National Open Access Monitor, Ireland
Inception Report

OpenAIRE AMKE

Ioanna Grypari, Leonidas Pispiringas, Natalia Manola

Contents

Abbreviations

APC	Article Processing Charge
EC	European Commission
EOSC	European Open Science Cloud
FAIR	Findable, Accessible, Interoperable, Reusable
FAQs	Frequently Asked Questions
FP	Framework Programme
FoS	Fields of Science
Graph	OpenAIRE Graph
IReL	The consortium of Irish research libraries
IT	Information Technology
Monitor	National Open Access Monitor, Ireland
MU	Maynooth University
NORF	National Open Research Forum
OA	Open Access
OS	Open Science
ORCID	Open Researcher and Contributor ID
PID	Persistent Identifier
RFO	Research Funding Organisation
RPO	Research Performing Organisation
SDGs	Sustainable Development Goals
UI	User Interface

Glossary

TERM	DEFINITION
Article Processing Charges (APC)	The fee charged by publishers in order to publish a research publication in an open access journal. These charges are meant to cover the costs of publication and ensure the work is freely accessible to all.
Research Outputs/Products	The four different types of research products in the OpenAIRE Graph: Publications, Research data, Research software, Other research products.
JOURNAL BUSINESS MODELS	
Fully OA	A journal that publishes only in OA
Subscription	A journal that charges for access to its articles.
Hybrid	A subscription journal where some of its articles are open access.
Transformative	"A Transformative Journal is a subscription/hybrid journal that is actively committed to transitioning to a fully Open Access journal. In addition, a Transformative Journal must gradually increase the share of Open Access content; and offset subscription income from payments for publishing services (to avoid double payments).[1]
JOURNAL APC BUSINESS MODELS	
Diamond OA	A fully OA journal that does not charge article processing charges (APCs).
OA TYPES/COLOURS[2]	
Green OA	Green articles are published in toll-access journals, but archived in an OA archive, or "repository". These repositories may be discipline-specific (like ArXiv) or institutional repositories operated by universities or other institutions. Green articles may be published versions or preprints, and can have any license or no license.
Bronze OA	Bronze articles are free to read on the publisher's website, without a license that grants any other rights. There may be a delay between publication and availability to read, and often articles can be removed unilaterally by the publisher.

---

[1]https://www.coalition-s.org/transformative-journals-faq/
[2]As requested by the project, we have used the Unpaywall definitions of OA colours.

Hybrid OA

Hybrid articles are free to read at the time of publication, with an open license. These are usually published in exchange for an article processing charge, or APC.

Gold OA

Gold articles have all the same characteristics as Hybrid articles, but are published in all-Open Access journals, which are in turn called "Gold journals", or just "OA journals".

Executive Summary

This Inception Report provides a comprehensive overview of the National Open Access Monitor, a dynamic platform designed to monitor and advance open access initiatives in Ireland. The report outlines the fundamental components, methodologies, and strategies that underpin the Monitor's functionality and data quality. It also sheds light on the coordination, engagement, and training aspects integral to the project's success.

1. Introduction

After a comprehensive open tender process overseen by IReL in conjunction with Education Procurement Services, OpenAIRE has been appointed to spearhead the development of Ireland's National Open Access Monitor (referred to as "the Monitor"). This seminal initiative aims to systematically analyse and track Ireland's progress towards the target of 100% Open Access (OA). In embarking on this venture, Ireland positions itself alongside its European peers, furthering the collective goal of robust Open Science (OS) monitoring. This inception report details each facet of our upcoming deliverables, providing a clear roadmap for the path ahead.

The report is structured around two central pillars:

1. **The National Open Access Monitor** – This section offers an in-depth exploration of the Monitor, detailing its dashboards and associated features. It elaborates on the specific data and indicators to be incorporated, the approach for code and data deposition, and the comprehensive documentation plan. Furthermore, it emphasizes the supportive engagement and training programmes and outlines the anticipated collaboration and contributions from IReL in its delivery.

2. **The National Open Access Report** (the "Report" from now on) – This section introduces the forthcoming report, a pivotal component accompanying the Monitor. This report serves as a holistic evaluation of the open access landscape in Ireland. It meticulously assesses the precision and inclusiveness of existing publication records, identifies enduring monitoring complexities, and delivers strategic recommendations for the future.

In collaboration with IReL and NORF, OpenAIRE is committed to delivering a monitoring platform that will not only measure, but also guide the progression of OA in Ireland. Our aim is to provide stakeholders with accurate data, insights, and strategies, ensuring an inclusive, structured and informed approach towards achieving a sustainable OA future.

The subsequent sections of the report delve deeper into the aforementioned pillars. The first part focuses on the Monitor, specifically addressing: its dashboards and functionalities (Section 2), the selected data and indicators (Section 3), the data and code deposition plan (Section 4) and the associated project coordination, engagement and training programmes (Section 5). Subsequently, Section 6 expands upon the National Open Access Report and Section 7 presents the project's Gantt chart.

2. Dashboards & Functionalities

This section outlines the platform's user interface (UI), detailing the distinct public dashboards tailored for various user personas: National, RPO (Research Performing Organisation), RFO (Research Funding Organisation), Repository, and Researcher. Each dashboard provides data and functionalities specifically crafted for its designated audience.

The Monitor's National dashboard, serving as the landing page, provides a comprehensive national overview. Delving deeper, the platform introduces specialised dashboards tailored to distinct stakeholder needs. The RPO and RFO dashboards cater to their respective groups, offering organisational indicators and a detailed view of research outputs. The Researcher dashboard is specifically designed for individuals, allowing users to search for specific researchers or browse their own record and Open Access uptake. Simultaneously, the Repository dashboard facilitates the tracking of records of an institutional repository and their associated usage statistics.

Across all these dashboards, users benefit from robust search and filtering functionalities and the capability to effortlessly export data and visualisations in widely accepted formats. Benchmarking metrics and visual tools are integrated to facilitate direct comparisons between organisations and countries. Supporting the Monitor's overarching commitment to openness, transparency, and replicability, detailed documentation pages are included, ensuring stakeholders have all the tools and knowledge to achieve their monitoring objectives.

The data and indicators showcased in the Dashboards are derived from the OpenAIRE Graph (referred to as 'the Graph', https://graph.openaire.eu/), a scientific knowledge database that integrates open data from over 126K data sources, encompassing 3M projects and hundreds of millions of publications, research data, software and other research products. Within the Graph's aggregation pipeline, every record undergoes rigorous cleaning, deduplication, enrichment, and interlinking to ensure accuracy and relevance. The "Data and Indicators" section includes additional details on the data shown.

For clarity in navigating and understanding the functionalities within the National Open Access Monitor, Ireland, we introduce two distinct sections within each dashboard: the **Indicators View** and the **Research Outputs View**.

