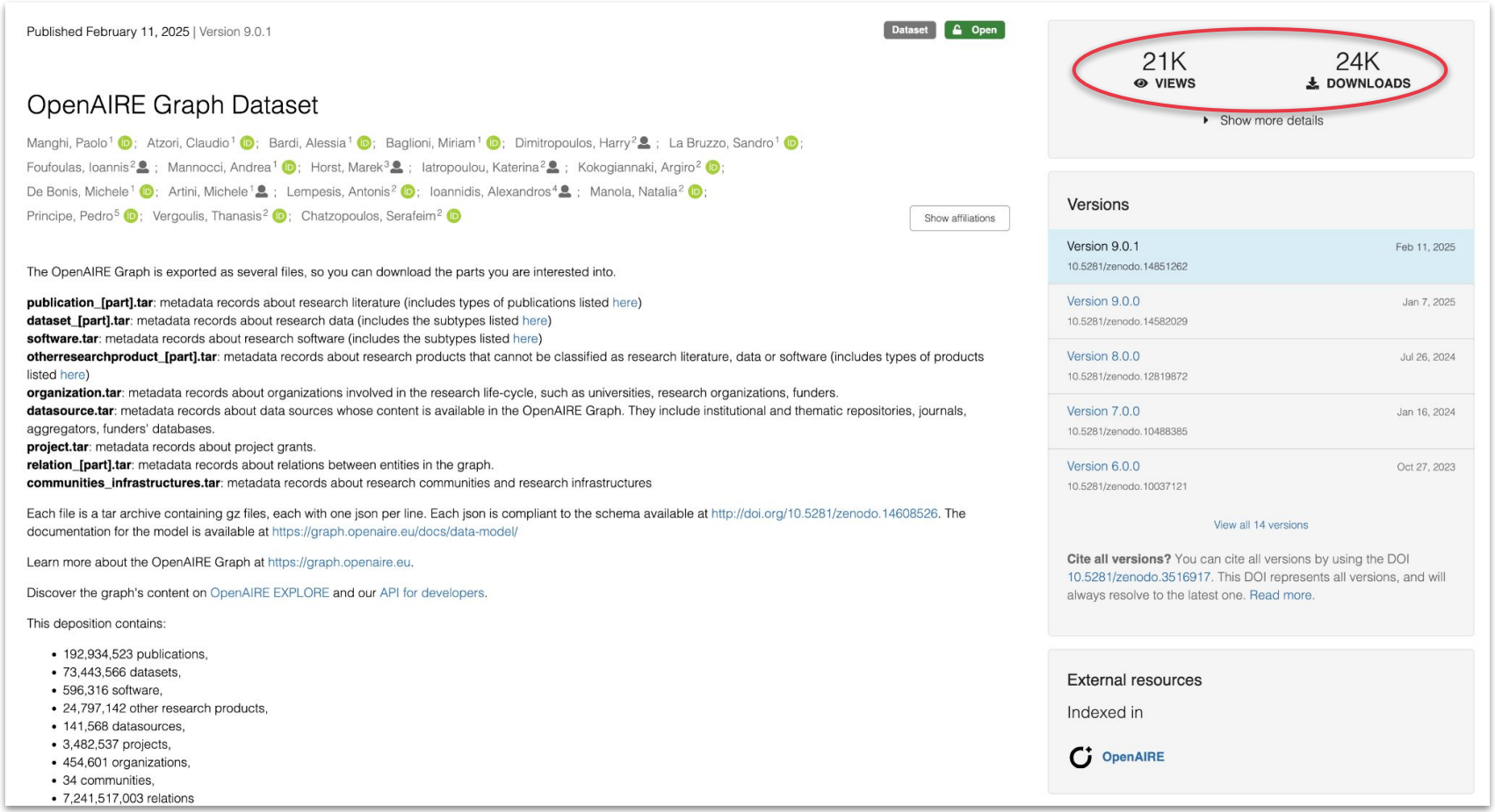
Indexed in

OpenAIRE

zenodo

https://zenodo.org/records/14851262

OpenAIRE Graph dataset



Published February 11, 2025 | Version 9.0.1

Dataset Open

OpenAIRE Graph Dataset

Manghi, Paolo[1]; Atzori, Claudio[1]; Bardi, Alessia[1]; Baglioni, Miriam[1]; Dimitropoulos, Harry[2]; La Bruzzo, Sandro[1]; Foufoulas, Ioannis[2]; Mannocci, Andrea[1]; Horst, Marek[3]; Iatropoulou, Katerina[2]; Kokogiannaki, Argiro[2]; De Bonis, Michele[1]; Artini, Michele[1]; Lempesis, Antonis[2]; Ioannidis, Alexandros[4]; Manola, Natalia[2]; Principe, Pedro[5]; Vergoulis, Thanasis[2]; Chatzopoulos, Serafeim[2]

Show affiliations

The OpenAIRE Graph is exported as several files, 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

Each file is a tar archive containing gz files, each with one json per line. Each json is compliant to the schema available at http://doi.org/10.5281/zenodo.14608526. The documentation for the model is available at https://graph.openaire.eu/docs/data-model/

Learn more about the OpenAIRE Graph at https://graph.openaire.eu.

Discover the graph's content on OpenAIRE EXPLORE and our API for developers.

This deposition contains:

- 192,934,523 publications,
- 73,443,566 datasets,
- 596,316 software,
- 24,797,142 other research products,
- 141,568 datasources,
- 3,482,537 projects,
- 454,601 organizations,
- 34 communities,
- 7,241,517,003 relations

21K VIEWS 24K DOWNLOADS

Show more details

Versions

Version 9.0.1 10.5281/zenodo.14851262	Feb 11, 2025
Version 9.0.0 10.5281/zenodo.14582029	Jan 7, 2025
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

View all 14 versions

Cite all versions? You can cite all versions by using the DOI 10.5281/zenodo.3516917. This DOI represents all versions, and will always resolve to the latest one. Read more.

External resources

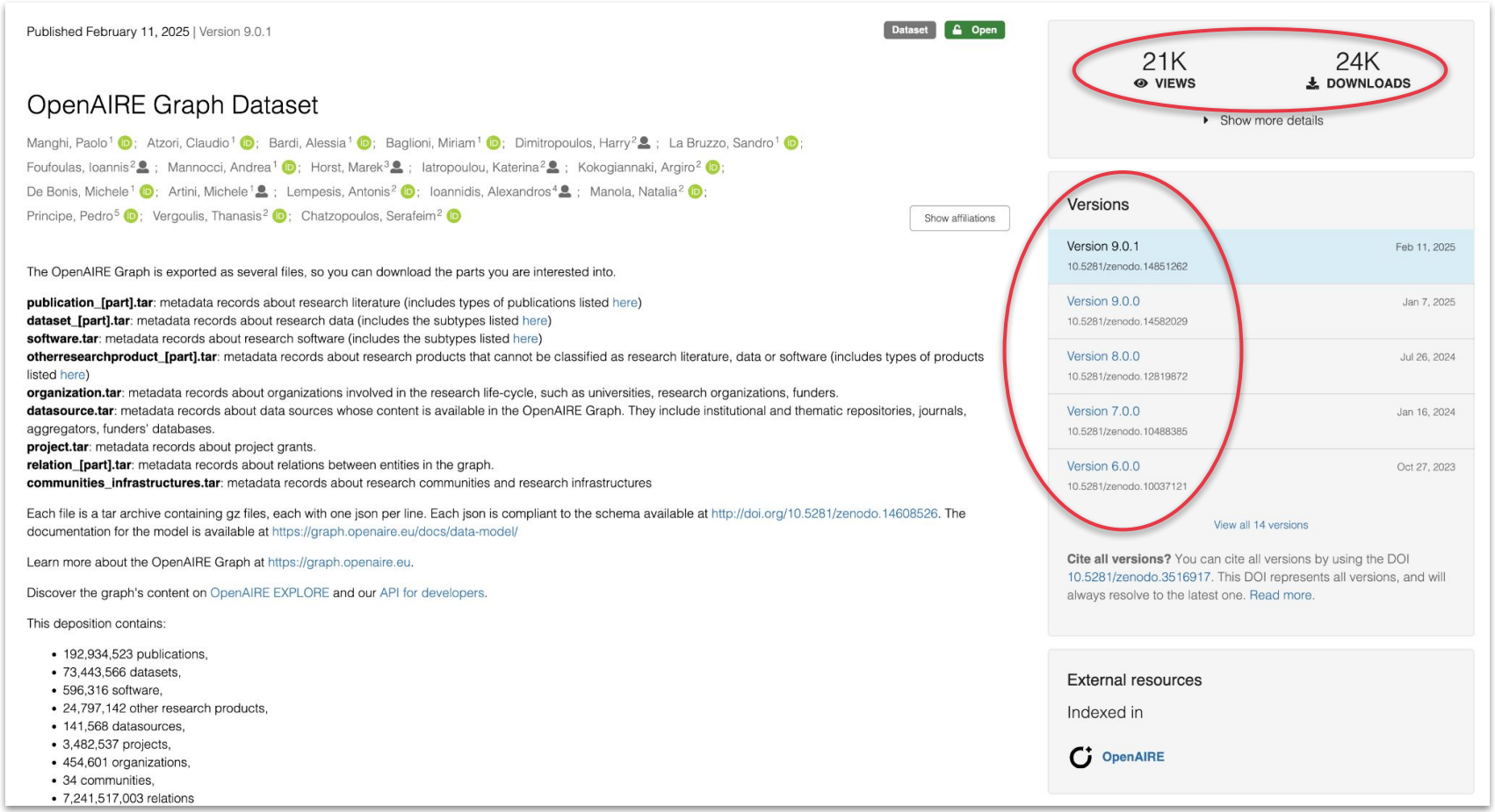
Indexed in

OpenAIRE

zenodo

https://zenodo.org/records/14851262

OpenAIRE Graph dataset



Published February 11, 2025 | Version 9.0.1

Dataset Open

OpenAIRE Graph Dataset

Manghi, Paolo[1]; Atzori, Claudio[1]; Bardi, Alessia[1]; Baglioni, Miriam[1]; Dimitropoulos, Harry[2]; La Bruzzo, Sandro[1]; Foufoulas, Ioannis[2]; Mannocci, Andrea[1]; Horst, Marek[3]; Iatropoulou, Katerina[2]; Kokogiannaki, Argiro[2]; De Bonis, Michele[1]; Artini, Michele[1]; Lempesis, Antonis[2]; Ioannidis, Alexandros[4]; Manola, Natalia[2]; Principe, Pedro[5]; Vergoulis, Thanasis[2]; Chatzopoulos, Serafeim[2]

Show affiliations

The OpenAIRE Graph is exported as several files, 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

Each file is a tar archive containing gz files, each with one json per line. Each json is compliant to the schema available at http://doi.org/10.5281/zenodo.14608526. The documentation for the model is available at https://graph.openaire.eu/docs/data-model/

Learn more about the OpenAIRE Graph at https://graph.openaire.eu.

Discover the graph's content on OpenAIRE EXPLORE and our API for developers.

This deposition contains:

- 192,934,523 publications,
- 73,443,566 datasets,
- 596,316 software,
- 24,797,142 other research products,
- 141,568 datasources,
- 3,482,537 projects,
- 454,601 organizations,
- 34 communities,
- 7,241,517,003 relations

21K VIEWS 24K DOWNLOADS

Show more details

Versions

Version 9.0.1 10.5281/zenodo.14851262	Feb 11, 2025
Version 9.0.0 10.5281/zenodo.14582029	Jan 7, 2025
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

View all 14 versions

Cite all versions? You can cite all versions by using the DOI 10.5281/zenodo.3516917. This DOI represents all versions, and will always resolve to the latest one. Read more.

