



# National Open Access Monitor, Ireland

Paving the road to 100% Open Access

## 1<sup>st</sup> RFO Training

Harry Dimitropoulos, Text Mining Expert & Technical Manager



# OUTLINE

- **Overview**
- **OpenAIRE Graph & Data Quality**
- **Linking functionality**
- **Funders in OpenAIRE**
- **Next Steps**

# THE MONITOR – 5 DASHBOARDS

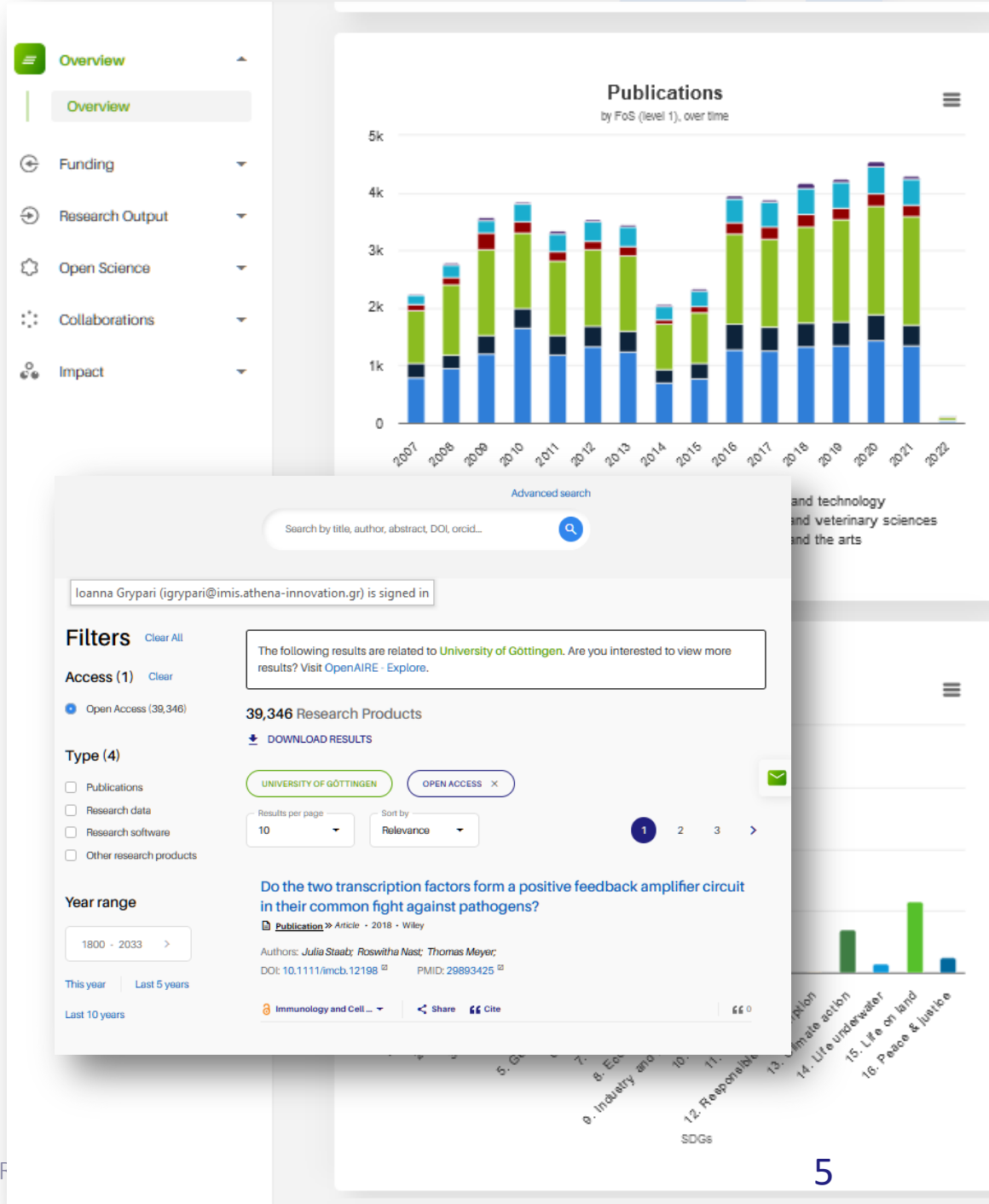


## OBJECTIVES

### RFOs

- **Why**
  - Benchmark against peers
  - Assess OA compliance
  - Monitor research impact
  - Shape strategic agendas

- Comprehensive insights into each Irish RFO's activities and OS performance
- **Indicators View**
  - Highlighting **OA uptake** indicators, enables **cross RFO benchmarking**
  - Offers detailed **breakdowns** and **filtering**
- **Research Outputs View**
  - Facilitate **browsing**, standard and advanced **searching**, and **filtering**
- **RFO Managers**
  - Access to the **Sandbox** to verify data quality



The icon consists of three interconnected circles forming a triangular shape, with a small plus sign inside the top circle.