- The **Indicators View** provides stakeholder-level indicators broken down by aspects of interest, facilitating a deeper understanding of overarching patterns and research trends. With clear visualizations, it allows users to grasp large datasets with clarity, providing a comprehensive overview of the research landscape. This view is essential for strategic decision-making and gauging overall progress.[3]

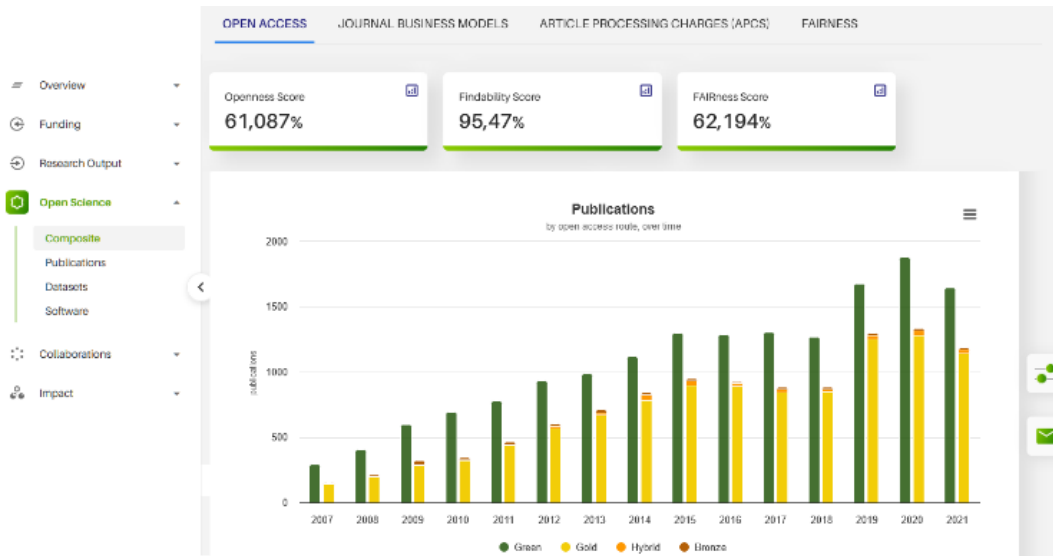[3] Examples of the Indicators View can be seen on the dashboards of https://monitor.openaire.eu/

Figure 1: Indicators View Sample from https://monitor.openaire.eu/

- The **Research Outputs View** offers the ability to zoom into specific research outputs, providing access to valuable information such as access rights, open access types, and other essential metadata. It offers a granular view that allows users to instantly verify the quality and accuracy of individual research data. The Research Outputs View is instrumental for stakeholders seeking in-depth insights.[4]
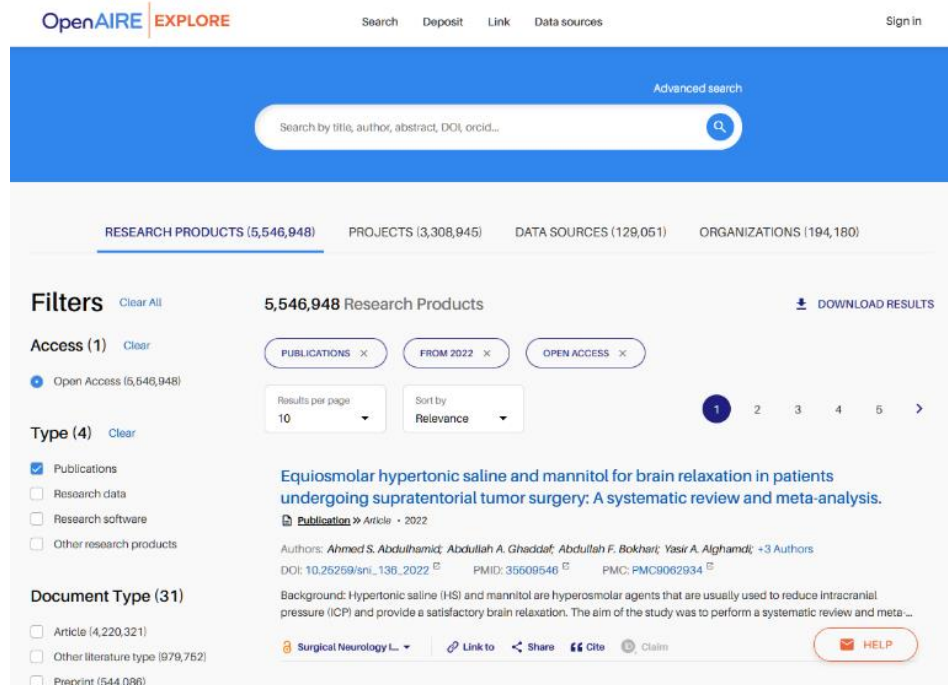

Figure 2: Indicators View Sample from https://explore.openaire.eu/

---

[4] Examples of the Research Outputs View can be seen on the dashboards of https://explore.openaire.eu/

Together, these sections provide a comprehensive and intuitive user experience, ensuring that users of the Monitor can both oversee the broader landscape and delve into specific research outputs with ease.

National Dashboard

The National Dashboard, which also functions as the landing page of the Monitor presents an overview of Ireland's research landscape and the country's OA uptake. Within the National Dashboard, two distinct views empower stakeholders with valuable insights.

Indicators View

- Showcases key OA uptake indicators to gauge Ireland's research openness.
- Enables cross-country benchmarking for informed comparisons.
- Offers detailed breakdowns and flexible filtering options, including by Research Performing Organizations (RPOs), Research Funding Organizations (RFOs), years, and more.

Research Outputs View

- Facilitates user-friendly browsing, standard and advanced searching, and effective filtering.
- Empowers users to explore specific research outputs with ease, supporting comprehensive research discovery.

Moreover, National Dashboard Managers from IReL gain access to a dedicated resource: the **Sandbox**. This beta interface provides a preliminary look at monitoring data and functionalities, offering a sneak peek into the upcoming developments. The data within the Sandbox is sourced from the OpenAIRE Graph's pre-release version (BETA), typically available two weeks before the main release, serving as the initial checkpoint for data changes and enhancements.

RPO & RFO Dashboards

The RPO and RFO dashboards offer comprehensive insights into each Irish RPO's and RFO's activities and performance within the realm of Open Science, and mirror in structure and features the National Dashboard.
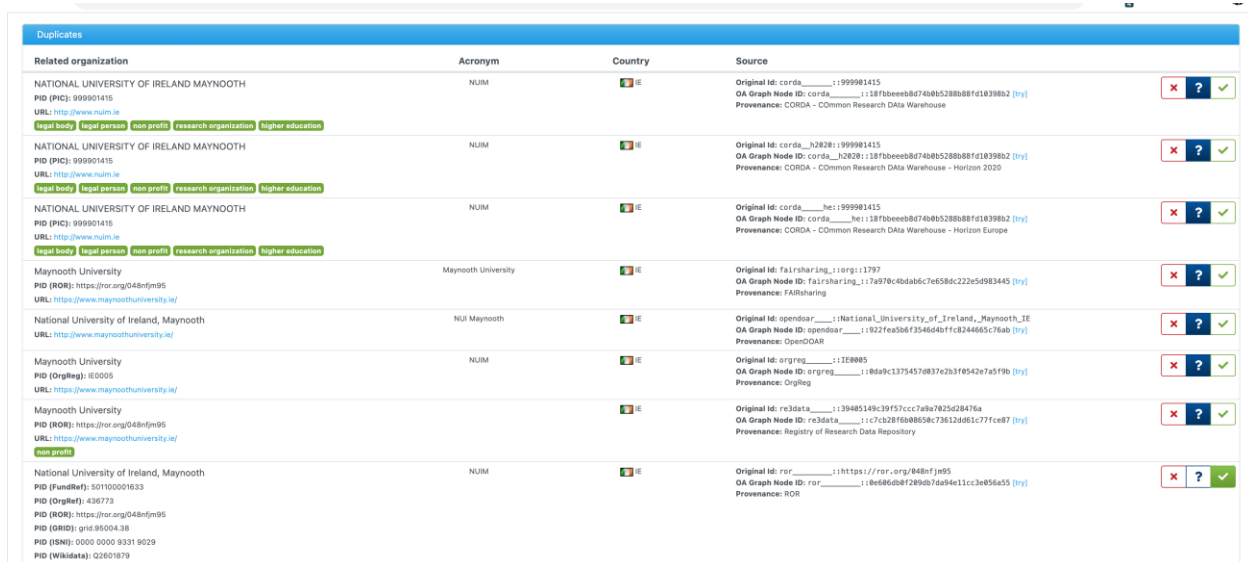
Indicators View: This view highlights essential open access (OA) uptake indicators, enabling cross RPO/RFO benchmarking. It offers detailed breakdowns and various filtering options for a thorough analysis.

Research Outputs View: The Research Outputs View facilitates browsing, standard and advanced searching, and filtering, enhancing the user experience.