External resources

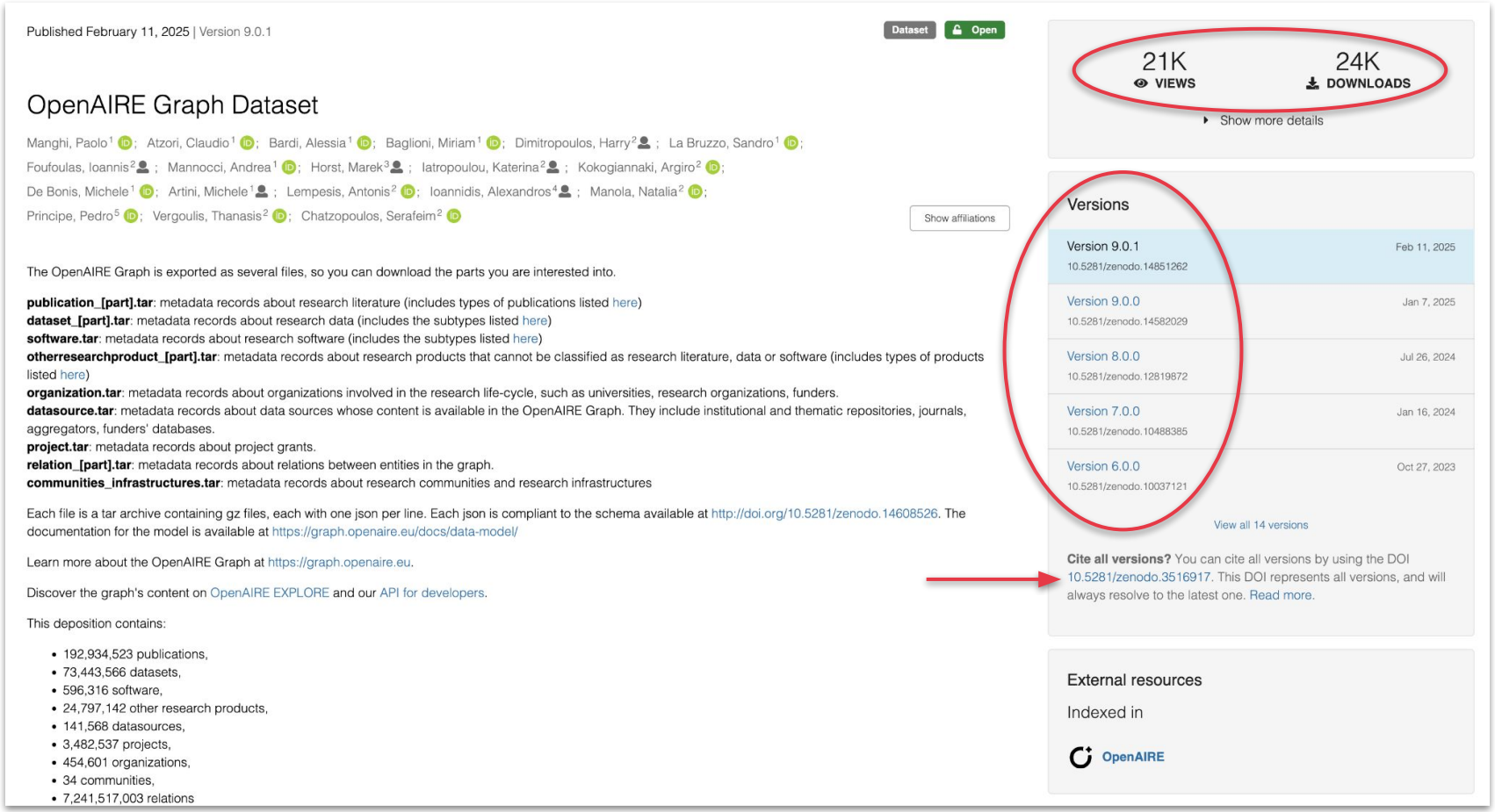
Indexed in

OpenAIRE

zenodo

https://zenodo.org/records/14851262

OpenAIRE Graph dataset



Published February 11, 2025 | Version 9.0.1

Dataset Open

OpenAIRE Graph Dataset

Manghi, Paolo[1]; Atzori, Claudio[1]; Bardi, Alessia[1]; Baglioni, Miriam[1]; Dimitropoulos, Harry[2]; La Bruzzo, Sandro[1]; Foufoulas, Ioannis[2]; Mannocci, Andrea[1]; Horst, Marek[3]; Iatropoulou, Katerina[2]; Kokogiannaki, Argiro[2]; De Bonis, Michele[1]; Artini, Michele[1]; Lempesis, Antonis[2]; Ioannidis, Alexandros[4]; Manola, Natalia[2]; Principe, Pedro[5]; Vergoulis, Thanasis[2]; Chatzopoulos, Serafeim[2]

Show affiliations

The OpenAIRE Graph is exported as several files, 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

Each file is a tar archive containing gz files, each with one json per line. Each json is compliant to the schema available at http://doi.org/10.5281/zenodo.14608526. The documentation for the model is available at https://graph.openaire.eu/docs/data-model/

Learn more about the OpenAIRE Graph at https://graph.openaire.eu.

Discover the graph's content on OpenAIRE EXPLORE and our API for developers.

This deposition contains:

- 192,934,523 publications,
- 73,443,566 datasets,
- 596,316 software,
- 24,797,142 other research products,
- 141,568 datasources,
- 3,482,537 projects,
- 454,601 organizations,
- 34 communities,
- 7,241,517,003 relations

21K VIEWS 24K DOWNLOADS

Show more details

Versions

Version 9.0.1 10.5281/zenodo.14851262	Feb 11, 2025
Version 9.0.0 10.5281/zenodo.14582029	Jan 7, 2025
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

View all 14 versions

Cite all versions? You can cite all versions by using the DOI 10.5281/zenodo.3516917. This DOI represents all versions, and will always resolve to the latest one. Read more.

External resources

Indexed in

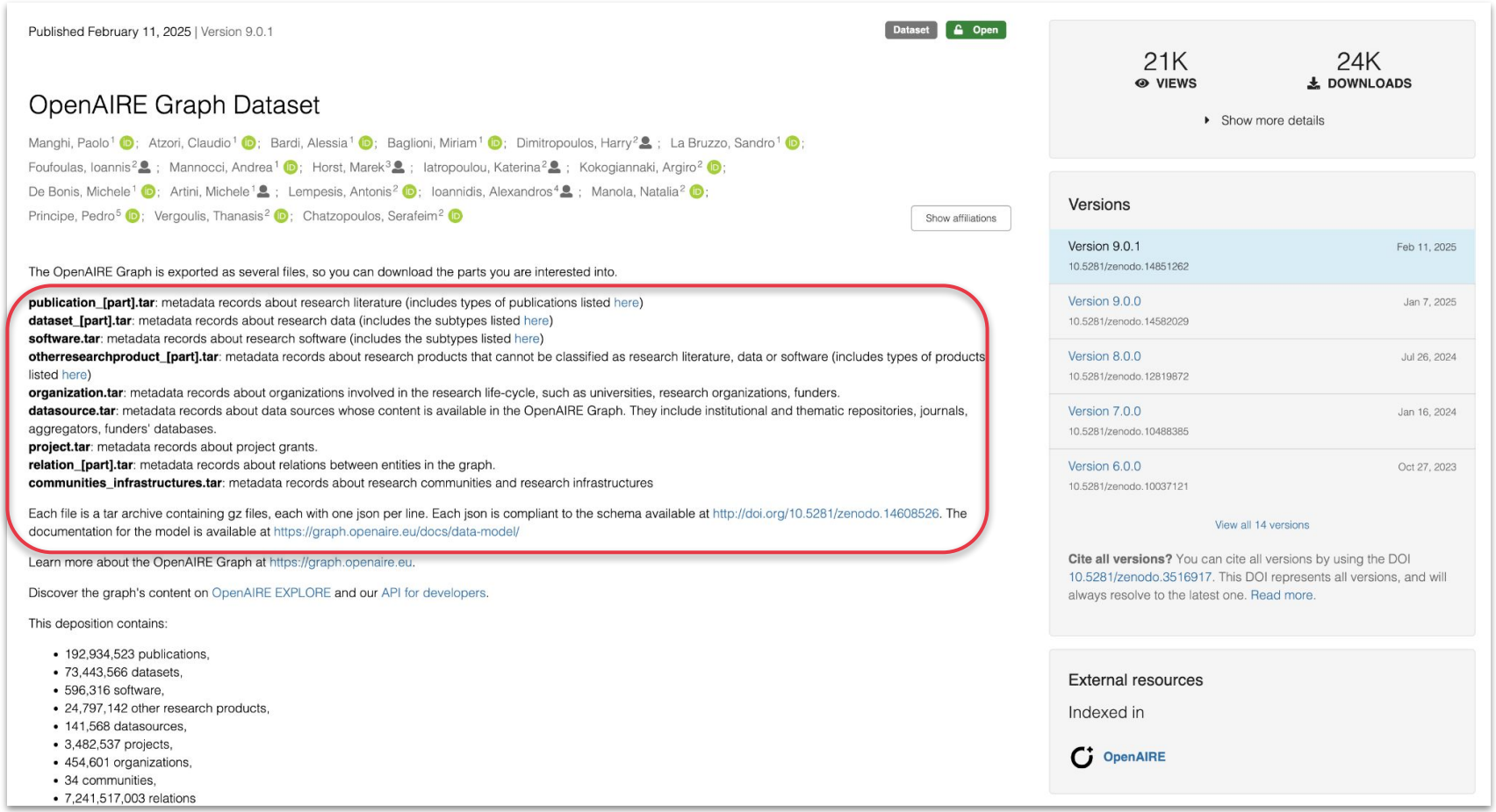
OpenAIRE

zenodo

https://zenodo.org/records/14851262

OpenAIRE

OpenAIRE Graph dataset



Published February 11, 2025 | Version 9.0.1

Dataset Open

OpenAIRE Graph Dataset

Manghi, Paolo[1]; Atzori, Claudio[1]; Bardi, Alessia[1]; Baglioni, Miriam[1]; Dimitropoulos, Harry[2]; La Bruzzo, Sandro[1]; Foufoulas, Ioannis[2]; Mannocci, Andrea[1]; Horst, Marek[3]; Iatropoulou, Katerina[2]; Kokogiannaki, Argiro[2]; De Bonis, Michele[1]; Artini, Michele[1]; Lempesis, Antonis[2]; Ioannidis, Alexandros[4]; Manola, Natalia[2]; Principe, Pedro[5]; Vergoulis, Thanasis[2]; Chatzopoulos, Serafeim[2]

Show affiliations

The OpenAIRE Graph is exported as several files, 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

Each file is a tar archive containing gz files, each with one json per line. Each json is compliant to the schema available at http://doi.org/10.5281/zenodo.14608526. The documentation for the model is available at https://graph.openaire.eu/docs/data-model/

Learn more about the OpenAIRE Graph at https://graph.openaire.eu.

Discover the graph's content on OpenAIRE EXPLORE and our API for developers.

This deposition contains:

- 192,934,523 publications,
- 73,443,566 datasets,
- 596,316 software,
- 24,797,142 other research products,
- 141,568 datasources,
- 3,482,537 projects,
- 454,601 organizations,
- 34 communities,
- 7,241,517,003 relations

21K VIEWS 24K DOWNLOADS

Show more details

Versions

Version 9.0.1 Feb 11, 2025
10.5281/zenodo.14851262

Version 9.0.0 Jan 7, 2025
10.5281/zenodo.14582029

Version 8.0.0 Jul 26, 2024
10.5281/zenodo.12819872

Version 7.0.0 Jan 16, 2024
10.5281/zenodo.10488385

Version 6.0.0 Oct 27, 2023
10.5281/zenodo.10037121

View all 14 versions

Cite all versions? You can cite all versions by using the DOI 10.5281/zenodo.3516917. This DOI represents all versions, and will always resolve to the latest one. Read more.