# OpenAIRE Graph & Data Quality

# THE OPENAIRE GRAPH

- **Scientific Knowledge Graph**

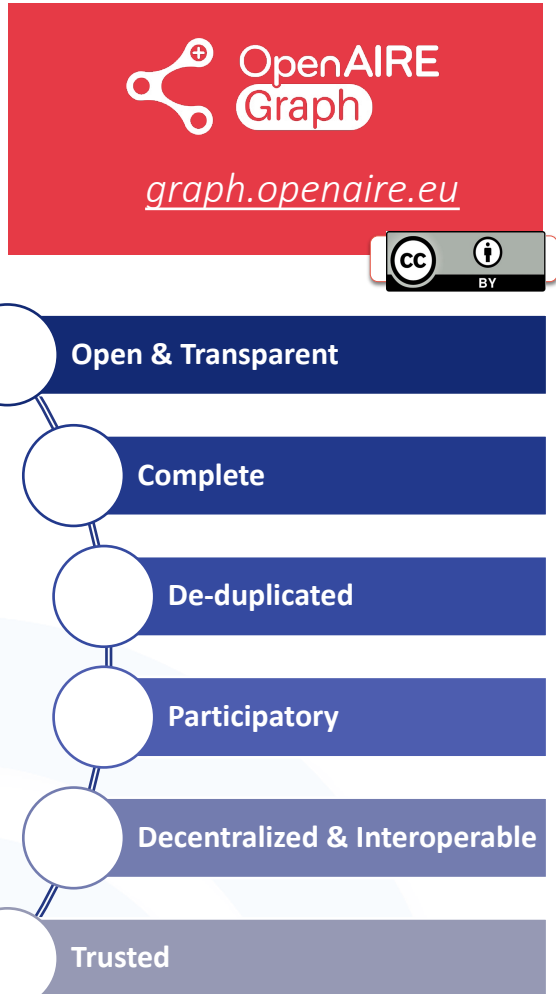
- *A collection of metadata describing objects in the research lifecycle and relationships among them*
- Timely and comprehensive coverage of research outputs
- 129K data sources, 3M projects, and 235M research outputs (publications, software, datasets, other research products)
- Near monthly updates