RPO/RFO Managers

IReL will define the procedure for how RPO/RFO managers will be assigned. Upon invitation to their respective RPO dashboards, these managers, by accepting their role, will consent to the Terms & Conditions associated with their position as stipulated by IReL. This includes agreeing to the (automatic) creation of a **public** log documenting any changes they make and adherence to standard OpenAIRE terms and conditions[5]. Additionally, these managers will have access to essential tools and features.

- **OpenORGS:** A user friendly tool for efficiently removing duplicate organisation names of each RPO.

Figure 3: Screenshot from OpenORGS

- **Sandbox**: The beta interface of the Monitor giving an early view of monitoring data from the pre-release version of the Graph (BETA). It will act as the first checkpoint for data modifications and improvements, such as new repository integration in OpenAIRE or organisation name deduplication via OpenORGS.

By providing these capabilities, we aim to bolster engagement and uplift the data quality across the dashboard, catering to varied user needs.

---

[5]All logs will be shared publicly on the Monitor.

Researcher Dashboard

ORCID Integration

The Researcher Dashboard's core strength lies in its tight integration with ORCID. Every researcher's profile is anchored to their ORCID iD, which can be used for logging in as well, ensuring that personal and professional details displayed are authentic and consistent.

- **Enhancements in Monitor:** The Monitor provides the ability to augment a researcher's ORCID record with any data found in the OpenAIRE Graph, using the **'Claim'** function. Consistent with Monitor's transparency, a public log of all modifications will be maintained. Users must explicitly consent to the creation of this log before confirming the edits. Enhancements will be shown on the researcher's dashboard after the upcoming Graph update.
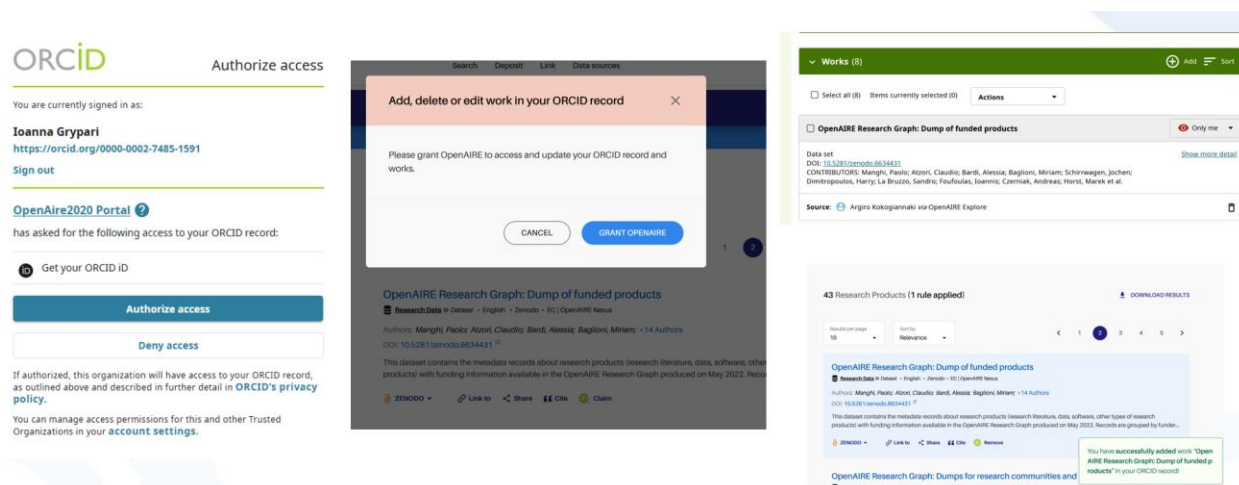
Figure 4: Screenshots of OpenAIRE's integration with ORCID and the Claim functionality.

- **Personal Data Updates:** For any changes to personal data[6], researchers will have to update their ORCID profile directly. These changes will reflect on the Monitor dashboard after the next OpenAIRE Graph update, which takes place monthly.

The Researcher Dashboard will include an **Indicators view** with a series of simple Open Science performance metrics for the researcher, followed by the **Research Outputs View** with a comprehensive list of their research works. Users will have the flexibility to browse, filter, export, and download their outputs.

---

[6] Name and email (only the name shows on the OpenAIRE Researcher profile).

Logged-In Section

Researchers, when logged into the Monitor, can:

- **Connect** their ORCID account directly, offering a seamless integration experience. This step is only taken once when setting up their profile.
- Grant OpenAIRE the permission to **enhance** their existing ORCID profiles.
- **Discover** research outputs and **claim** them as their own as described above.

Repository Dashboard

OpenAIRE allows the registration and aggregation of any **institutional** data sources (Repositories, Data Repositories, CRIS, Open Access Journals). In the case of an institution maintaining both an institutional repository and a CRIS with different research products then they will have separate Repository Dashboards for each data source.

To empower institutional repository managers, the Repository Dashboard has been designed as an essential tool to oversee, manage, and enhance their institutional repository's/CRIS performance in alignment with Open Science standards.

Optional Add-on: For an institutional repository/CRIS to *fully* benefit from the Repository Dashboard, two steps can be taken:

1. **Integration with OpenAIRE**: The institutional repository/CRIS should be registered with and integrated into OpenAIRE. This connection is essential to harness the full suite of features and insights that the dashboard presents. Metadata records from these registered repositories should align with the OpenAIRE Guidelines (https://guidelines.openaire.eu/) and be accessible for aggregation via the OAI-PMH protocol. Adhering to the OpenAIRE Guidelines has the following added benefits:
   - Achieve greater **interoperability**: Present richer, more contextualized content through links and relationships with varied research outcomes, such as PIDs. Enjoy flexibility from diverse and enhanced vocabularies and deeper integration with the R&I ecosystem, aligning with Open Science mandates, standards, and established metadata schemas and namespace abbreviations.
   - Ensure they are aligning with **FAIR** principles.
   - Establish their initial pathway to **EOSC**, facilitating onboarding onto the integrated EOSC Portal Catalogue and Marketplace platform.

2. **Activation of OpenAIRE UsageCounts Service**: To leverage detailed usage metrics repository managers must personally activate the OpenAIRE UsageCounts service. By having the service

activated, managers receive a granulated, de-duplicated set of metadata views and downloads. This data, in adherence to the COUNTER standards, seamlessly integrates into the Monitor, offering item-based statistics. Moreover, the service provides diverse COUNTER compliant report formats, deepening the scope of usage analysis.

The Repository Dashboard, provides a panoramic view of the repository's research landscape, this includes:

Indicators View:

- **Metrics & Indicators**: This segment showcases a suite of metrics surrounding the research output of that data source. Moreover, *when step 2 (as detailed above) is completed*, the dashboard will feature indicators derived from usage reports. These indicators emphasize the repository's impact, illustrated through statistics on views and downloads.

Research Outputs View:

- **Browsing Capabilities**: Users can seamlessly navigate through the repository's affiliated research outputs. They can use advanced filters, offering a more in-depth look, and have exporting options consistent with the features in the "Common features" section.

OpenAIRE will provide dashboards from repositories that are not currently registered with OpenAIRE, as long as these repositories maintain an active OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) protocol. Subsequent to the integration process, a comprehensive examination will be conducted to assess the compliance with the OpenAIRE Guidelines. Moreover, this examination shall identify any potential issues arising from the lack of compliance with the OpenAIRE guidelines and ascertain the impact such non-compliance may have on the Monitor Data. These potential impacts encompass the absence of persistent identifiers within records' metadata, the omission of ORCID iDs, insufficient access rights and licensing information, among others.