External resources

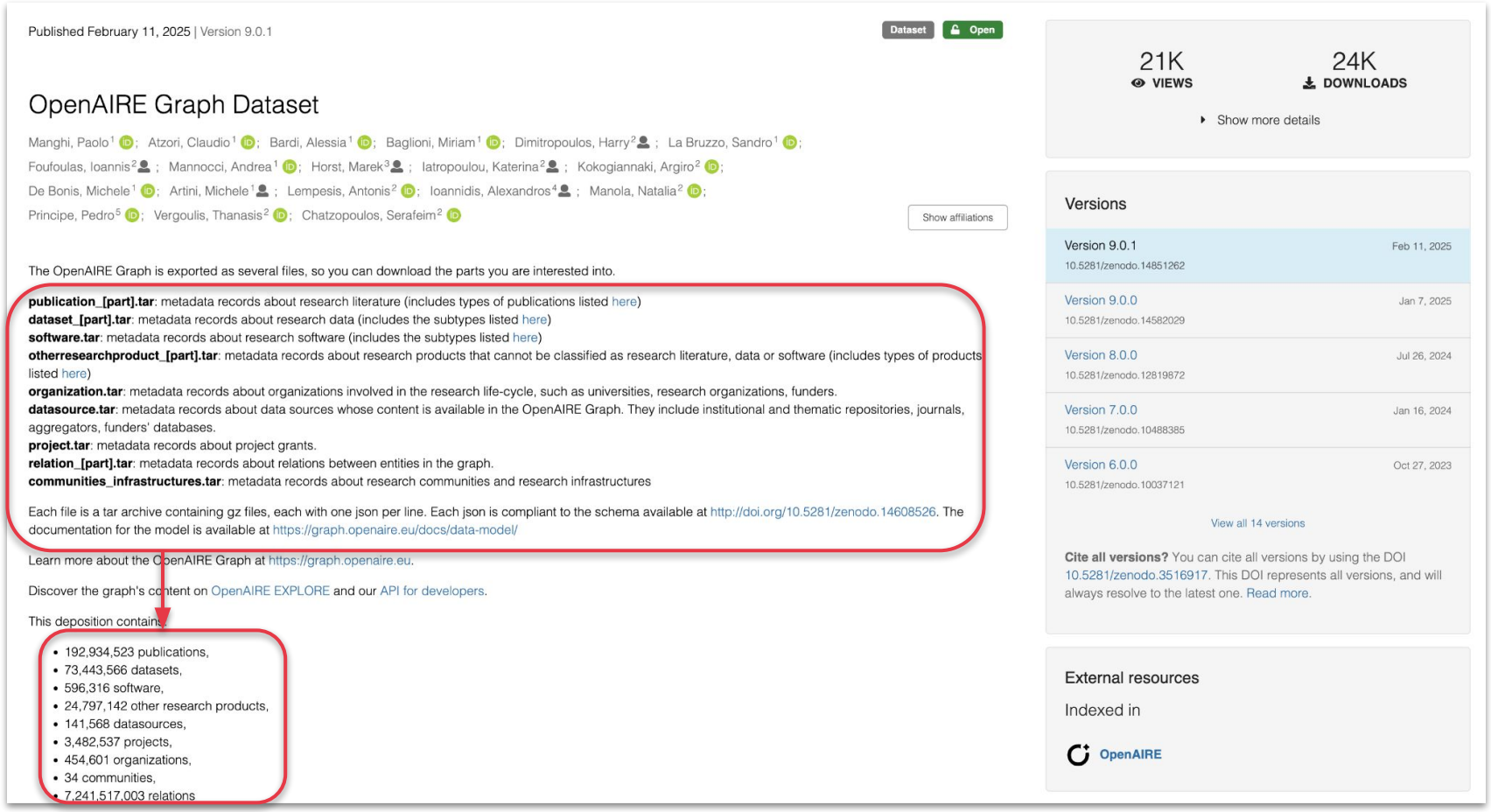
Indexed in

OpenAIRE

zenodo

https://zenodo.org/records/14851262

OpenAIRE Graph dataset



Published February 11, 2025 | Version 9.0.1

Dataset Open

OpenAIRE Graph Dataset

Manghi, Paolo[1]; Atzori, Claudio[1]; Bardi, Alessia[1]; Baglioni, Miriam[1]; Dimitropoulos, Harry[2]; La Bruzzo, Sandro[1]; Foufoulas, Ioannis[2]; Mannocci, Andrea[1]; Horst, Marek[3]; Iatropoulou, Katerina[2]; Kokogiannaki, Argiro[2]; De Bonis, Michele[1]; Artini, Michele[1]; Lempesis, Antonis[2]; Ioannidis, Alexandros[4]; Manola, Natalia[2]; Principe, Pedro[5]; Vergoulis, Thanasis[2]; Chatzopoulos, Serafeim[2]

Show affiliations

The OpenAIRE Graph is exported as several files, 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

Each file is a tar archive containing gz files, each with one json per line. Each json is compliant to the schema available at http://doi.org/10.5281/zenodo.14608526. The documentation for the model is available at https://graph.openaire.eu/docs/data-model/

Learn more about the OpenAIRE Graph at https://graph.openaire.eu.

Discover the graph's content on OpenAIRE EXPLORE and our API for developers.

This deposition contains:

- 192,934,523 publications,
- 73,443,566 datasets,
- 596,316 software,
- 24,797,142 other research products,
- 141,568 datasources,
- 3,482,537 projects,
- 454,601 organizations,
- 34 communities,
- 7,241,517,003 relations

21K VIEWS

24K DOWNLOADS

Show more details

Versions

Version 9.0.1 — Feb 11, 2025 — 10.5281/zenodo.14851262
Version 9.0.0 — Jan 7, 2025 — 10.5281/zenodo.14582029
Version 8.0.0 — Jul 26, 2024 — 10.5281/zenodo.12819872
Version 7.0.0 — Jan 16, 2024 — 10.5281/zenodo.10488385
Version 6.0.0 — Oct 27, 2023 — 10.5281/zenodo.10037121

View all 14 versions

Cite all versions? You can cite all versions by using the DOI 10.5281/zenodo.3516917. This DOI represents all versions, and will always resolve to the latest one. Read more.

External resources

Indexed in

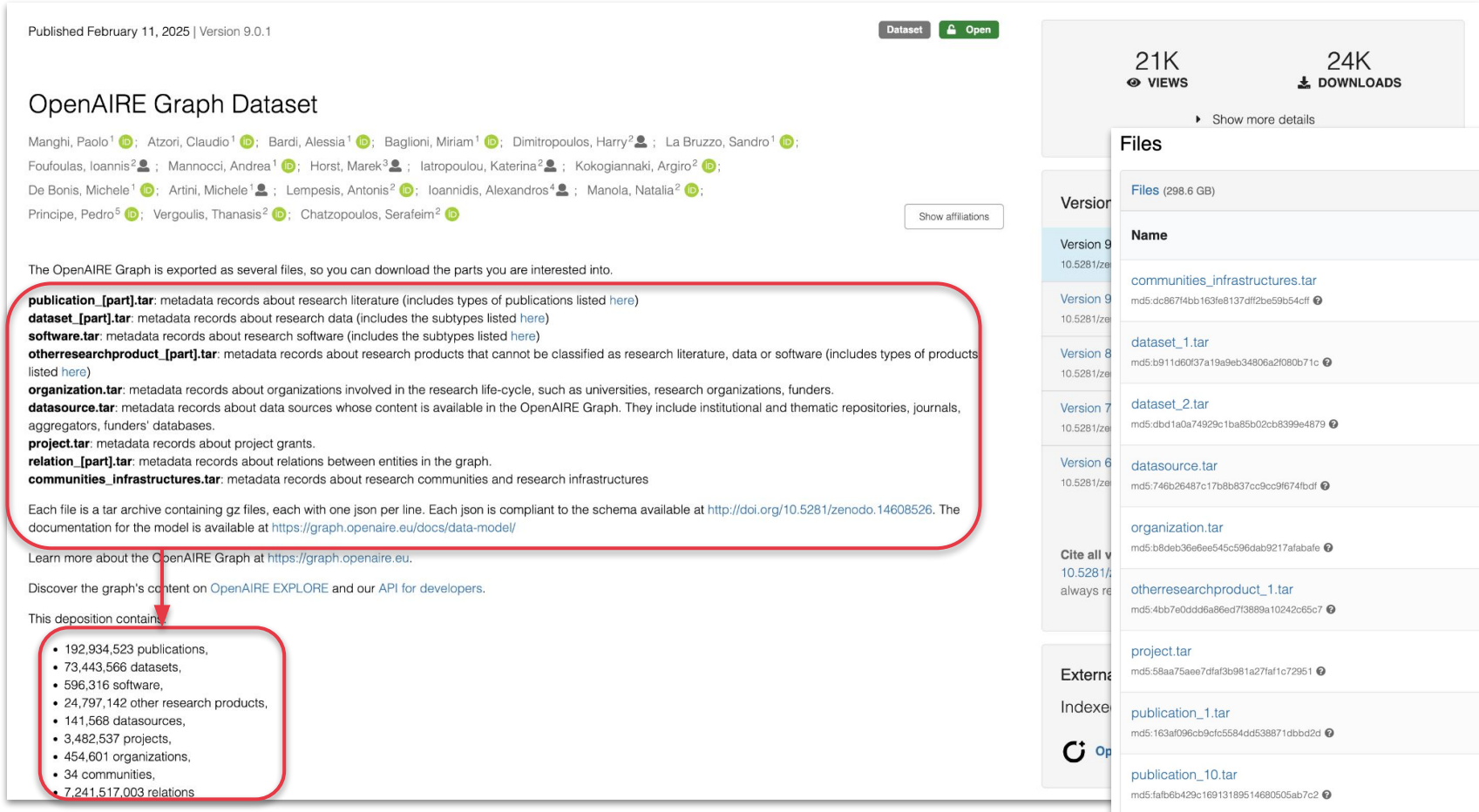
OpenAIRE

zenodo

https://zenodo.org/records/14851262

OpenAIRE

OpenAIRE Graph dataset



Published February 11, 2025 | Version 9.0.1

Dataset Open

OpenAIRE Graph Dataset

Manghi, Paolo[1]; Atzori, Claudio[1]; Bardi, Alessia[1]; Baglioni, Miriam[1]; Dimitropoulos, Harry[2]; La Bruzzo, Sandro[1]; Foufoulas, Ioannis[2]; Mannocci, Andrea[1]; Horst, Marek[3]; Iatropoulou, Katerina[2]; Kokogiannaki, Argiro[2]; De Bonis, Michele[1]; Artini, Michele[1]; Lempesis, Antonis[2]; Ioannidis, Alexandros[4]; Manola, Natalia[2]; Principe, Pedro[5]; Vergoulis, Thanasis[2]; Chatzopoulos, Serafeim[2]

Show affiliations

The OpenAIRE Graph is exported as several files, 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

Each file is a tar archive containing gz files, each with one json per line. Each json is compliant to the schema available at http://doi.org/10.5281/zenodo.14608526. The documentation for the model is available at https://graph.openaire.eu/docs/data-model/

Learn more about the OpenAIRE Graph at https://graph.openaire.eu.

Discover the graph's content on OpenAIRE EXPLORE and our API for developers.

This deposition contains:

- 192,934,523 publications,
- 73,443,566 datasets,
- 596,316 software,
- 24,797,142 other research products,
- 141,568 datasources,
- 3,482,537 projects,
- 454,601 organizations,
- 34 communities,
- 7,241,517,003 relations

21K VIEWS 24K DOWNLOADS

Show more details

Version

Version 9 10.5281/ze

Version 9 10.5281/ze

Version 8 10.5281/ze

Version 7 10.5281/ze

Version 6 10.5281/ze

Cite all v 10.5281/ always re

Externa

Indexe

Op

zenodo

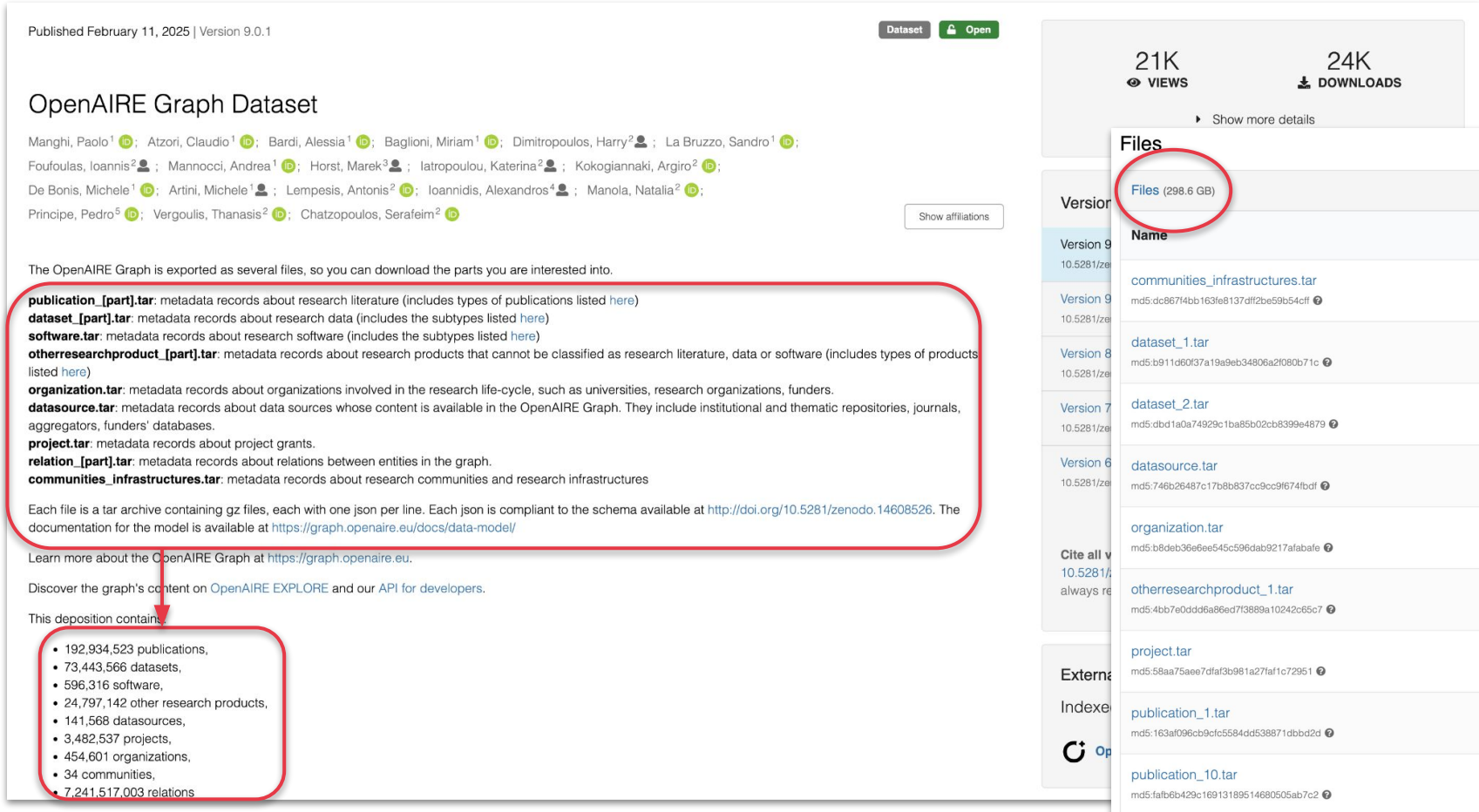
Files

Files (298.6 GB)

