OpenAIRE Graph

An open resource for Open Science

Claudio Atzori, PhD
CNR-ISTI
Research evaluation and Open Science
a transition in the making

- Pre – Internet era: Science Citation Index (SCI), Current Contents, Journal Citation Reports
- WOS: monopoly based on sole database focused on the “core” science, shared algorithms, proprietary tools
- Scopus: sold for having more content, rising questions about the coverage
Research evaluation and Open Science
a transition in the making

- Google Scholar + H-index + publish or perish: bibliometrics @ home
- Google Scholar researcher profiles
- Question about openness
Research evaluation and Open Science
a transition in the making
OpenAIRE Graph

Putting research into context, making the connections
Foundations of a complete, open, and transparent scholarly publishing infrastructure

Reproducing and transparency require tracking of all products of research, their “context”, and their semantic links.

Monitoring and assessing quality, impact, and “open scienceness” of research must be a transparent, reproducible process, inclusive of research “context” and metrics.

Discovery of reproducible research must find new ways, moving beyond the “article-driven” search, taking into account metrics and intent.
Overview of the OpenAIRE Graph

“A collection of metadata describing objects in the research lifecycle and relationships among them”
**Overview**

The OpenAIRE Graph (formerly known as the OpenAIRE Research Graph) is one of the largest open scholarly record collections worldwide, key in fostering Open Science and establishing its practices in the daily research activity. Conceived as a public and transparent good, populated out of data sources trusted by scientists, the Graph aims at bringing discovery, monitoring, and assessment of science back in the hands of the scientific community.

Imagine a vast collection of research products all linked together, contextualized and openly available. For the past years OpenAIRE has been working to gather this valuable record. It is a massive collection of metadata and links between scientific products such as articles, datasets, software, other research products, entities like organisations, funders, funding sources, projects, communities, and data sources.

The OpenAIRE Graph aggregates millions of metadata records collected from trusted data sources, including:

- Open Access journals registered in DOAJ
- Crossref
- Unpaywall
- ORCID
- Microsoft Academic Graph
- Datacite

Open. Transparent. Interconnected.

OpenAIRE Graph is an open resource that aggregates a collection of research data properties (metadata, links) available within the OpenAIRE Open Science infrastructure for funders, organizations, researchers, research communities and publishers to interlink information by using a semantic graph database approach.
Data sources

- GitHub
- Applications Database
- Software Heritage
- elixir bio tools
- … and more tools and software sources
- frontiers
- F1000
- PLOS
- Springer
- … and more publishers
- CORDIS
- NSF
- DFG
- … and more European and international funders
- unpaywall
- Microsoft
- OpenCitations
- … and more research graphs
- OpenAIRE
- Graph
- re3data.org
- Crossref
- ROR
- … and more registries
- LA Referencia
- DOAJ
- DIRECTORY OF OPEN ACCESS JOURNALS
- arXiv.org
- zenodo
- HAL
- PubMed
- HAL
- archives-ouvertes.fr
- … and more thematic and institutional repositories
- RePEc
- EGI
- DARIAH-EU
- elixir
- … and more einfra and RI sources
OpenAIRE Graph data model

more @ graph.openaire.eu/docs/data-model
OpenAIRE Graph: the supply chain
Access to the OpenAIRE Graph

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.

Datasets @
graph.openaire.eu/docs/category/downloads

- OpenAIRE Graph
- OpenAIRE Funded products
- OpenAIRE COVID-19
- OpenAIRE communities
- DOIBoost
- OpenAIRE ScholeXplorer

zenodo.org/communities/openaire-research-graph
WHY MONITOR?

Know Thyself
- Resources
- Research Output
- Open Science Uptake
- Collaborations
- Visibility
- Impact

Understand Thyself
- Pathways
- Insights
- Opportunities

Position Thyself
- Decision making
- Reporting
- Story telling
Why OpenAIRE MONITOR?

- **Relevance for the community**
  - Co-develop indicators that make sense to all

- **It’s all about open science and open data**
  - Inclusiveness, transparency and replicability

- **Full coverage of Open Science**
  - Beyond publications: Research Data, Software, Other Research Products
  - Linked science

- **Fully embedded in EOSC infrastructure**
  - Starting from content providers to included metrics
How? Methodological Principles

- **Openness & transparency:** Assumptions are openly and clearly presented

- **Coverage & accuracy:** Based on data from multiple authoritative data sources with proven AI-driven mechanisms

- **Clarity & replicability:** Indicators documented, verified and used by the scholarly communication community

- **Readiness & timeliness:** Big data infrastructure operational workflows to warrant timely results
What? Dashboard on Demand

Institutions

Funders

Research Initiatives

Countries

Your OpenAIRE expert
Leonidas Pispiringas

Your OpenAIRE expert
Harry Dimitropoulos

Your OpenAIRE expert
Alessia Bardi

Your OpenAIRE expert
Ioanna Grypari

https://monitor.openaire.eu

https://osobservatory.openaire.eu
What?
Indicator themes
What? Indicators Breakdown

By fields of interest

- research product type
- domains (FoS)
- time
- countries
- data sources
- funders
- ...
What? Features

• **User-friendly** dashboards
  - interactive visualizations
  - exporting capabilities (download data & visualizations in different formats)
  - filtering
  - browsing via direct links to
  - editing of visualizations

• **Control** how you share
  - **Public** indicators for external stakeholders – *Showcasing*
  - **Restricted** indicators for team-members – *Internal Monitoring*
  - **Private** indicators for “work in progress” - *Reviewing*

• **Customize & validate** in one-on-one sessions with our experts
Si presentano qui alcuni dati tratti dall’Open science monitor di OpenAIRE. La dashboard dedicata all’ateneo (in versione beta) registra fra gli altri dati il modello di business delle pubblicazioni open access, le politiche di accesso e i download (anche per tipologia di file scaricato).
Can I trust the indicators?  
a.k.a can I trust the processes for their calculation?  
a.k.a can I trust the underlying data?

Challenges
• Data access / availability
• Data incompleteness, format inconsistency, default values
• Duplicate data
• Old / stale data
• Inconsistent record identification
• Coverage

The Graph as a trustable resource
• Vocabulary based cleaning
• Automation of the aggregation processes
• Provenance information
• Transparency of the processes
  • Documentation
• Repeatability

“Not everything that counts can be counted, and not everything that can be counted counts.”
Supporting projects

Several extensions in the context of ongoing & future projects

FAIRCORE4EOSC
- Intelligent NL and impact-based search, community recommendation profiles, support for and inference of RAIDs
- Adopting the RDA SKG common data representation for Scientific Knowledge Graphs

SciLake
- metadata enhancements & indicators derived from community specific use case pilots leveraging SKGs

GraspOS
- Innovate the research assessment process leveraging the OpenAIRE Graph contents

CraftOA
- Diamond institutional open access publishing & plugins for the OJS platform (broker and guidelines)

OSTrails
- Become part of a federation of cross-border and cross-discipline SKGs, interoperable with DMP and FAIR assessment platforms
Thank you

On behalf of the OpenAIRE Graph team

claudio.atzori@isti.cnr.it

info@openaire.eu