Logged-In Section for Repository Dashboard:

If registered with OpenAIRE PROVIDE (step 1 above), when logging in, Repository Dashboard managers will be redirected to their dashboard in the *OpenAIRE PROVIDE Service* (https://provide.openaire.eu), the primary gateway for repository managers into the OpenAIRE Graph. This step is crucial, as the PROVIDE dashboard is part of the initial phase of the OpenAIRE Graph production workflow, ensuring the data integrity for the Monitor.[7]

The OpenAIRE PROVIDE dashboard, extends an enriched array of features, allowing managers to

---

[7] https://graph.openaire.eu/docs/graph-production-workflow/

- **Register, validate, and enrich their records:** Through the OpenAIRE PROVIDE dashboard, Repository managers can validate (incl. FAIRness) their records against the OpenAIRE Guidelines and enrich their records with potential richer metadata that come from other data sources that OpenAIRE harvests from.
- **View details on the aggregation history of their repository.**

Common Features

Filtering the Data

Indicators View

The dashboard employs a two-fold approach to data presentation and exploration, designed to meet a diverse range of user requirements.

Breakdowns: The dashboard is configured to showcase detailed breakdowns using key categorizations. With central emphasis on aspects like RPO, RFO, year, data source type and so on, users can access a comprehensive view of the data.

Filters: To complement the breakdowns, the dashboard will integrate additional filters allowing the user to further refine the data. The filters that will be incorporated in the first Monitor release will include at a minimum year range and publicly-funded. While obtaining a list of public RPOs and RFOs from IReL is not necessary, doing so could substantially improve data curation and provide users with a more tailored overview.

Employing a combination of detailed breakdowns with select additional filters, the dashboard is designed for optimal drill-down functionality. This ensures that users can probe into specific data points while the interface remains user-friendly and uncluttered.
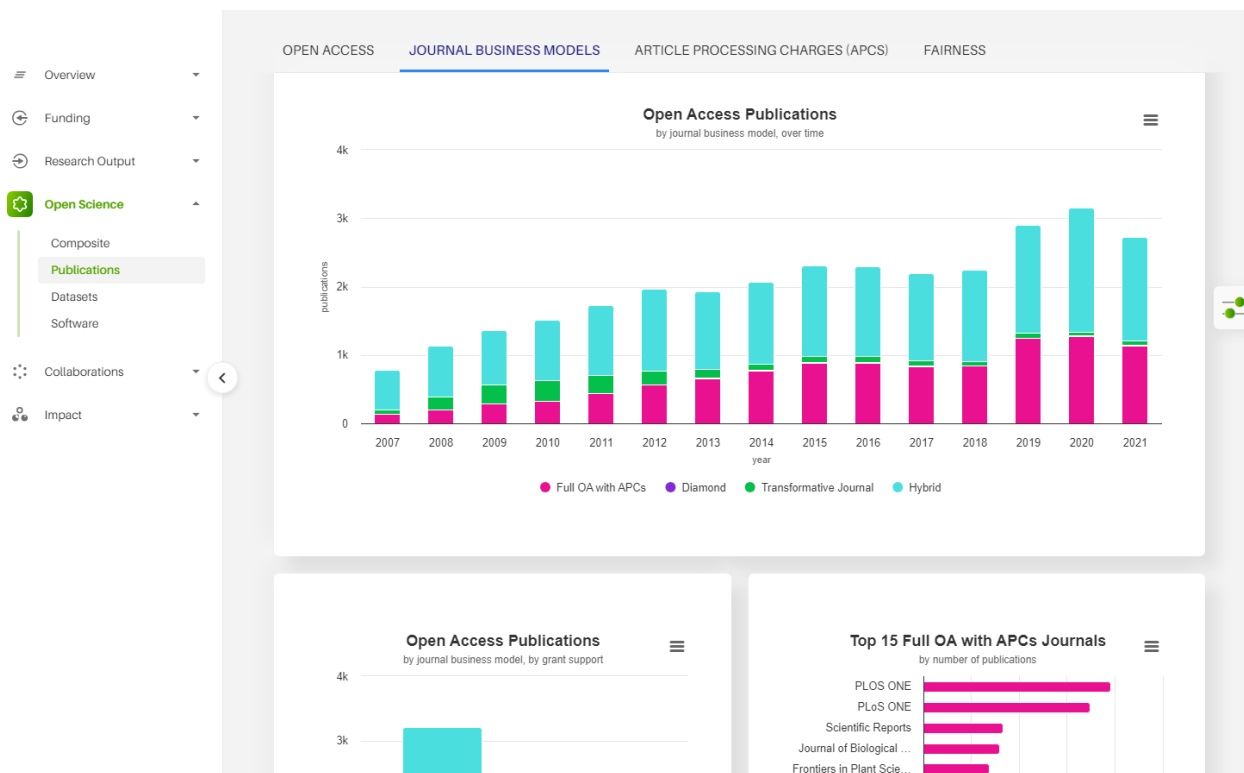
Figure 5: Indicator categorisation and breakdown with additional filters (button on the right)

Research Outputs View

The **Research Outputs View** sections of the portal will present enriched metadata records, detailing both access rights and the open access types/colours (Green, Gold, Hybrid, Bronze, Diamond OA) for each research output.

The **advanced search** feature provides a refined approach to navigation with the following *search filters:*

- RFO, grant award ID, project acronym
- RPO, RPO of the corresponding author[8]
- year of publication (range)
- Fields of Science (FoS)
- publisher, journal
- author name, ORCID iD

Users can combine these filters in any arrangement to pinpoint specific results.

---

[8] Examination of the availability of data on the corresponding author will be detailed in the Report.
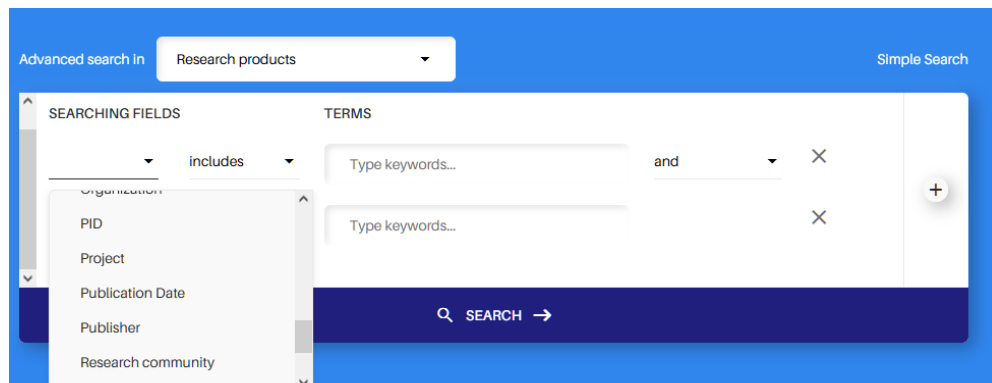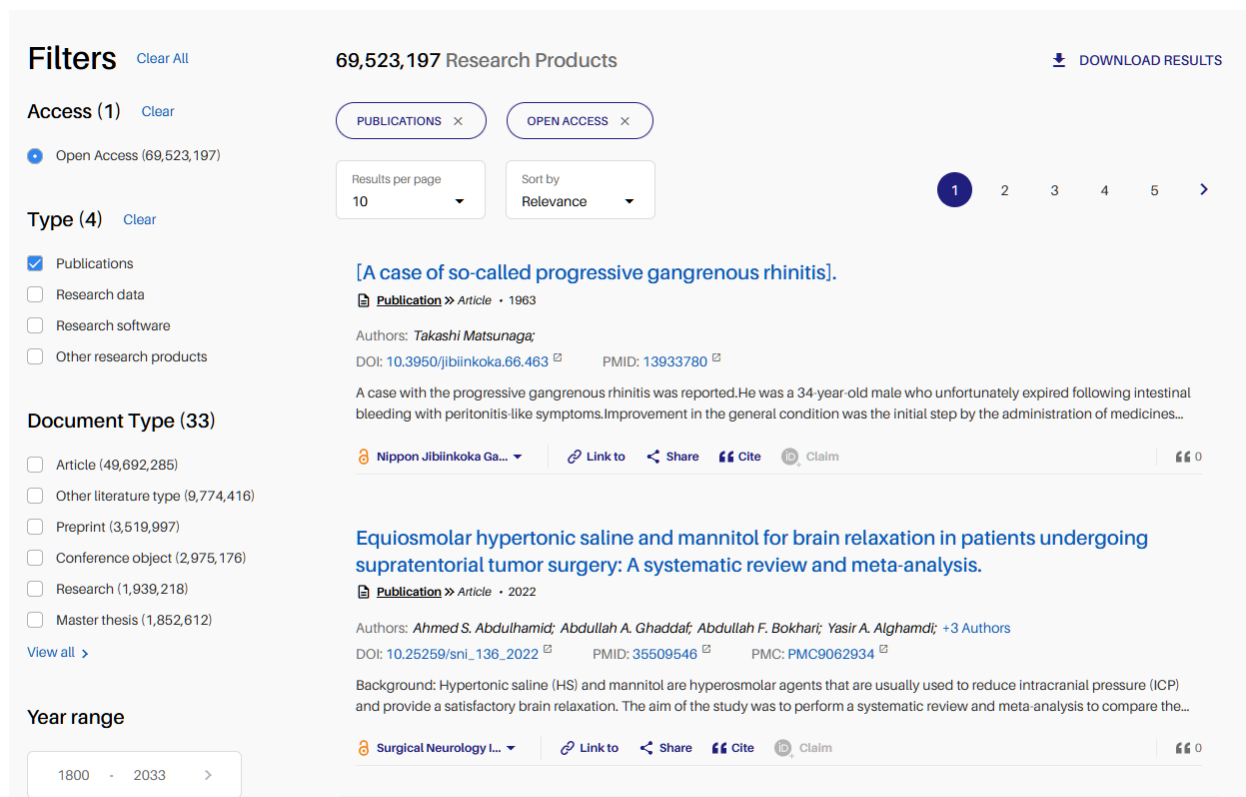
Figure 6: Advanced search