Name	Size	Download all
communities_infrastructures.tar md5:dc867f4bb163fe8137dff2be59b54cff	41.0 kB	Download
dataset_1.tar md5:b911d60f37a19a9eb34806a2f080b71c	10.8 GB	Download
dataset_2.tar md5:dbd1a0a74929c1ba85b02cb8399e4879	7.1 GB	Download
datasource.tar md5:746b26487c17b8b837cc9cc9f674fbdf	14.0 MB	Download
organization.tar md5:b8deb36e6ee545c596dab9217afabafe	33.9 MB	Download
otherresearchproduct_1.tar md5:4bb7e0ddd6a86ed7f3889a10242c65c7	9.1 GB	Download
project.tar md5:58aa75aee7dfaf3b981a27faf1c72951	577.9 MB	Download
publication_1.tar md5:163af096cb9cfc5584dd538871dbbd2d	10.8 GB	Download
publication_10.tar md5:fafb6b429c16913189514680505ab7c2	10.8 GB	Download
publication_11.tar md5:dca2418236a00e0d25a3002d832136ab	10.8 GB	Download
publication_12.tar md5:b8fce350a73716912ef10e47e27fd1b5	10.8 GB	Download

OpenAIRE

OpenAIRE Graph dataset



Published February 11, 2025 | Version 9.0.1

Dataset Open

OpenAIRE Graph Dataset

Manghi, Paolo[1]; Atzori, Claudio[1]; Bardi, Alessia[1]; Baglioni, Miriam[1]; Dimitropoulos, Harry[2]; La Bruzzo, Sandro[1]; Foufoulas, Ioannis[2]; Mannocci, Andrea[1]; Horst, Marek[3]; Iatropoulou, Katerina[2]; Kokogiannaki, Argiro[2]; De Bonis, Michele[1]; Artini, Michele[1]; Lempesis, Antonis[2]; Ioannidis, Alexandros[4]; Manola, Natalia[2]; Principe, Pedro[5]; Vergoulis, Thanasis[2]; Chatzopoulos, Serafeim[2]

Show affiliations

The OpenAIRE Graph is exported as several files, 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

Each file is a tar archive containing gz files, each with one json per line. Each json is compliant to the schema available at http://doi.org/10.5281/zenodo.14608526. The documentation for the model is available at https://graph.openaire.eu/docs/data-model/

Learn more about the OpenAIRE Graph at https://graph.openaire.eu.

Discover the graph's content on OpenAIRE EXPLORE and our API for developers.

This deposition contains:

- 192,934,523 publications,
- 73,443,566 datasets,
- 596,316 software,
- 24,797,142 other research products,
- 141,568 datasources,
- 3,482,537 projects,
- 454,601 organizations,
- 34 communities,
- 7,241,517,003 relations

21K VIEWS 24K DOWNLOADS

Show more details

Version

Version 9 10.5281/ze
Version 9 10.5281/ze
Version 8 10.5281/ze
Version 7 10.5281/ze
Version 6 10.5281/ze

Cite all v 10.5281/ always re

Externa

Indexe

Op

zenodo

Files

Files (298.6 GB)

Name	Size	Download all
communities_infrastructures.tar md5:dc867f4bb163fe8137dff2be59b54cff	41.0 kB	Download
dataset_1.tar md5:b911d60f37a19a9eb34806a2f080b71c	10.8 GB	Download
dataset_2.tar md5:dbd1a0a74929c1ba85b02cb8399e4879	7.1 GB	Download
datasource.tar md5:746b26487c17b8b837cc9cc9f674fbdf	14.0 MB	Download
organization.tar md5:b8deb36e6ee545c596dab9217afabafe	33.9 MB	Download
otherresearchproduct_1.tar md5:4bb7e0ddd6a86ed7f3889a10242c65c7	9.1 GB	Download
project.tar md5:58aa75aee7dfaf3b981a27faf1c72951	577.9 MB	Download
publication_1.tar md5:163af096cb9cfc5584dd538871dbbd2d	10.8 GB	Download
publication_10.tar md5:fafb6b429c16913189514680505ab7c2	10.8 GB	Download
publication_11.tar md5:dca2418236a00e0d25a3002d832136ab	10.8 GB	Download
publication_12.tar md5:b8fce350a73716912ef10e47e27fd1b5	10.8 GB	Download

OpenAIRE

OpenAIRE Graph dataset

Suitable

- If you want to have **full control** on data and custom preprocessing
- If you have **technical skills** to process raw JSON data
- If you have adequate **computational power** and storage
  - Server, cluster
  - Cloud computing
  - DBMS, Graph DB, …

Unsuitable

- If not comfortable getting your hands dirty
- If a local computer is all you have at disposal

OpenAIRE Graph
Beginner’s kit

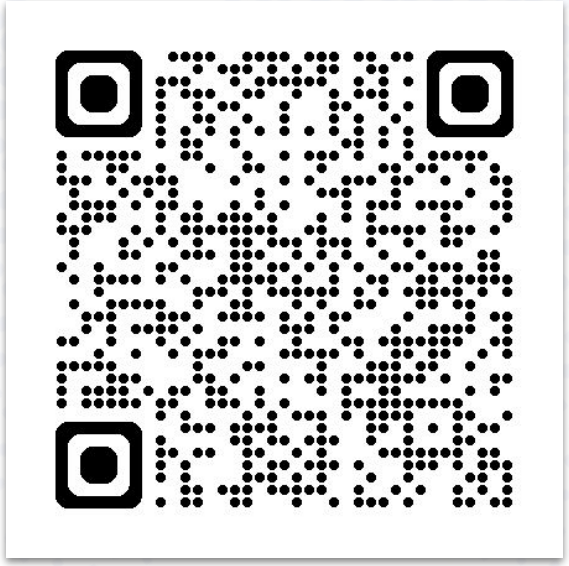
OpenAIRE Beginner's kit

Underlying ideas

- Familiarise with data and tools
- On your standard computer
- Cost-free
- Hassle-free (almost zero configuration)
- Catch: “small” subset of data
- Scale up later

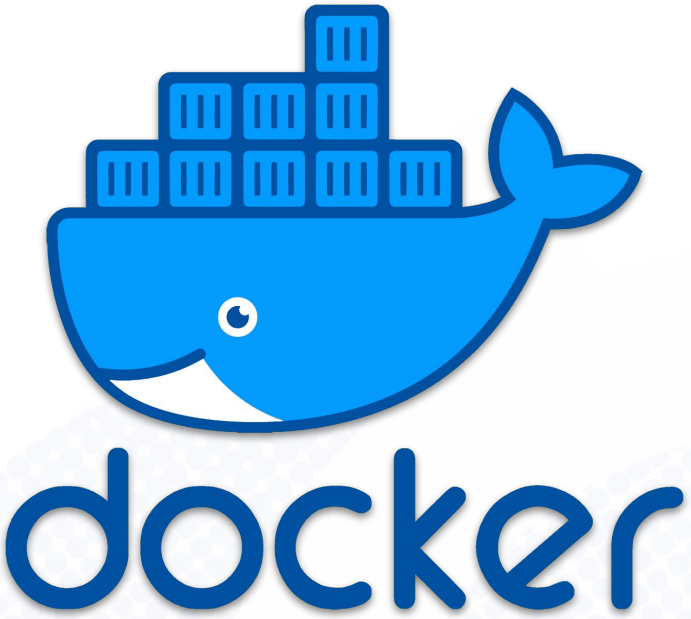
GitHub → https://github.com/openaire/beginners-kit

Zenodo → https://zenodo.org/doi/10.5281/zenodo.10841263



OpenAIRE

OpenAIRE Beginner's kit



Simple setup

- Download and install Docker engine
- Build image

```
docker build --rm -t openaire-beginners-kit .
```

- Run the container

```
docker run --name kit-container -p 8889:8889 --rm openaire-beginners-kit
```

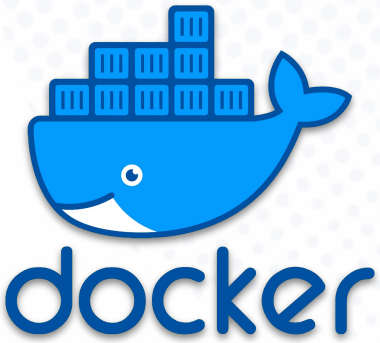
OpenAIRE Beginner's kit

What did I just do?

- Created a **Docker container** virtualising an Apache Hadoop cluster and providing a **Jupyter Lab** instance for playing with the data from a Python notebook

From the **Python notebook**

- Fetch a Graph subset (JSON files) from Zenodo, https://zenodo.org/doi/10.5281/zenodo.7490191
- Virtualise the presence of relational DB via SparkSQL
- Perform **queries over JSON files with SQL-like syntax**