- **Precision, depth & processing**

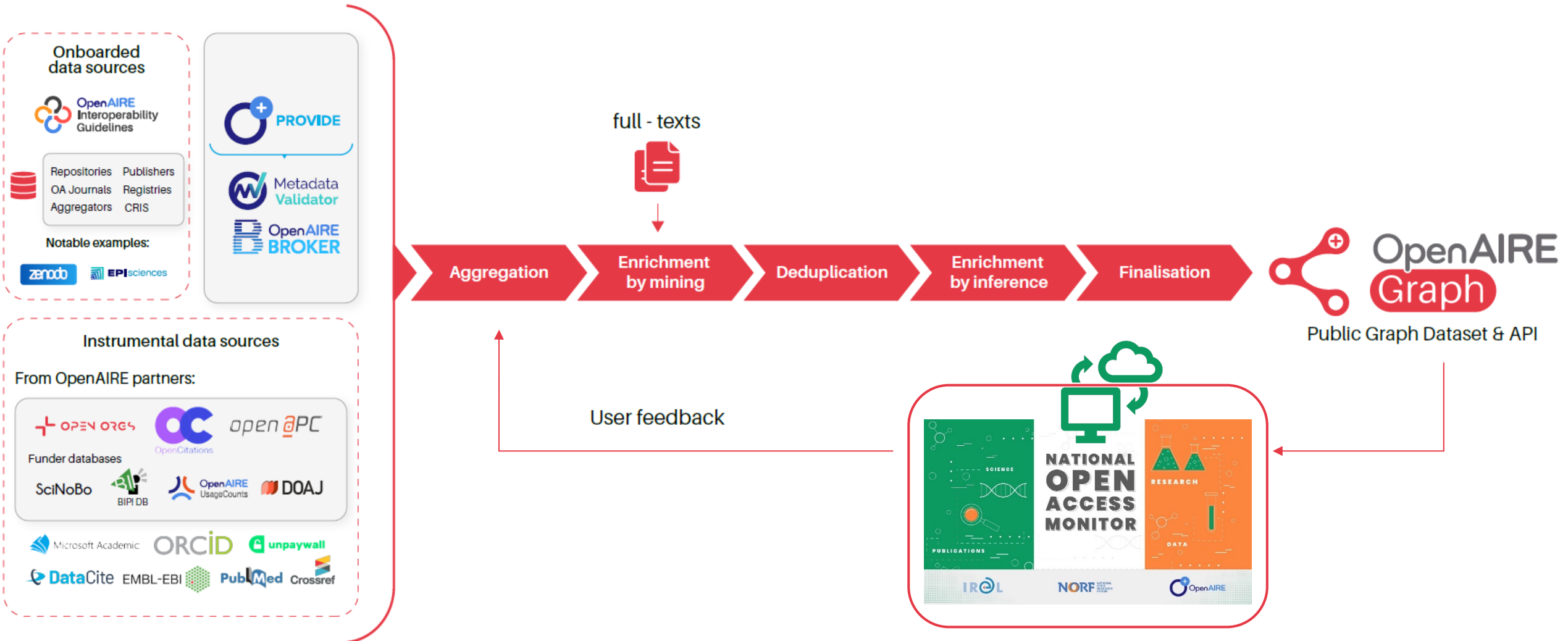
- Rigorous cleaning, deduplication, and enrichment for optimal accuracy
- Enriched metadata links: Research results to projects, author affiliations, and classifications (FoS, SDG)
- Techniques for duplicate identification in addition to OpenAIRE's curation tools

- **Robustness & openness**

- Maintenance, load balancing, backups, overseen by the OpenAIRE technology centre
- Open data & transparent methodologies