Additionally, when **browsing** research products the following *filters* are available:

- RFO
- year of publication (range)
- FoS
- country of affiliated RPO
- language
- data source
- Research Community
- Type of research output: Options include peer-reviewed publications (which are preselected as the default filter), all publications, datasets, software, and other research products.
    - Within the Monitor, "peer-reviewed publications" are set as the default filter in alignment with the project's focus. Yet, the system offers flexibility, allowing users to adjust or deselect this filter based on their preferences or to explore broader research categories.

Figure 7: Browsing research outputs with filters on the left

Exporting & Integration

The Monitor provides users with several options for data exporting and integration, catering to a variety of research and analytical needs.

Data Exporting & Downloads

The Monitor offers extensive functionality for accessing research product data and indicators. Users have the flexibility to:

- Download **research outputs** metadata records and **indicators** for quick access.
    - For datasets above 2,000 entries, users are redirected to the data dumps stored in Zenodo.
- Access the **monthly data dumps** in Zenodo.
- Access "**Download From**" links on the Research Output sections of the dashboards which provide the Open Access links for a publication, allowing users to directly download the PDF.

- *Repository Manager Only:* Download enriched metadata records of their institutional repository, via the OpenAIRE Broker service, through OpenAIRE PROVIDE. https://graph.openaire.eu/develop/broker.html
- **RESTful Data Access:** The Monitor service possesses the capability to seamlessly integrate Monitor Data with RFO or RPO platforms via OpenAIRE APIs specifically customised for the Monitor Data[9], standard protocols, and metadata schemas,[10] including harvesting metadata records from Monitor via OAI-PMH.

Visualisation Download & Integration

- Users can **download** indicator visualizations in PNG, SVG, JPEG, and PDF formats.
- Users can **embed** an indicator or a set of indicators to their website by copying the given inline frame and adding it to their HTML page accordingly.

Login Options & Additional Functionalities

While the Monitor is publicly open, enhanced functionalities are available for users who choose to log in through the **OpenAIRE AAI Service**. Supported login methods include institutional email (via EDUGAIN), GitHub, Google, LinkedIn, ORCID, or a dedicated OpenAIRE account.

Upon successful login:

- Users have the ability to **upload DOIs** to the Monitor, checking their presence and retrieving associated Open Access types and key metadata. If dealing with substantial datasets, the monthly data dump is the most efficient method. For direct comparisons, there's a cap of **100 DOIs**. However, if users are willing to navigate results page-by-page, they can upload up to **2,000 DOIs**, though this approach will not provide a complete view of all results simultaneously.
- They are granted access to the **LINKING** functionality, which empowers users to associate specific research outputs with projects, research communities, or other research outputs. Any established links will be reflected in the subsequent OpenAIRE Graph update. To maintain transparency, a log of all added links will be made publicly available. Users will be prompted for their consent during the final step of the linking process to ensure their understanding and agreement.

---

[9] Customised APIs for the Monitor Data will be based on the existing functionalities/documentation of the OpenAIRE APIs dedicated to return results from the Monitor Data.

[10] A documented case-study is the integration of OpenAIRE with SyGMA, the EC participant portal. The latter is using the OpenAIRE APIs to collect information about publications and datasets resulting from funded projects to facilitate the reporting process and acquire a reliable and authoritative source to verify this information. More information: https://www.openaire.eu/reporting-research-outputs-to-the-ec-using-the-openaire-api, https://www.openaire.eu/reporting-to-the-ec.
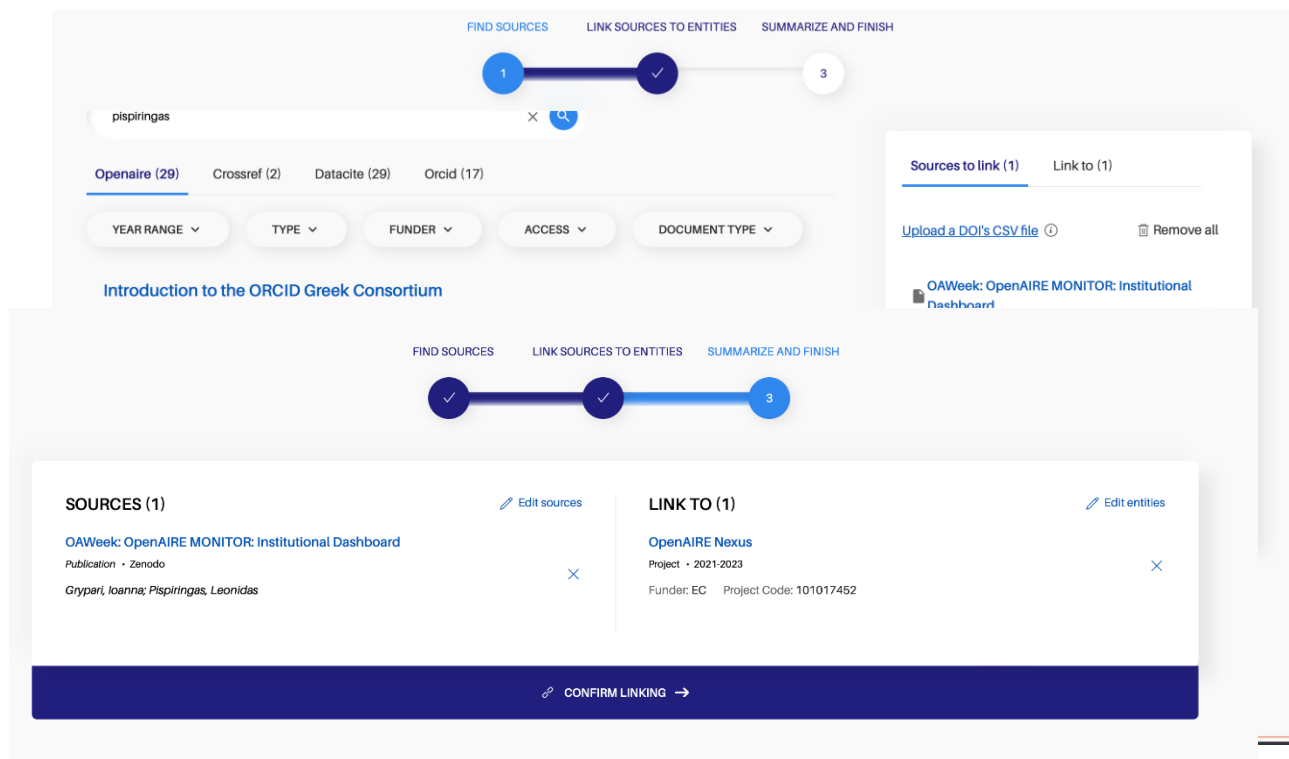
Figure 8: Screenshots from the Linking functionality