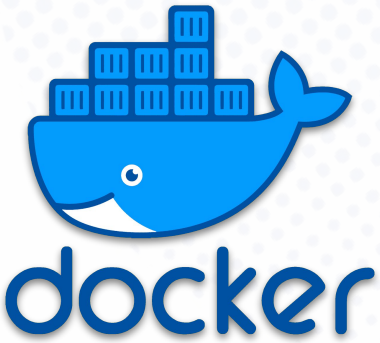


OpenAIRE

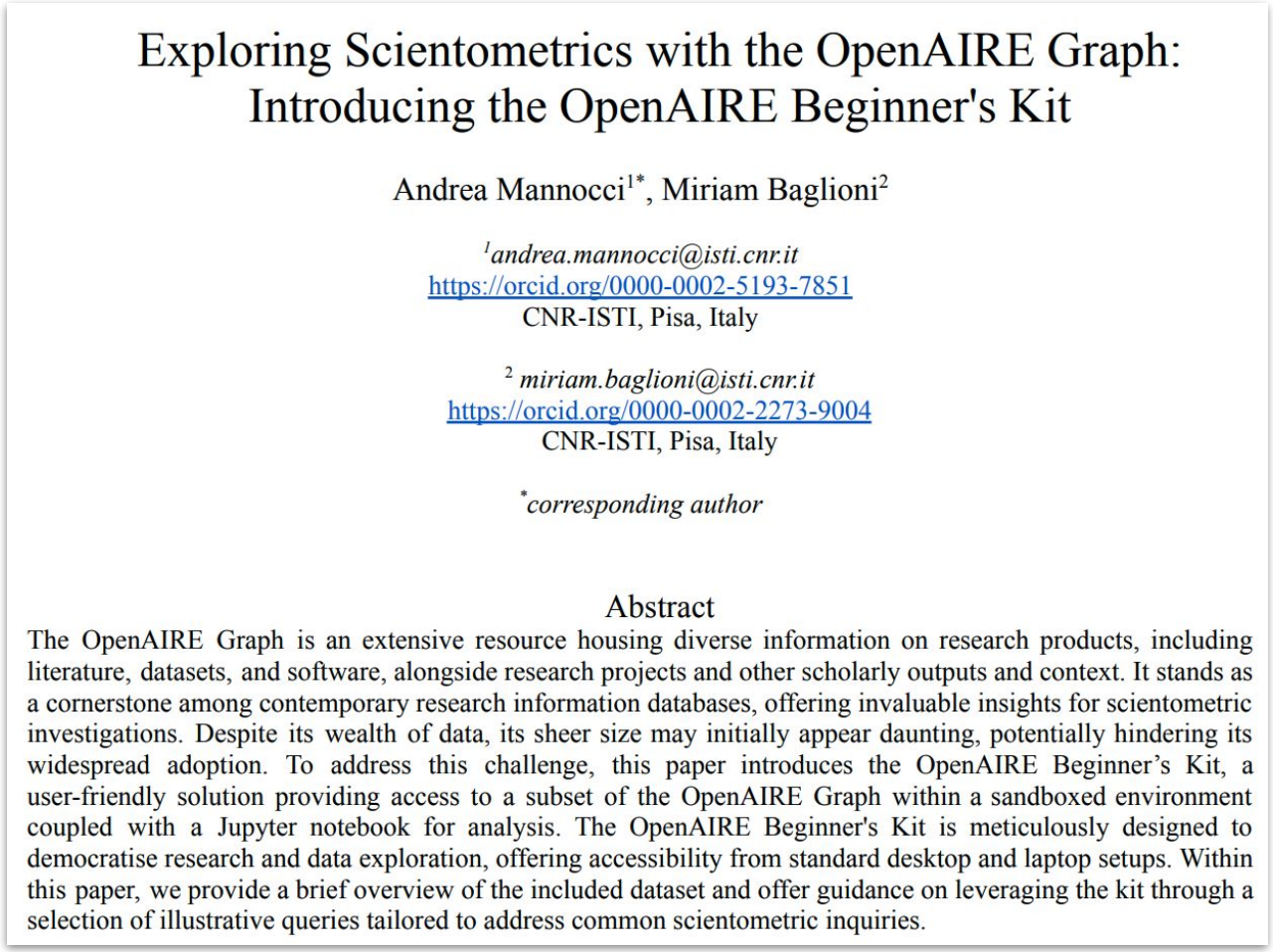
OpenAIRE Beginner's kit

For example

```
SELECT publications.id, pid.value, COUNT(*) AS count
FROM publications
JOIN relations
 ON publications.id = relations.source
WHERE reltype.name = 'IsCitedBy'
GROUP BY publications.id, pid.value
ORDER BY count DESC
```



To read further



Exploring Scientometrics with the OpenAIRE Graph: Introducing the OpenAIRE Beginner's Kit

Andrea Mannocci[1*], Miriam Baglioni[2]

[1] *andrea.mannocci@isti.cnr.it*
https://orcid.org/0000-0002-5193-7851
CNR-ISTI, Pisa, Italy

[2] *miriam.baglioni@isti.cnr.it*
https://orcid.org/0000-0002-2273-9004
CNR-ISTI, Pisa, Italy

[*] *corresponding author*

Abstract

The OpenAIRE Graph is an extensive resource housing diverse information on research products, including literature, datasets, and software, alongside research projects and other scholarly outputs and context. It stands as a cornerstone among contemporary research information databases, offering invaluable insights for scientometric investigations. Despite its wealth of data, its sheer size may initially appear daunting, potentially hindering its widespread adoption. To address this challenge, this paper introduces the OpenAIRE Beginner's Kit, a user-friendly solution providing access to a subset of the OpenAIRE Graph within a sandboxed environment coupled with a Jupyter notebook for analysis. The OpenAIRE Beginner's Kit is meticulously designed to democratise research and data exploration, offering accessibility from standard desktop and laptop setups. Within this paper, we provide a brief overview of the included dataset and offer guidance on leveraging the kit through a selection of illustrative queries tailored to address common scientometric inquiries.

Mannocci, A., & Baglioni, M. (2024, September 19). Exploring Scientometrics with the OpenAIRE Graph: Introducing the OpenAIRE Beginner's Kit. **28th International Conference on Science, Technology and Innovation Indicators (STI2024)**, Berlin, Germany.

https://doi.org/10.5281/zenodo.13942507

OpenAIRE Graph
Cloud Access

Google Cloud Platform

- Ecosystem of commercial tools for “cloud computing”
- Regular services, such as hosting, dedicated virtual machines, databases
- Different pricing schemes for the tools: mix of **processing and storage**
- All tools integrated in a common authentication system: easy transfer
- Used in the web browser or as an SDK installed locally

Special applications relevant for us

- **BigQuery** data warehousing
- **Jupyter Notebooks** (literate programming) as Google Colaboratory
- (Google Cloud Storage)

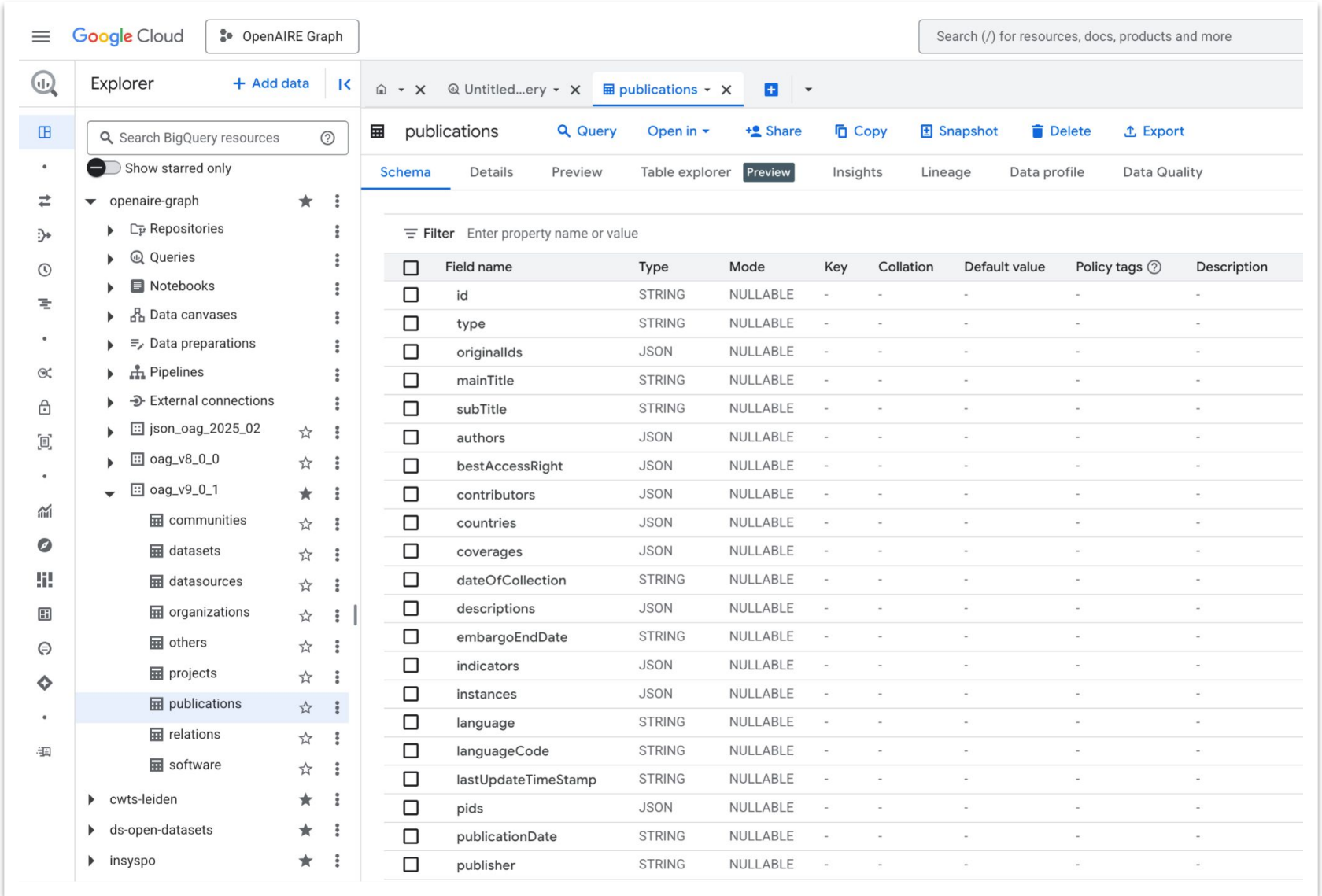
Google BigQuery

- Serverless data warehouse
- Several tools for data ingress/egress/streaming
- Integration with Google Cloud Storage and Colab notebooks
- **Low maintenance cost, charged mostly by queries**
- Easy to collaborate and share datasets
- Private and public datasets
- Capacity to query petabytes
- Tabular data with arbitrarily complex data structures (arrays, JSON)
- **SQL-like** data manipulation and access language
- Web interface and API

Basic tasks

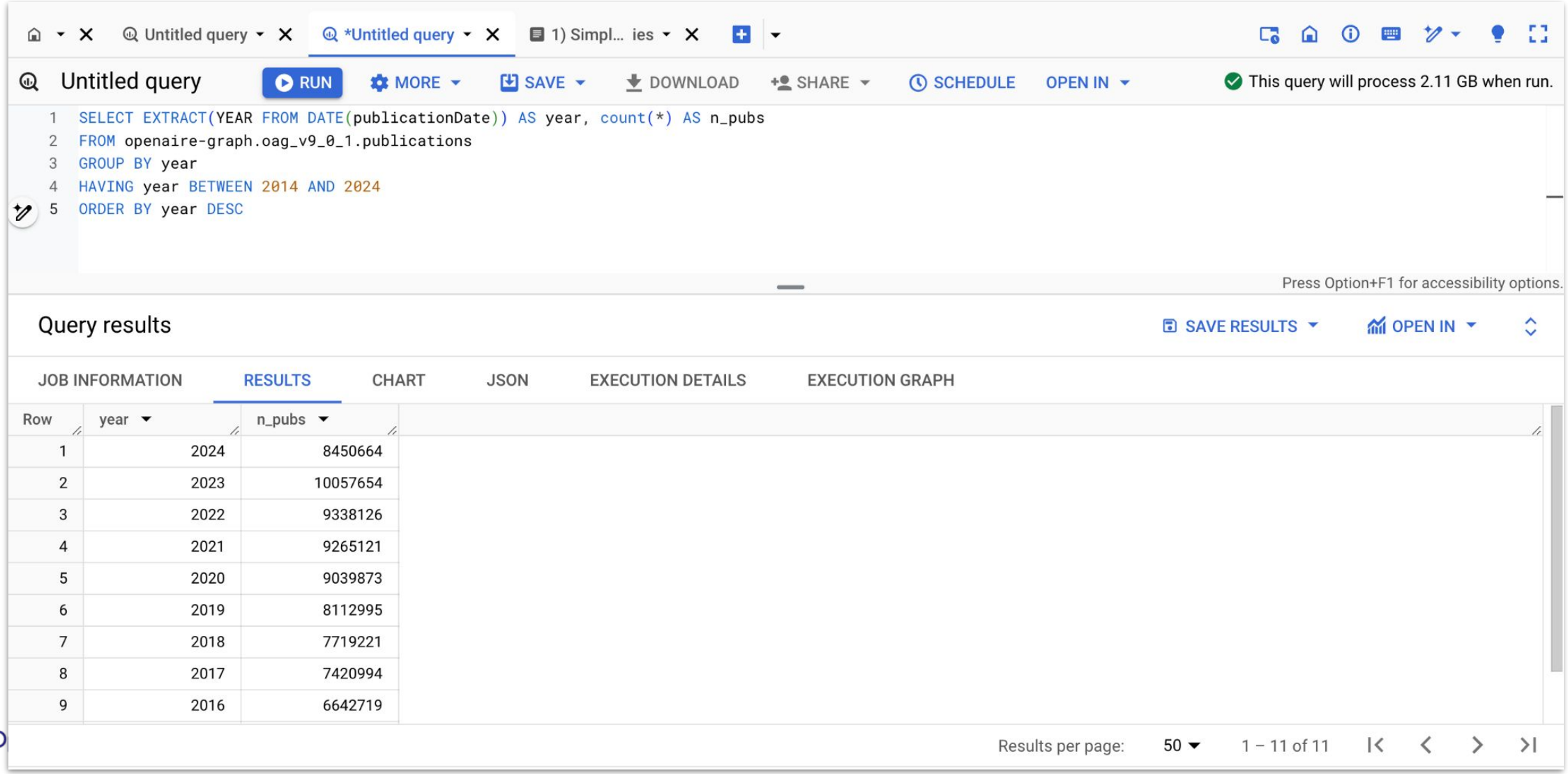
- Import, export of data
- Create or copy datasets and tables
- Query data
  - Query window in **BigQuery Studio**
  - Embedded Google **Colaboratory Notebook**