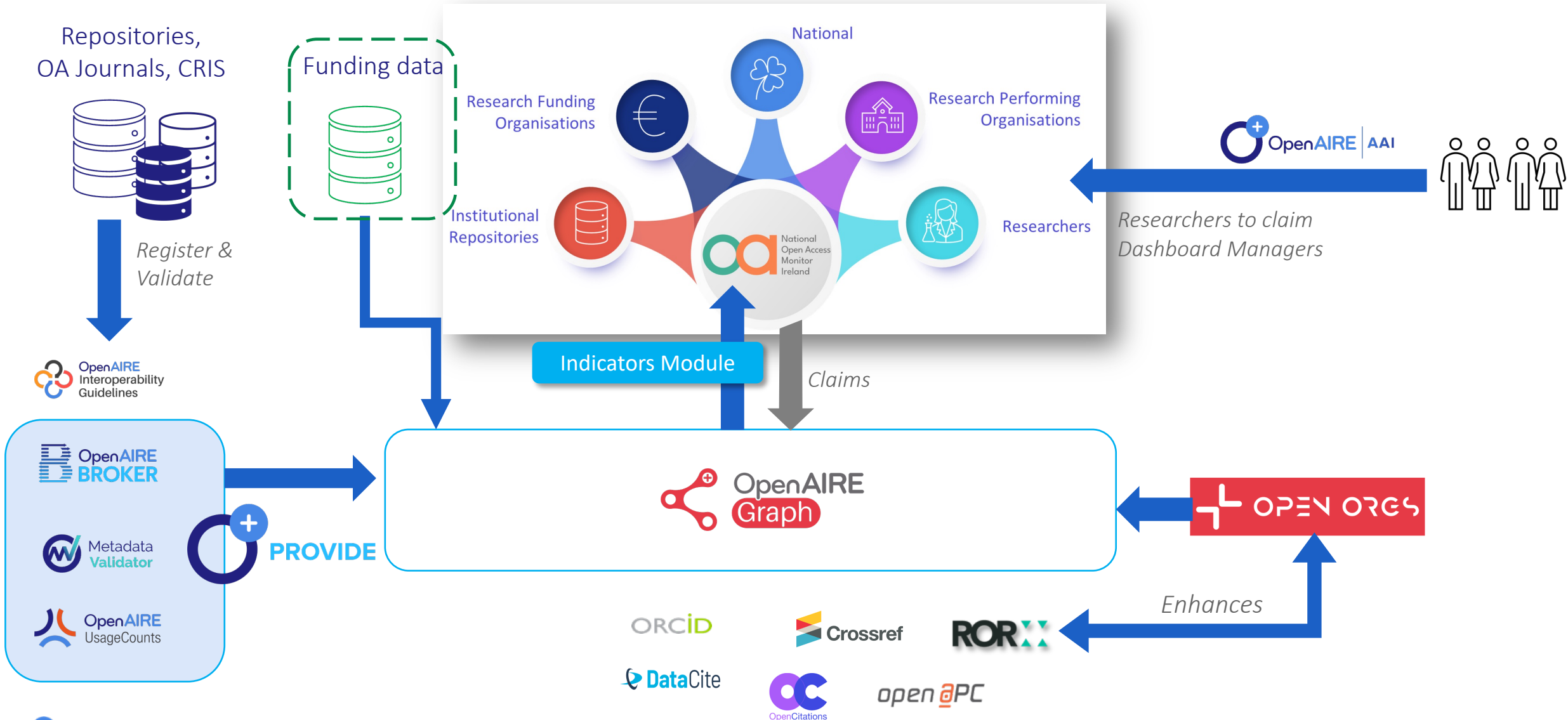


# GRAPH PIPELINE

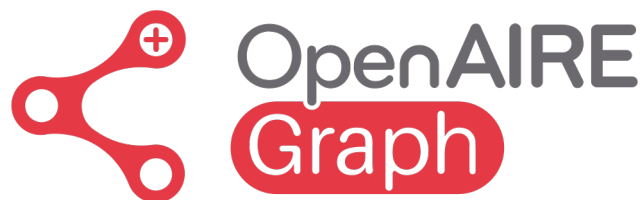
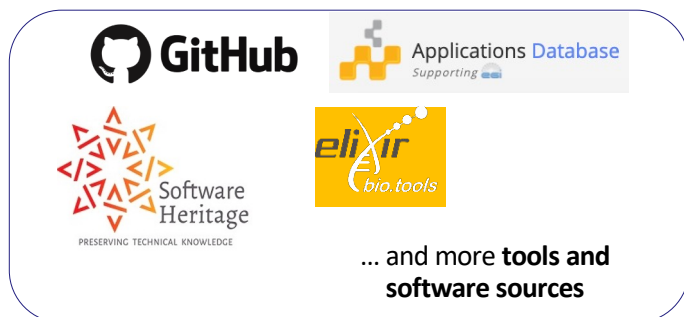




# THE ARCHITECTURE



# COVERAGE: SOURCES CONTRIBUTING TO THE GRAPH



## DATA QUALITY

- **Deduplication:** OpenAIRE Graph merges duplicate records of the same scholarly work.
- **Enrichment through TDM:** affiliation, citation, and concept extraction with document classification (FoS, SDG) and similarity assessment
- **Cleaning:** Independent, continuous aggregation processes utilize vocabularies to harmonize diverse data source records, ensuring consistent and accurate bibliographic records.
- **Disambiguation:** Additional disambiguation of Journals, Publishers and Licenses
- More @ <https://graph.openaire.eu/docs/graph-production-workflow/>

# DATA QUALITY

## Additional Data Quality Enhancements

- **All:** via linking functionality
- **RFOs:** Provide Funding data for custom text mining

# LINKING

# LINKING FUNCTIONALITY

- **Log in** (EDUGAIN, ORCID, Google, GitHub, LinkedIn, OpenAIRE) to access the **LINKING** functionality
  - Enables users to associate **research outputs** with **projects, communities, or other outputs**, enhancing data representation in OpenAIRE Graph and the Monitor
  - Links show up in next OpenAIRE Graph update
  - A **log** will be **publicly visible** with all added links, with **permission** asked during final linking step

# LINKING FUNCTIONALITY

FIND SOURCES LINK SOURCES TO ENTITIES SUMMARIZE AND FINISH



pispiringas ✕ 🔍

Openaire (29) Crossref (2) Datacite (29) Orcid (17)