Support

Helpdesk

Users can provide feedback directly from the Monitor, which is then channelled to the OpenAIRE Helpdesk, powered by Zammad[11], for thorough issue tracking. Within the Helpdesk, a specific section is devoted to the Monitor, mirroring the structure seen at OpenAIRE Helpdesk (https://www.openaire.eu/helpdesk). It is important to note that none of the issues raised will be made open access to safeguard against potential inappropriate content.

Technical Support & System Testing

A Redmine-based ticketing system will promptly notify OpenAIRE's technical team and IReL's technical support contact about any system disruptions or technical difficulties. Additionally, every four months, we schedule regular system testing, which includes VM backups for stakeholders, VM restore tests, and offsite restore tests.

---

[11] https://zammad.org

Documentation on the Platform

The Monitor will feature **dedicated documentation pages**, focusing particularly on Frequently Asked Questions (FAQs), and the terminology and construction of indicators. These sections will be curated to ensure clarity and comprehension for all users.[12]
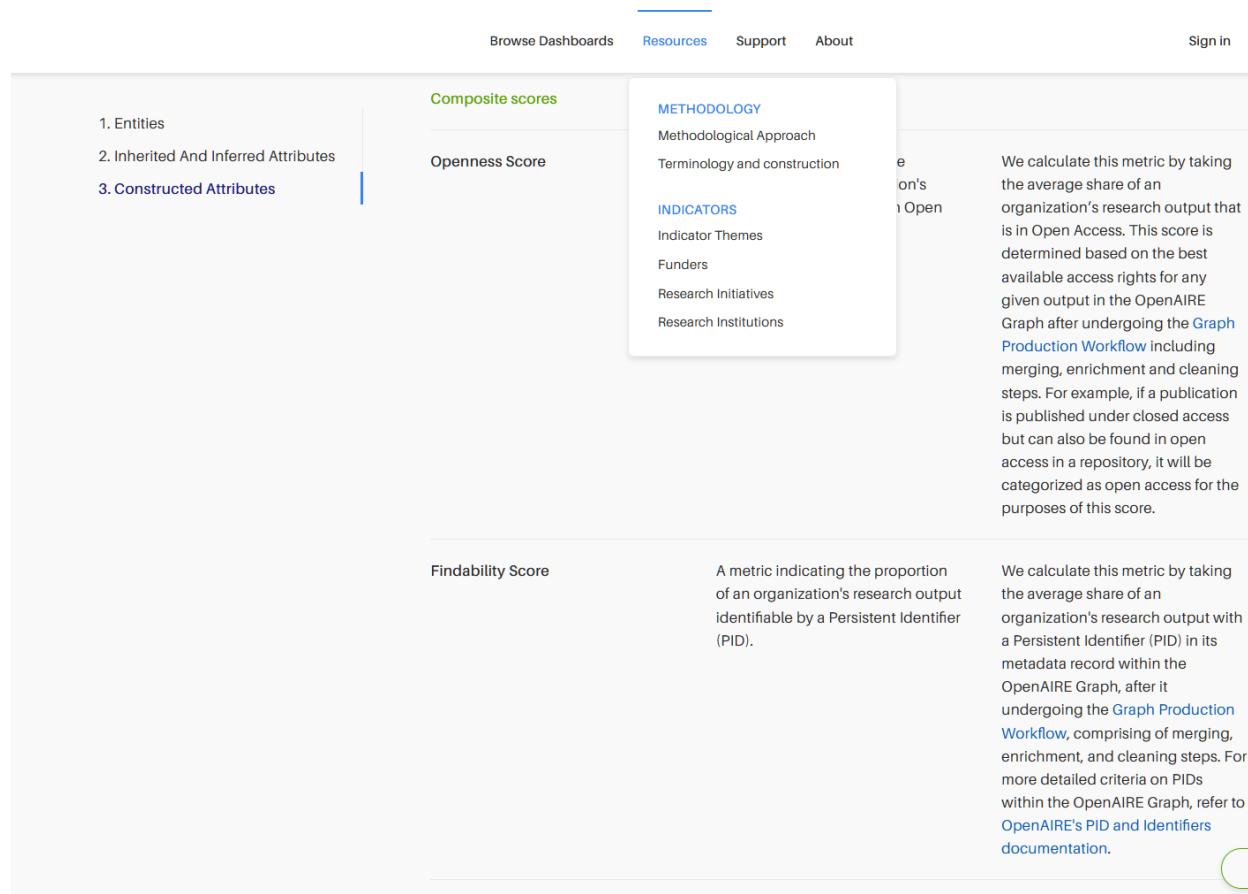
Figure 9: Sample documentation page

Integration of Matomo for Usage Analytics

The Monitor service employs the Matomo analytics platform to track usage and deliver web statistics. When relevant, it also supplies statistics to assess user engagement. The Matomo dashboard, showcasing these insights, will be publicly accessible, ensuring transparency while upholding privacy standards. COUNTER statistics are not applicable for measuring the engagement of the users with the service, as counter statistics report usage of the electronic resources (research outputs views and downloads) in a consistent way.

---

[12] Comprehensive documentation will also supplement the Code and Data deposition outlined in the relevant section below.

3. Data and Indicators

This section details the primary data and indicators vital to the Monitor's operation. We also discuss methods to improve data quality and the associated tools. An in-depth analysis of data quality, along with improvement suggestions, will be highlighted in the upcoming Report.

OpenAIRE, in alignment with the project will present a thorough representation of Irish research outputs using only public open data. While stakeholders have the option to integrate with OpenAIRE for open metadata quality enhancement, it is by no means a requirement. While complete data integration is targeted by month 6, we anticipate the continuous evolution of data quality improvements.

OpenAIRE Graph: Foundation of the Monitor

The Monitor is built upon the OpenAIRE Graph (https://graph.openaire.eu). An open resource that aggregates a collection of research data properties (metadata, links) available for funders, organizations, researchers, research communities and publishers to interlink information by using a semantic graph database approach.

The Graph aggregates around 450 million metadata records from more than 120,000 trusted scholarly communication sources worldwide, including Crossref, Unpaywall, ORCID, institutional and thematic repositories (registered in OpenDOAR, re3data.org and FAIRSharing.org), Open Access journals, data archives, and the EOSC Service Catalogue[13]. These metadata records are harvested and enriched with links between research results and projects, author affiliations, subject classifications, and links to domain-specific databases using dedicated inference algorithms. OpenAIRE's metadata records are cleaned, deduplicated, enriched, and transformed according to the OpenAIRE internal metadata model, generating the final OpenAIRE Graph[14]. A new version of the OpenAIRE Graph is available every month. The OpenAIRE Graph is available for download and reuse under a CC-BY license.