BigQuery interface

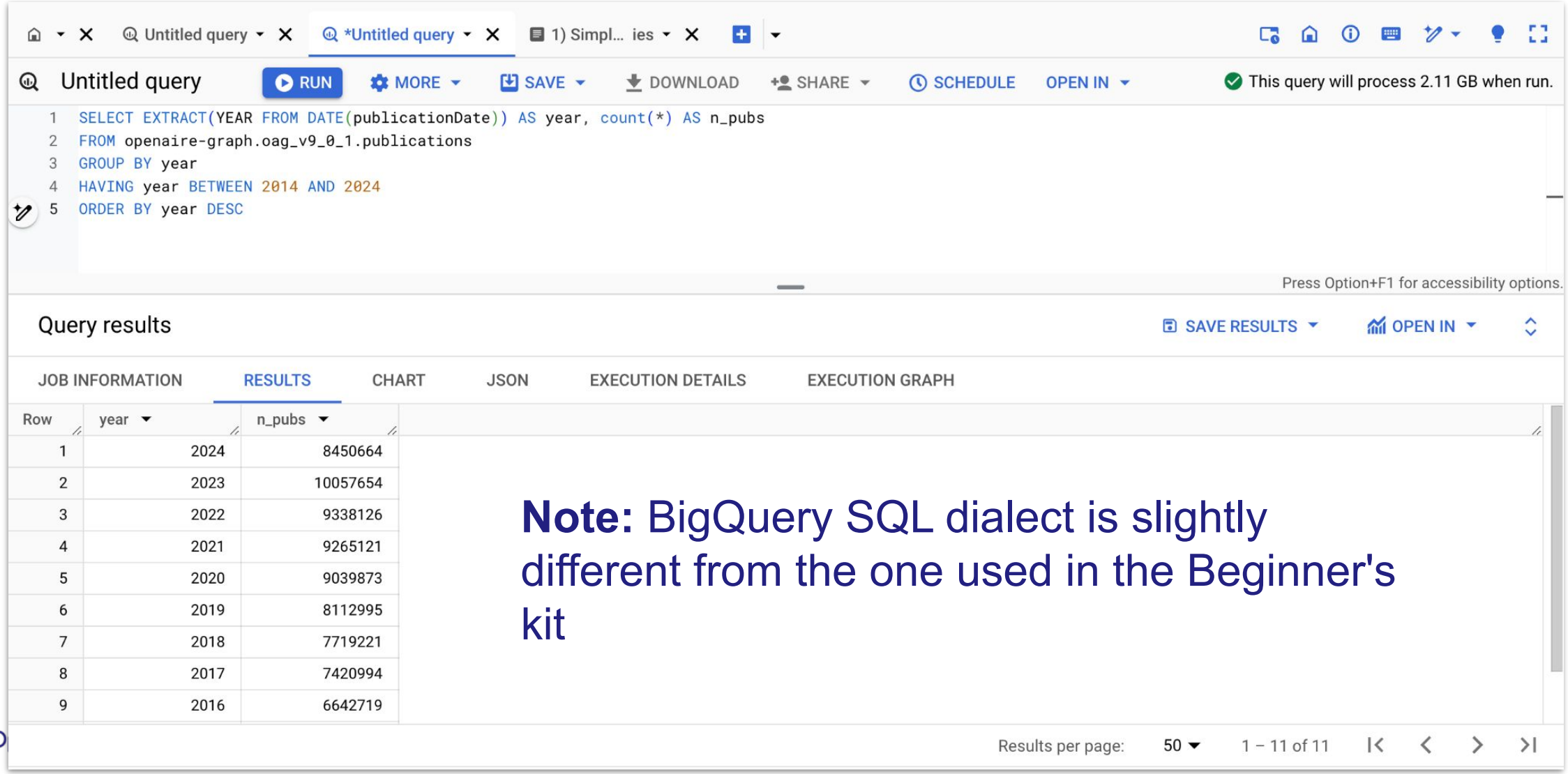


OpenAIRE

BigQuery interface

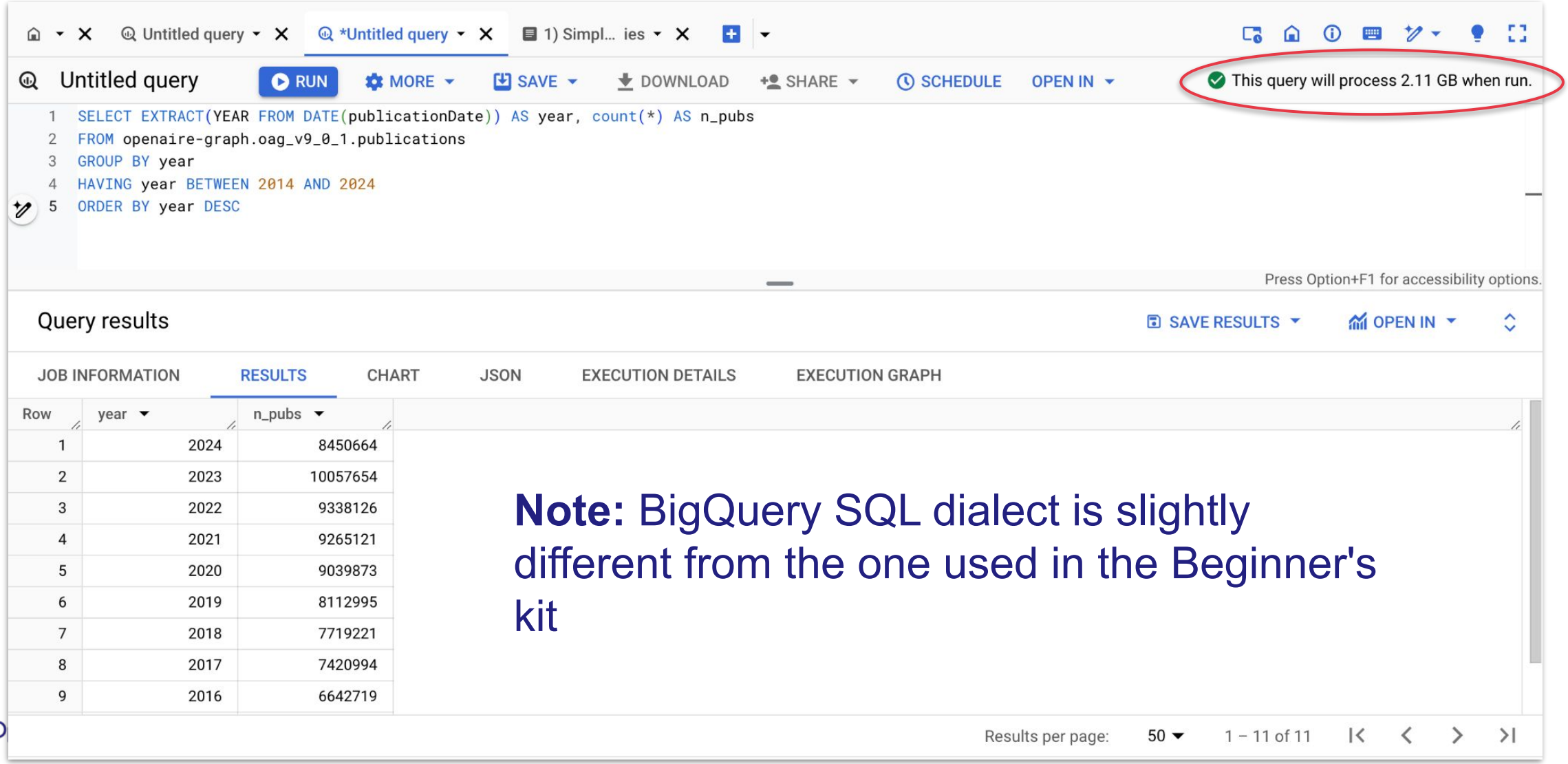


BigQuery interface



Note: BigQuery SQL dialect is slightly different from the one used in the Beginner's kit

BigQuery interface



Note: BigQuery SQL dialect is slightly different from the one used in the Beginner's kit

Budget information

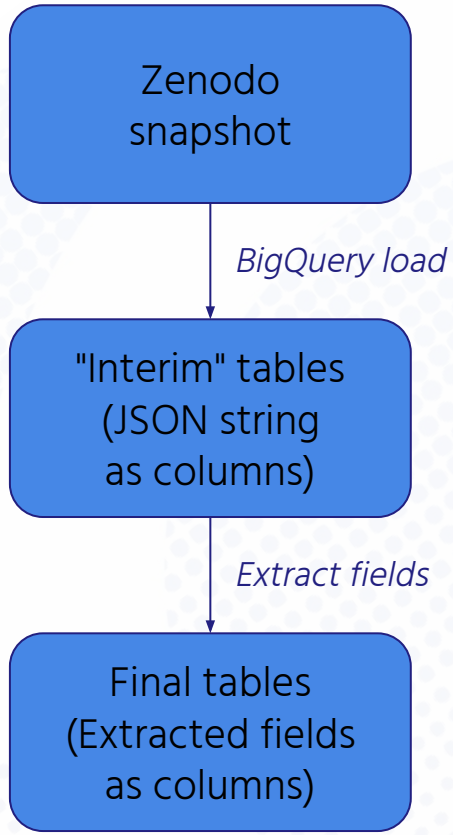
- **"Data parking" cost are covered by OpenAIRE**
  - The public dataset on BigQuery mirrors the OpenAIRE Graph dataset already published on Zenodo
- **Processing costs** for querying data are billed to individual accounts and projects

Budget information

- For every Google account, there is a **US$ 300.00 credit**, activated using a credit card (not charged). **Trial period for 3 months or credit expired.**
- Possible to **apply for research credits**, https://edu.google.com/intl/ALL_us/programs/credits/research
- Enough for large number of queries (several terabytes), a good solution for experimentation or courses
  - **Free 1st Terabyte** of query data processed **per month**
  - Roughly **$6.25 per Terabyte**

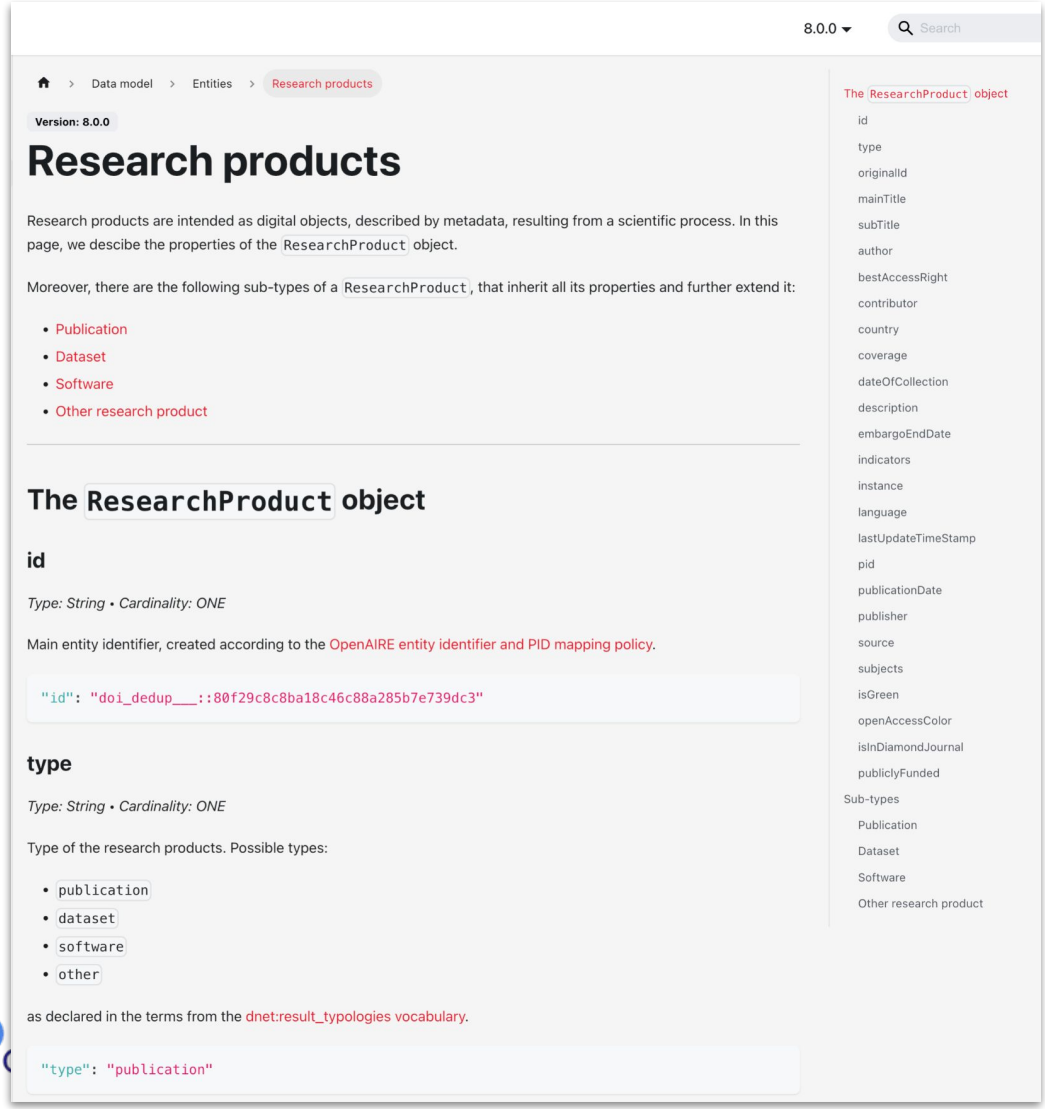
OpenAIRE Graph @ Google

- Zenodo snapshot from **December 2025**, https://zenodo.org/records/17725827
- Documentation online
  - → https://graph.openaire.eu/docs/data-model
  - → https://graph.openaire.eu/docs/cloud-access