YEAR RANGE ▾ TYPE ▾ FUNDER ▾ ACCESS ▾ DOCUMENT TYPE ▾

## Introduction to the ORCID Greek Consortium

Publication • ORCID • 2019

Pispiringas, Leonidas

## OAWeek: OpenAIRE MONITOR: Institutional Dashbo

Publication • Zenodo • 2022

Grypari, Ioanna; Pispiringas, Leonidas

## OAWeek: OpenAIRE MONITOR: Institutional Dashbo

Publication • Zenodo

Grypari, Ioanna; Pispiringas, Leonidas

## Research data: accessible infrastructures and innova

Publication • Zenodo • 2020

Pispiringas, Leonidas; Chatzigeorgakidis, Georgios; Terrovitis, Manolis



Sources to link (1) Link to (1)

[Upload a DOI's CSV file](#) ⓘ 🗑️ Remove all

OAWeek: OpenAIRE MONITOR: Institutional Dashboard

Grypari, Ioanna; Pispiringas, Leonidas ✕

FIND SOURCES LINK SOURCES TO ENTITIES SUMMARIZE AND FINISH



SOURCES (1) ✎ Edit sources

OAWeek: OpenAIRE MONITOR: Institutional Dashboard

Publication • Zenodo

Grypari, Ioanna; Pispiringas, Leonidas ✕

LINK TO (1) ✎ Edit entities

OpenAIRE Nexus

Project • 2021-2023

Funder: EC Project Code: 101017452 ✕

CONFIRM LINKING →

# Funders in OpenAIRE



# FUNDERS IN OPENAIRE

OpenAIRE | EXPLORE

Search Deposit Link Data sources Funders Sign in

Funders

## Be an integral part of the open R&I ecosystem

Welcome to the Funders' page on OpenAIRE Explore. Discover key details about each funder, their commitment to open access, and the impactful research they enable. Our aim? To foster transparency, inspire collaboration.

[JOIN US](#)

**3M+** GRANTS    **26** FUNDERS    **4M+** FUNDED RESEARCH OUTPUTS

Show All funders    Sort by Alphabetically Asc.    Results per page 10

Search funder

<p>85% Open Access</p> <p>ACADEMY OF FINLAND</p> <p>AKA Academy of Finland</p> <p>MONITOR DASHBOARD</p> <p>PROJECTS (28,416) →</p> <p>RESEARCH PRODUCTS (138, 164) →</p>	<p>95% Open Access</p> <p>asap</p> <p>Aligning Science Across Parkinson's</p> <p>PROJECTS (35) →</p> <p>RESEARCH PRODUCTS (64) →</p>	<p>69% Open Access</p> <p>Australian Government Australian Research Council</p> <p>PROJECTS (29,472) →</p> <p>RESEARCH PRODUCTS (97,457) →</p>
<p>93% Open Access</p> <p>FWF</p> <p>Der Wissenschaftsfonds.</p> <p>FWF Austrian Science Fund (FWF)</p> <p>MONITOR DASHBOARD</p> <p>PROJECTS (17,731) →</p> <p>RESEARCH PRODUCTS (134,221) →</p>	<p>80% Open Access</p> <p>CIHR IRSC</p> <p>Canadian Institutes of Health Research</p> <p>RESEARCH PRODUCTS (88,032) →</p>	<p>87% Open Access</p> <p>chist-era</p> <p>CHISTERA CHIST-ERA</p> <p>MONITOR DASHBOARD</p> <p>PROJECTS (108) →</p> <p>RESEARCH PRODUCTS (1,204) →</p>
<p>68% Open Access</p> <p>HRZZ</p> <p>HRZZ Croatian Science Foundation (CSF)</p> <p>MONITOR DASHBOARD</p> <p>PROJECTS (1,832) →</p> <p>RESEARCH PRODUCTS (9,898) →</p>	<p>84% Open Access</p> <p>EC</p> <p>European Commission</p> <p>MONITOR DASHBOARD</p> <p>PROJECTS (70,206) →</p> <p>RESEARCH PRODUCTS (1,717,230) →</p>	<p>100% Open Access</p> <p>European Environment Agency</p> <p>PROJECTS (7) →</p> <p>RESEARCH PRODUCTS (23) →</p>

## FUNDERS IN OPENAIRE

- Directly joining OpenAIRE is the recommended avenue for representation in the OpenAIRE Graph
- Benefits
  - Ensures a comprehensive presence, with project level granularity
  - Guarantees the precision of the data incorporated
- Process
  - RFO provides Funding Data (list of research projects with metadata)
  - A tailored text mining algorithm is created for data extraction
  - Meticulous curation of project-publication links to ensure accuracy
  - <https://www.openaire.eu/funders-how-to-join-guide>