---

[13] The Microsoft Academic Graph, while no longer in production as of 2021, remains a significant component of the OpenAIRE Graph, providing valuable historical data up to its cessation point.
[14] https://graph.openaire.eu/docs/
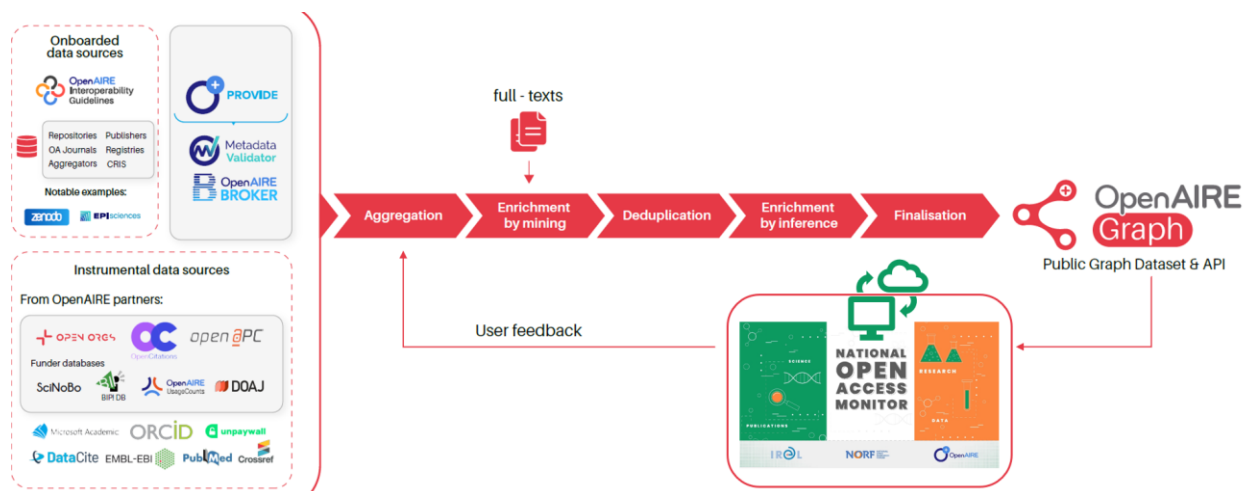
Figure 10: The Monitor and the OpenAIRE Graph Pipeline

Identifying Irish Research Outputs

This section outlines the processes through which the Monitor compiles and utilizes data from Irish Research Performing Organizations (RPOs) and Research Funding Organizations (RFOs), ensuring comprehensive and high-quality information in the Monitor.

Identification of Irish RPO Research Output

The Monitor leverages the comprehensive affiliation information already present in the OpenAIRE Graph to identify Irish RPO research output. The provenance of affiliation links in the Graph includes

- Institutional data sources registered in OpenAIRE (repositories, CRIS, Open Access Journals)
- Metadata from harvested data sources such as Crossref.
- Inferred links via text mining.
- Links created via the claim and link functionalities in OpenAIRE EXPLORE.[15]

Additionally, OpenORGS is utilized for name deduplication, ensuring consistency and precision in recognizing and distinguishing distinct RPOs and their outputs.

Identification of Irish RFO Research Output

To guarantee a thorough representation of funded research outputs, the OpenAIRE Graph establishes links between publications and their associated funding data through a variety of methods:

- Harvesting links from repositories, OA Journals, CRIS systems.

---

[15] See previous section for a description.

- Merging information from CrossRef's Open Funder Registry[16]
- Collecting links from users via the "link" functionality.
- Exchanging data with the EC's IT systems for EC/FP funding.
- Text mining of full text publications to identify the grants for 30+ funders that have joined OpenAIRE (see next paragraph).
    - Science Foundation Ireland (SFI) is one of them.

Optional Add-on: Funding Classifiers for Enhanced Data Quality

For funders that join OpenAIRE[17], we utilize advanced text mining techniques and create specialized funding classifiers to analyse publication abstracts. This process establishes links between funders, projects, and research outputs, resulting in a significant enhancement of data quality for RFOs within the Graph. OpenAIRE is open to developing custom classifiers for Irish funders upon request. Importantly, this collaboration is not obligatory for RFOs to access their dashboard and display research outputs in the Monitor. Instead, it represents an optional opportunity to elevate data quality.

Refining the Monitor Dataset

Identification of Irish OA Peer-Reviewed Publications

The Monitor further refines the set described in the previous section, to identify Irish peer-reviewed publications, their access rights and Open Access types/colours using the following criteria.

Peer-Reviewed: OpenAIRE aggregates various publication types based on the OpenAIRE Guidelines. To discern peer-reviewed publications and conference proceedings, a detailed algorithm is utilized. This algorithm factors in multiple criteria: it filters out non-relevant publication types like letters, considers publications from established journal data sources, and places importance on the presence of a DOI from Crossref. Notably, a significant criterion is the number of references a publication has. If a publication's reference count falls below a certain threshold, it's less likely to be considered as a peer-reviewed publication, given the intrinsic variations in reference counts across fields. We are also actively investigating the incorporation of DOIs from DataCite. The results of this investigation, along with the most optimal selection criteria, will be detailed in the Report. Additionally, any adjustments to this methodology can be smoothly integrated, with subsequent Monitor updates reflecting these changes.

---

[16] https://www.crossref.org/services/funder-registry/
[17] https://www.openaire.eu/funders-how-to-join-guide

Open Access: OpenAIRE uses the Budapest Open Access Initiative definition of "open access": *"By "open access" to this literature, we mean its free availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself."*[18]

Open Access Types/Colours: OpenAIRE and OpenAIRE MONITOR adopt the "open access types" definitions from the Unpaywall construction methodology.[19]

Use of Persistent Identifiers (PIDs): The Monitor uniquely defines an Irish scholarly publication as one that contains a persistent identifier (PID) of an Irish organization, whether it is in the publication metadata, PID metadata, or within the publication content. OpenAIRE seamlessly integrates various PIDs for research outputs and organizations. The deduplication process ensures that metadata records from different data sources are merged with publicly displayed provenance information. This guarantees comprehensive coverage and data consistency.[20]

Data Disambiguation Techniques

Deduplication in OpenAIRE: The OpenAIRE Graph collects metadata records about scholarly works from different providers, which can carry different information. To provide accurate statistics, OpenAIRE merges duplicate records of the same scholarly work. The deduplication process is described in detail in the following link: https://graph.openaire.eu/docs/graph-production-workflow/deduplication/

Organizations: Organizations within OpenAIRE are aggregated from diverse registries and undergo a deduplication process via OpenOrgs. This tool merges automation with a "human in the loop" mechanism. It is designed to cluster records that are more likely to be analogous, employing both URL-based and title-based functions. Through the process of grouping duplicates, representative organizations not only inherit all attributes from the combined records but also maintain a record of their origin. As discussed earlier, managers overseeing the national, RPO, and RFO dashboards will have access to OpenOrgs, empowering them to deduplicate Irish RPO records. It is imperative to note

---

[18] https://www.budapestopenaccessinitiative.org/read/
[19] See Glossary for definitions.
[20] Additional documentation on the various identifiers utilised in the OpenAIRE Graph can be found here https://graph.openaire.eu/docs/data-model/pids-and-identifiers.

that all modifications made will be publicly recorded. These changes will first be showcased in the Sandbox environment, providing an avenue for review and validation.

Journals, Publishers and Licenses: To ensure clarity and accuracy in its statistics, the Monitor disambiguates journals by employing their ISSN numbers and is engaged in a systematic approach to normalize licenses. Leveraging Crossref metadata, such as ROR ids and DOI prefixes, among others, the team is in the process of disambiguating publishers. This initiative is further enhanced with custom text similarity algorithms. A significant effort is set to take place by M6, marking an important milestone in this endeavour. However, the refinement will be a continuous process, persisting beyond the release to provide the most coherent and meaningful data presentation possible by M11.

Authors: Researcher dashboards and corresponding research outputs will be based on ORCID IDs as presented in the previous sections.

Indicators

The Monitor provides a multifaceted perspective on Open Science through a selection of key indicators, each illuminating a distinct aspect of the research ecosystem.