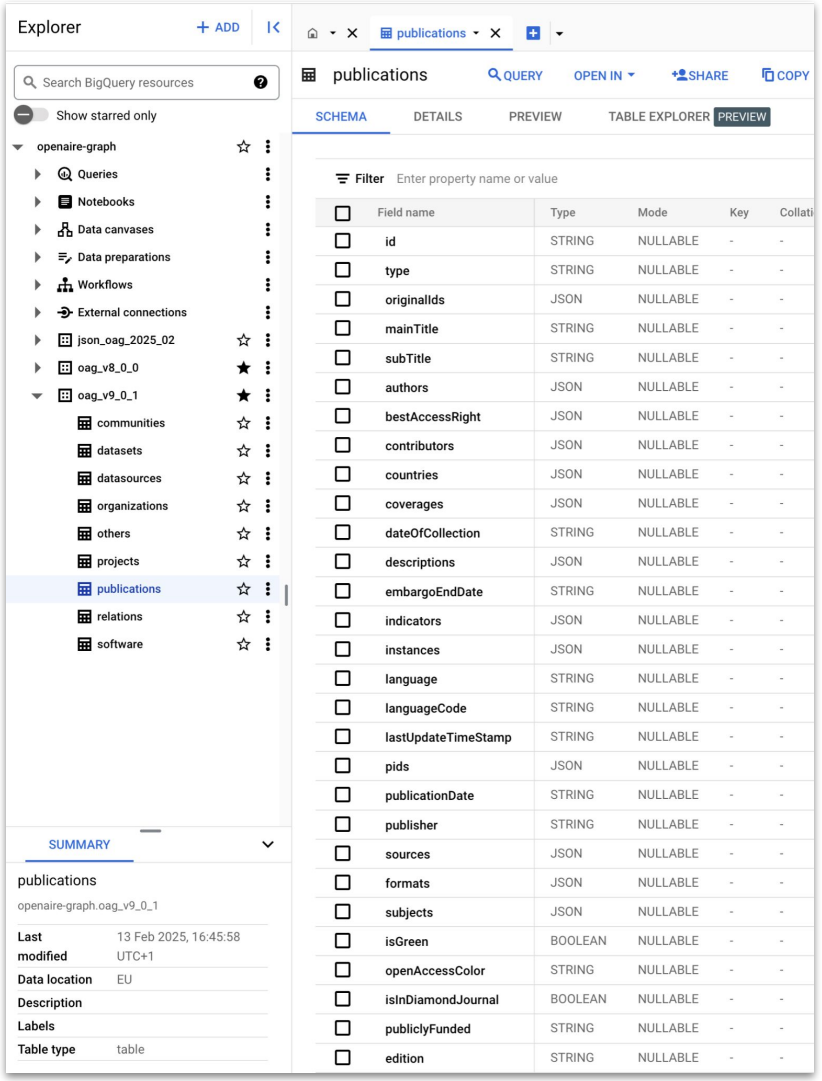


OpenAIRE

OpenAIRE Graph @ Google



"1-on-1 mapping"

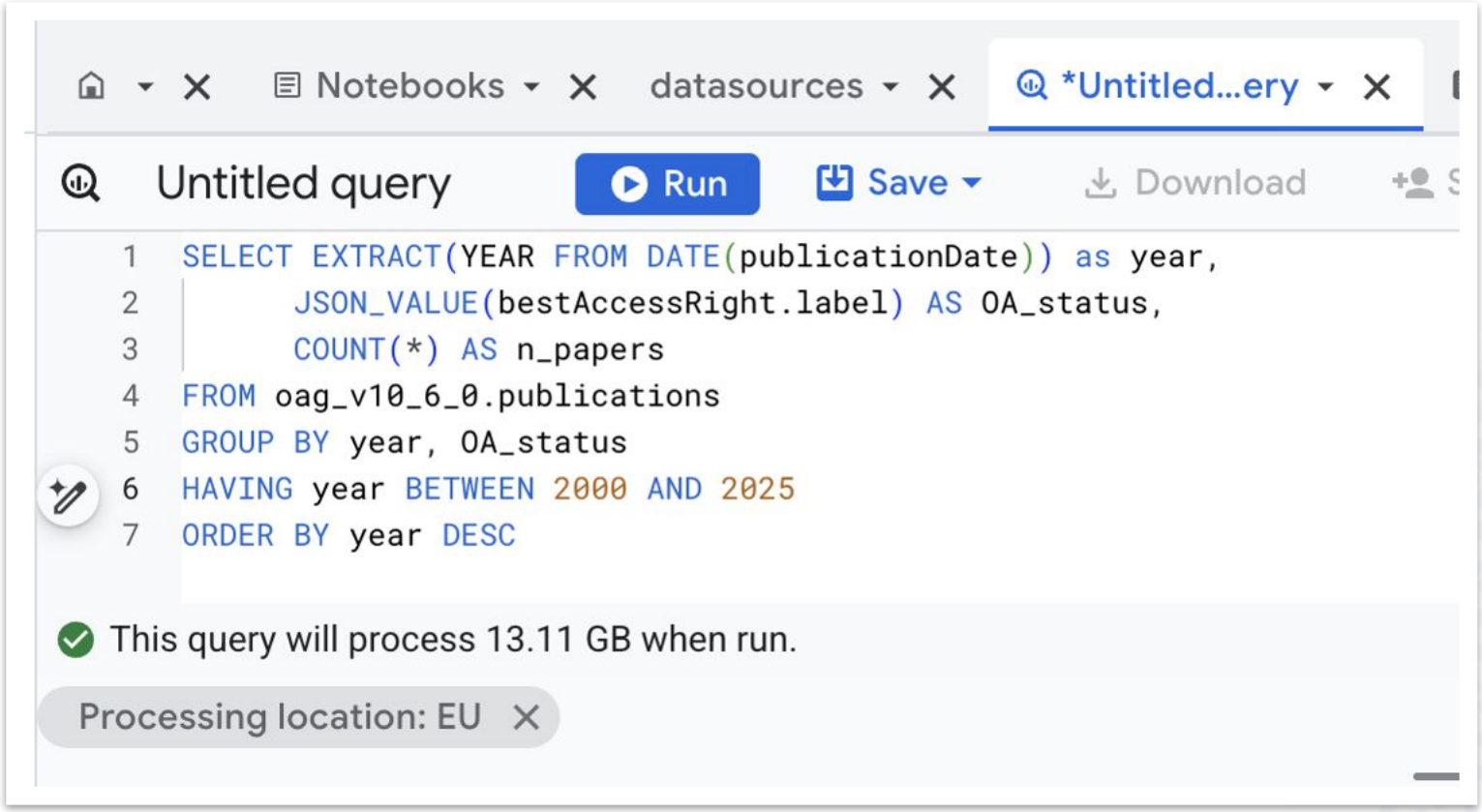


OpenAIRE Graph @ Google BigQuery

Tables on BigQuery

- **publications** - research products (e.g., article, thesis, peer-review, blog posts, books, reports, patents, etc.);
- **datasets** - research products referring to self-contained, persistently-identified digital assets intended for processing (e.g., files containing raw data, tables, metadata collections, dumps; persistent dynamic queries to scientific databases);
- **software** - research products referring to source code files, algorithms, scripts, computational workflows, and executables that were created during the research process or for a research purpose;
- **others** - any digital asset, uniquely identified, whose nature does not fall under the first three types described above;
- **datasources** - metadata referring to services where published material (metadata and files) is stored, preserved, and made discoverable and accessible;
- **organizations** - metadata about academic institutions, research centres, funders, or any other institutions taking part in the research process;
- **projects** - metadata about funding awarded to a person or an organisation by a funding body;
- **communities** - contains metadata about the research communities/infrastructures registered in OpenAIRE;
- **relations** - contains the relations between couples of source and target IDs. The semantics of the relations follow the DataCite semantic definition, extended by specific semantics.

Query example



OpenAIRE

Query example

Notebooks | datasources | *Untitled...ery

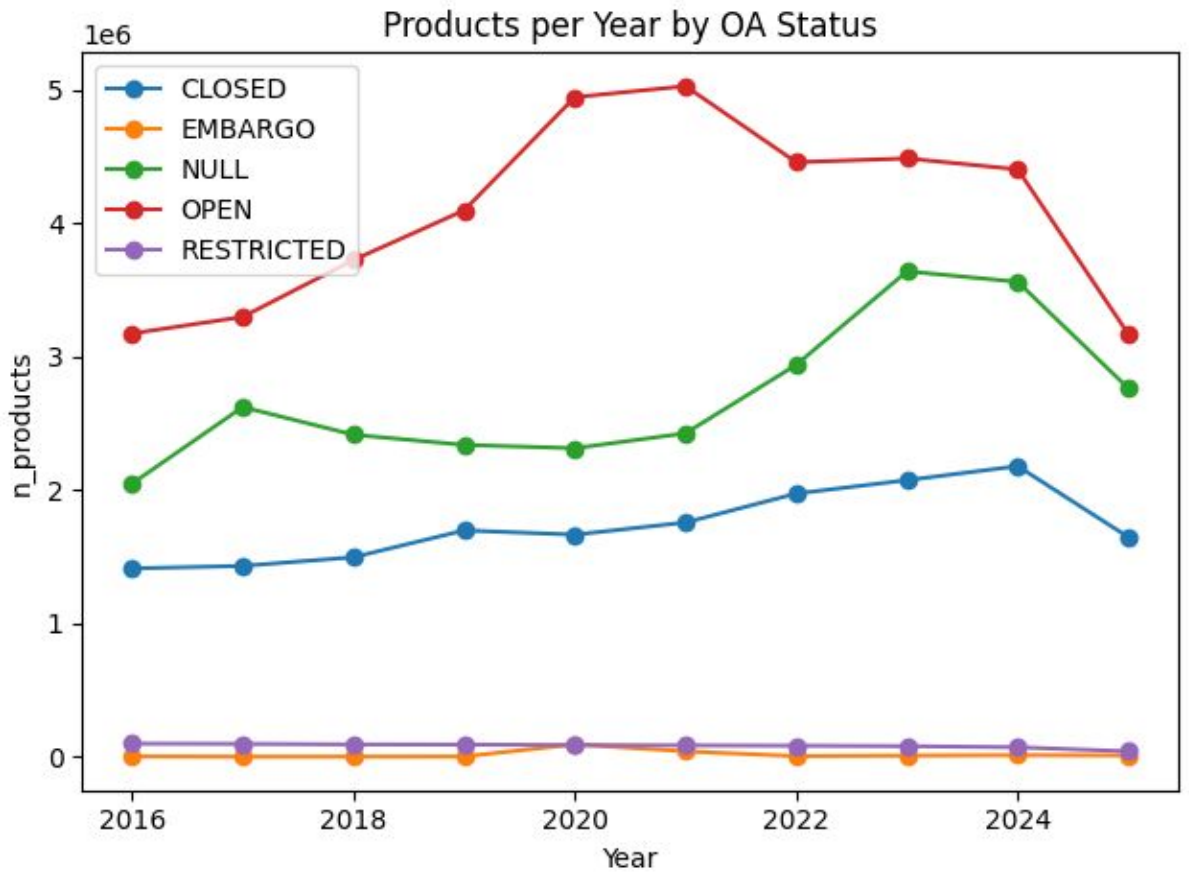
Untitled query

Run | Save

```
1  SELECT EXTRACT(YEAR FROM DATE(publicationDate))
2        JSON_VALUE(bestAccessRight.label) AS OA_s
3        COUNT(*) AS n_papers
4  FROM oag_v10_6_0.publications
5  GROUP BY year, OA_status
6  HAVING year BETWEEN 2000 AND 2025
7  ORDER BY year DESC
```