## IRISH FUNDERS IN OPENAIRE GRAPH

- Alternative avenue for representation via Open Funder Registry (OFR)
  - <https://www.crossref.org/services/funder-registry/>
  - Allows funders to be associated with publications via valid DOIs in the Graph
    - No projects currently registered with Crossref, we have only funder-publications links
    - Text mining module cannot be built without list of projects
    - We are working on creating funder-publication links for IRC, but with unidentified projects.
  - Does not offer the granularity of the direct integration
  - Quality not guaranteed, no curation
- 157 Irish RFOs integrated via OFR
  - 141 with publications

## SCIENCE FOUNDATION IRELAND

- SFI joined OpenAIRE in 2015 via the direct route
- Project data from <https://www.sfi.ie/about-us/governance/open-data/>
- 6,391 projects
- Currently 16,750 publications in <https://explore.openaire.eu/>
  - Increasing to 18,618 in the next few days

## IRISH RESEARCH COUNCIL

- Started looking into IRC, without their project data
- Created a text mining module to identify publications funded by IRC at the funder level (funder-publication links, not project-publication links)
- Found 5K publications
- Investigated the "text snippet" statements provided by the mining
  - E.g., “... was funded by the *Irish Research Council*, (*GOIPG/2013/1110*)”.
  - Identified IRC **projects** in statements, they are well structured and easy to mine
- Given the full project list, IRC mining would return near perfect accuracy

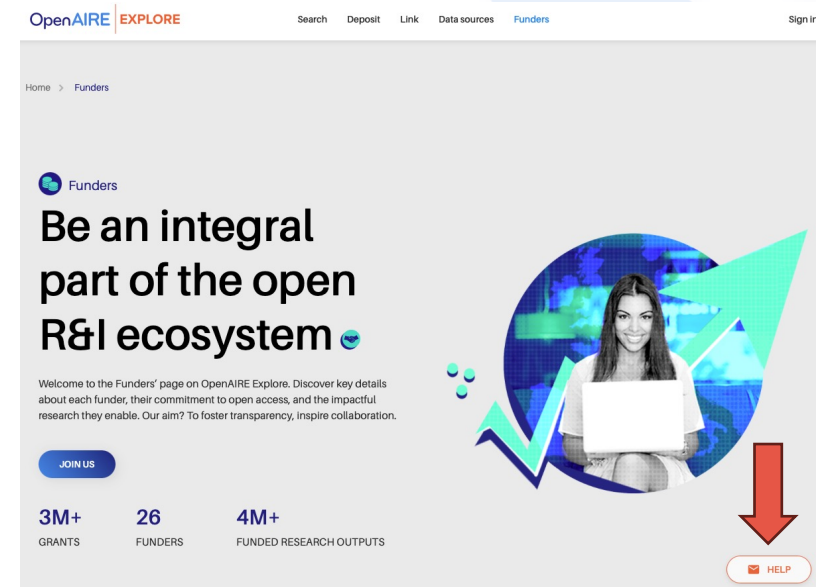
# HOW TO JOIN OPENAIRE

Join us by either:

- Filling in the simple HELP form at the bottom right of <https://explore.openaire.eu/funders>
- Email [info@openaire.eu](mailto:info@openaire.eu) or [mining@openaire.eu](mailto:mining@openaire.eu)
- Register as an RFO manager

We will then get back to you for:

- Providing your project dataset (metadata) and future updates as needed
  - By file (standard spreadsheet), via an API, your Website, or by Registering Projects with OFR
- Give examples of suggested project acknowledgement statements to facilitate the text and data mining algorithms



# Next Steps

## NEXT STEPS

- Release of **pilot** MONITOR on January 10th, but able to preview data in a few days at [explore.openaire.eu](https://explore.openaire.eu)
- Use linking functionality
- Integrate your data with OpenAIRE
  - Join OpenAIRE and provide project data – the sooner the better
  - Consult OpenAIRE guidelines
    - for specifying funding information and sharing essential public data with relevant Repositories/CRIS systems (<https://guidelines.openaire.eu/>)



Thank you!  
*[harryd@athenarc.gr](mailto:harryd@athenarc.gr)*