Open Science Indicators

- Access rights, Open Access types (Green, Gold, Hybrid, Bronze, Closed)
- Diamond vs OA with APCs
- Unrealised Open Access
- Embargoed vs. Immediate Open Access
- APCs, Plan S, Transformative Agreements

Bibliometrics

- Citations, Views, Downloads

The upcoming draft Report will delve further into the quality of these indicators, reinforcing the Monitor's commitment to accuracy and depth. The Monitor will have the capacity to incorporate additional indicators in the future.

Historical Snapshots

OpenAIRE will provide through the historical snapshots of the Monitor metrics to measure the evolution in the share (%) of OA publications. Thus, users will be able to view progress in Open Science uptake over time, using data from different versions of the Monitor data.

These historical snapshots will be taken monthly and the first one will be taken upon the release of the Monitor.

4. Code, Data Deposition & Documentation

Code & Data Deposition

In compliance with transparency and OS standards, OpenAIRE ensures the following principles throughout the Monitor's development.

Documentation and Data Preservation:

- All processes involved in delivering the Monitor will be documented and deposited in Zenodo. These documents will be publicly available under an open license.
- The data that underpins the National Open Access Monitor will be deposited in Zenodo under an open license. This deposit will occur before the contract terms conclude or renew. OpenAIRE will regularly provide data dumps, which include all data and indicators, to Zenodo.

Source Code Openness:

- OpenAIRE will release the Monitor's source code, the data dump, the Monitor dashboard code and the existing software frameworks APIs and libraries that will be used to the public under an open-source license in Zenodo.

Documentation

The Monitor will be accompanied by comprehensive documentation covering various essential aspects, including:

- **Data Sources:** Detailed information about the data sources, including their origin (RPO Repository, RFO Data), and an overview of the Monitor Data's characteristics.
- **Methodological Approach**: A clear description of the approach employed in the Monitor's development. This approach is based on principles such as inclusion, transparency, quality, and state-of-the-art technology. It encompasses methods, tools, and processes employed for data collection, analysis, and visualization.
- **Terminology:** An explanation of the terminology used within the Monitor. This will clarify the entities, attributes derived from the data (inherited, inferred, or constructed), and definitions of relevant concepts, indicators, and metrics.

As described in the "Common Features" section, the Monitor will also include dedicated documentation pages for usability and accessibility, adopting content from the following sources.

- OpenAIRE Graph:
  - https://graph.openaire.eu/docs/

- OpenAIRE MONITOR:
    - https://monitor.openaire.eu/methodology/methodological-approach
    - https://monitor.openaire.eu/methodology/terminology#entities

5. Project Coordination, Engagement & Training

Project Coordination

IReL and OpenAIRE project teams currently engage in regular online meetings every two to three weeks, ensuring the project progresses effectively. While meetings typically occur at fixed times, adjustments can be made to accommodate participants' schedules. Each meeting follows a predetermined agenda, distributed to participants beforehand. The agenda typically encompasses:

- revisiting the project's timeline and deliverables,
- reviewing progress made since the prior meeting,
- addressing challenges encountered by both teams and discussing resolutions,
- exchanging feedback aimed at refining the Monitor's operations,
- plotting the tasks each team should tackle next, and
- setting the date and time for the following meeting.

A specific moderator oversees each meeting, ensuring its productivity and relevance. This moderator then circulates a summary of the key points and decisions from the meeting to all participants.

In addition to these meetings, we have established a shared workspace on Microsoft Teams where team members have access to meeting minutes, shared documents, and other relevant materials. This collaborative space fosters transparent communication and ensures everyone remains updated.

Monitor Advisory Group

Moreover, OpenAIRE will meet with the Monitor Advisory Group. This group offers invaluable feedback on the Monitor's development. Such interactions are synchronised with the project's milestones and deliverables, as detailed in the project plan. Once vetted by project partners, all relevant documents, including meeting minutes, are made publicly available on Zenodo.

Stakeholder Engagement

Stakeholder involvement is foundational to the success of the Monitor. Active participation ensures that the Monitor's data is not only of high quality but also that the platform is relevant and inclusive, providing true value to its users.

Pursuit of Data Excellence

While the Monitor leverages public open data, stakeholder collaboration is paramount to elevate the quality, depth, and relevance of this data foundation.

While it's not necessary for the delivery of the Monitor, the collaboration with IReL can enhance data quality and representation. When IReL provides OpenAIRE with a list of stakeholders, it will help in refining the data within the dedicated dashboards for all key participants. Stakeholders can improve the quality of their public data through direct sources like Crossref or via OpenAIRE's specialized tools, including OpenOrgs. Registering with OpenAIRE and sharing project data can deepen the connections and insights. However, these steps are entirely voluntary and aim to optimize data quality.

Inclusivity through Active Engagement

The Monitor is conceived to align seamlessly with its users. Through targeted stakeholder engagement, we endeavour to cater to the distinct needs of a diverse user community. This approach not only amplifies the Monitor's proficiency but also refines the user experience, producing a solution shaped by and tailored to its users.

IReL spearheads stakeholder engagement activities, with OpenAIRE on hand to assist with presentations, training, and additional discussions. This collaborative effort ensures that the Monitor resonates with those it is crafted for.

To enhance this collaborative spirit, specialized training sessions are in the pipeline. These sessions aim to equip stakeholders with pivotal knowledge and skills, nurturing an integrated collaboration.

Stakeholders interested in participating will be asked to sign either one or two consent forms, depending on their specific engagement: one from IReL and/or another from OpenAIRE. These forms authorize the use, storage, and preservation of their contributions throughout the Monitor's development. Additionally, the project meticulously tracks all stakeholder engagements to ensure that feedback is captured comprehensively and that every key stakeholder is aptly represented.

Training

OpenAIRE, in partnership with IReL, is developing a training programme for the Monitor, catering to the distinct needs of various user groups.

Planned Training Sessions

- Mini-Training Sessions for RFOs and RPOs: These specialised sessions offer insights and guidance tailored for Research Funding and Performing Organisations.

On-Demand Training Areas

Based on feedback from engagement activities, led by IReL, additional training sessions can be offered as required, on the following topics

- *Platform Administration:* Introducing users to the Monitor's functionalities and navigation.
- *Data Quality & Metadata Standards*
- *Indicators*: Equipping users with the knowledge to effectively utilise and interpret data presented on the Monitor.
- *Open Science Topics*: Providing insights on areas such as Open Access publishing, data management, and the EOSC. The exact topics will be refined post-project initiation based on feedback.

IReL will oversee the coordination, ensuring training sessions (in webinar format) align with the needs of stakeholders.

If there is significant interest, OpenAIRE is equipped to conduct "train the trainer" sessions. This approach equips participants with the ability to disseminate this knowledge within their organisations.

6. National Open Access Report

The Report, a key project deliverable, will outline the current landscape of OA in Ireland. Its primary objectives are:

- **Baseline Analysis:** The report will provide a clear overview of OA in Ireland, covering both the national context and specific domains.
- **Data Evaluation:** Assessing the quality and coverage of the current data will be vital for a comprehensive understanding:
    - The OpenAIRE Graph will be used to identify existing data sources and potential gaps.
    - The quality and accuracy of data will be examined, with a particular emphasis on researcher affiliations.
    - The Metadata Validator will be deployed to check the completeness and FAIRness of records from repositories and OA journals.
- **Challenges & Recommendations:** Identifying issues and offering solutions will be central:
    - The report will highlight potential obstacles in long-term OA monitoring and other Open Science indicators.
    - Practical recommendations will be provided, including data validation techniques and enrichment workflows, aiming for continuous and consistent OA monitoring in Ireland.

A first draft of this report will be provided to the Monitor Advisory Group on October 9th for feedback.

7. Timeline and Milestones

Lastly, The Gantt chart presented below presents the project timeline, showing progress by month and the respective milestones.