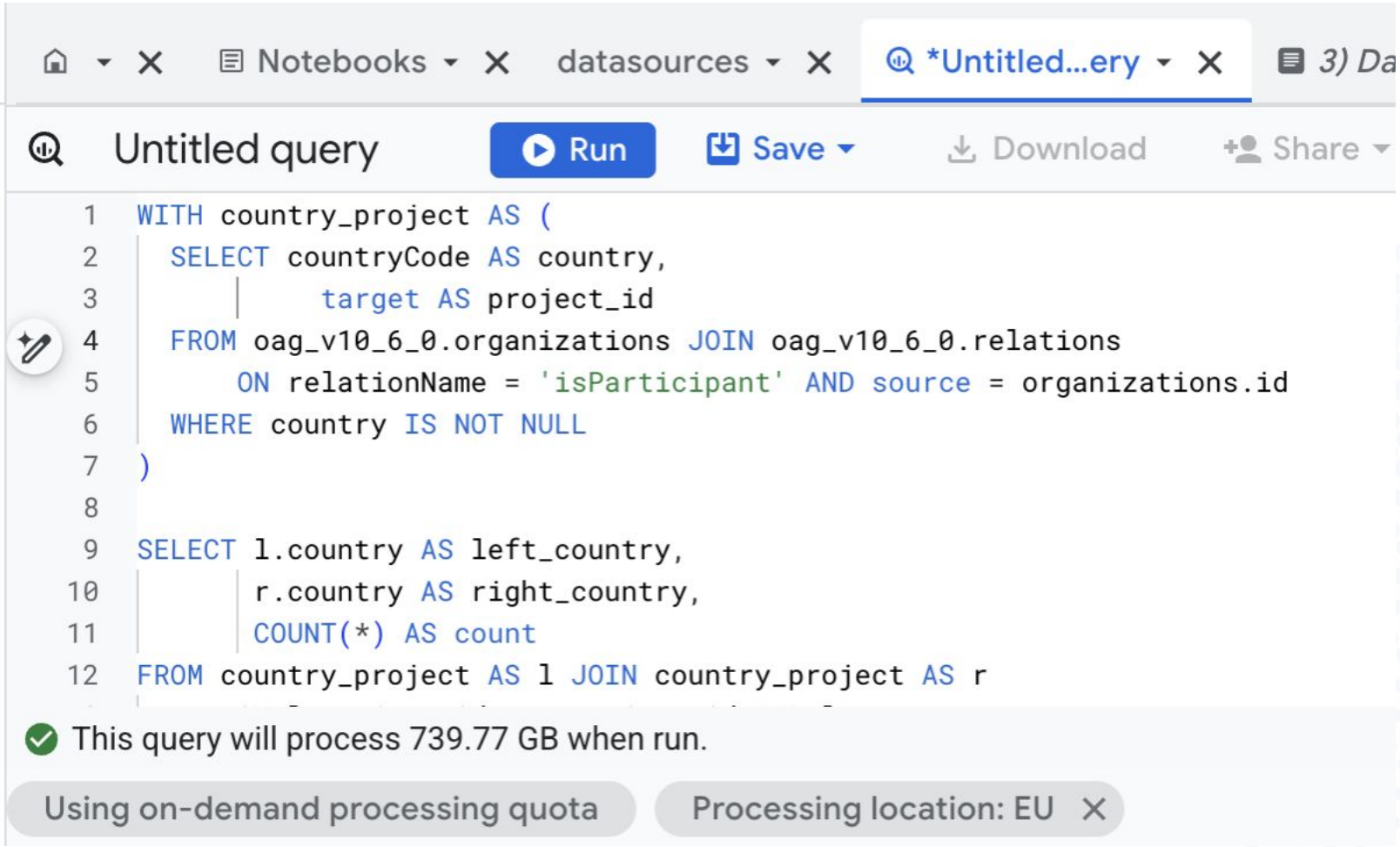
This query will process 13.11 GB when run.

Processing location: EU



OpenAIRE

Query example



OpenAIRE

Query example

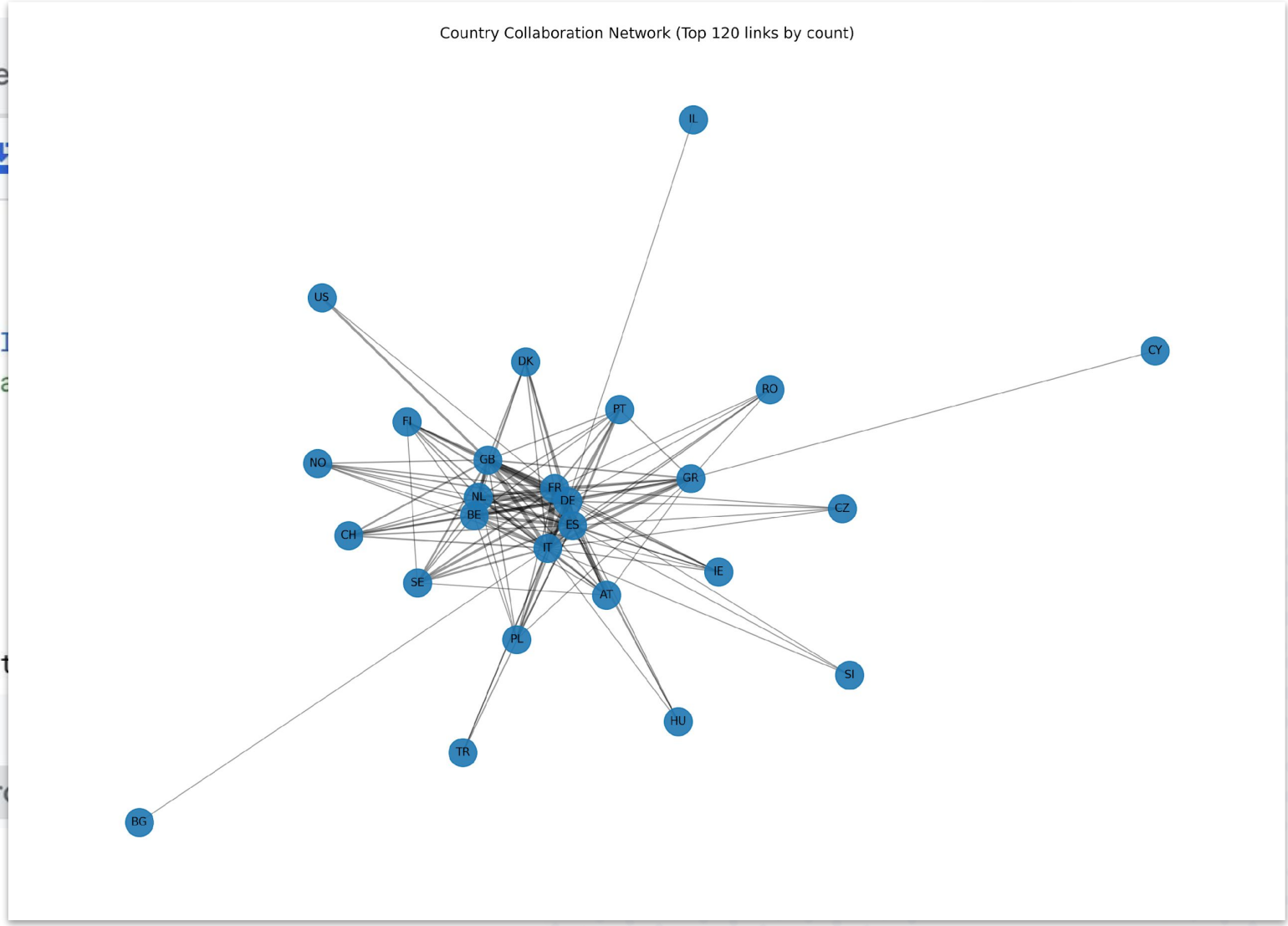
Notebooks ✕ datasource

Untitled query Run

```
1  WITH country_project AS (
2    SELECT countryCode AS country,
3         |     target AS project_id
4    FROM oag_v10_6_0.organizations JOI
5        ON relationName = 'isParticipa
6    WHERE country IS NOT NULL
7  )
8
9  SELECT l.country AS left_country,
10       | r.country AS right_country,
11       | COUNT(*) AS count
12 FROM country_project AS l JOIN count
```

This query will process 739.77 GB when run.

Using on-demand processing quota Pr



OpenAIRE

Current research lines

Charting the Landscape of Italian Diamond Open Access Publishing

Simone Angioni[1, [0000-0002-6682-3419]], **Miriam Baglioni**[1, [0000-0002-2273-9004]], **Alessia Bardi**[1, [0000-0002-1112-1292]], **Paolo Manghi**[1,2, [0000-0001-7291-3210]], **Andrea Mannocci**[1, [0000-0002-5193-7851]], **and Gina Pavone**[1, [0000-0003-0087-2151]]

[1]**CNR-ISTI — National Research Council, Institute of Information Science and Technologies "Alessandro Faedo", 56124 Pisa, Italy**
[2]**OpenAIRE AMKE, Athens, Greece**
[*]**Corresponding author: andrea.mannocci@isti.cnr.it**

ABSTRACT

Diamond Open Access (DOA) is a non-commercial model of scholarly publishing that removes financial barriers for authors and readers. While international studies have outlined the global uptake of DOA, this paper investigates the presence and characteristics of the DOA landscape in Italy. We conducted a quantitative analysis on the 168 Italian journals classified as Diamond by EZB, and we studied their publishing volume, disciplinary distribution and citation impact by integrating information from the following open resources: DOAJ, OpenAIRE, ROAD, SCImago Journal Rank, and the ANVUR classification used in the Italian Research Assessment Framework (VQR). Key findings include the significant growth of DOA journals, particularly in the social sciences and humanities, as well as the high level of international citations, indicating strong global relevance. The study also highlights challenges such as the need for better indexing and comprehensive data to fully capture the DOA landscape.

Accepted for publication in Quantitative Science Studies (QSS)
Preprint: https://zenodo.org/records/17484911

OpenAIRE

Latest research lines

Operationalising Bibliodiversity via the OpenAIRE Graph: A Multidimensional Demonstration

Simone Angioni[0000-0002-6682-3419]1, Miriam Baglioni[0000-0002-2273-9004]1, Andrea Mannocci[0000-0002-1234-5678]1*

[1]CNR-ISTI, Pisa, Italy.

Abstract

Bibliodiversity has emerged as a central concept in debates on open science, scholarly communication, and research evaluation, highlighting the need to sustain linguistic, geographic, epistemic, and infrastructural diversity in knowledge production and dissemination. While its normative foundations are well established, operationalising bibliodiversity at scale remains methodologically challenging and strongly dependent on the availability of inclusive, interoperable data infrastructures.

This paper presents a methodological demonstration of how bibliodiversity can be systematically analysed using the OpenAIRE Graph, a large-scale, open Scholarly Knowledge Graph that aggregates metadata from publishers, repositories, and research infrastructures worldwide. Drawing on the multidimensional framework proposed in the literature, we illustrate how key dimensions of bibliodiversity – linguistic, geographic, disciplinary, publisher, business model, and infrastructure diversity – can be operationalised using the OpenAIRE data model and queried reproducibly in Google BigQuery. Through a series of exploratory analyses, we show how the Graph captures the "long tail" of scholarly communication, including non-English products, diverse research formats such as datasets and software, and contributions from institutional and thematic repositories, particularly beyond the Global North. Rather than identifying new empirical trends, this work demonstrates the analytical potential of the OpenAIRE Graph as an open, community-governed infrastructure for bibliodiversity research. By making diversity measurable across multiple, interconnected dimensions, the OpenAIRE Graph provides a robust empirical basis for future studies and supports evidence-informed policies to foster a more inclusive and equitable scholarly communication ecosystem.

Submitted to Scientometrics Special Issue on Oper Research Information

Latest research lines

- Identification of author collaborations clusters and author roles
- Automated verbalization using data from the OpenAIRE Graph
  - Descriptive blobs of arrays of research products
  - Narrative CVs
- Web of Science DCI VS. OpenAIRE: comparison on datasets with Honami Numajiri and Takayuki Hayashi

Consiglio Nazionale delle Ricerche

ISTITUTO DI SCIENZA E TECNOLOGIE DELL'INFORMAZIONE "A. FAEDO"

infrascience

Thank you!

Andrea Mannocci

andrea.mannocci@isti.cnr.it

Images licence
https://www.istockphoto.com/legal/license-agreement

OpenAIRE
SCIENCE. SET FREE.