National Open Access Monitor Ireland

Report

Ioanna Grypari, Leonidas Pispiringas, Natalia Manola

OpenAIRE AMKE

March 2024

Executive Summary[1]

Ireland is entering the next phase of the NORF National Action Plan for Open Research, and to navigate this journey successfully, a comprehensive understanding of the current state of Open Access (OA) within the country is crucial. The National Open Access Monitor, Ireland (https://oamonitor.ireland.openaire.eu/)[2] represents a significant leap forward in this regard. It serves as an innovative platform designed to promote and comprehend OA research and scholarly publishing within Ireland. This report, accompanying the Monitor, offers a thorough analysis of OA in Ireland, aiming to provide valuable insights and recommendations for the scholarly community[3].

Key Observations and Insights

Scholarly Production and Research Landscape Of the total **423,893** Irish publications, **333,404** (**78.7%**) are peer-reviewed, showcasing a consistent upward trend in peer-reviewed output. This reflects a dynamic research environment and increased adoption of digital practices.

Medical and Health Sciences dominate publication volume, reflecting global research trends. Significant contributions also come from Natural Sciences, Engineering, and Technology, highlighting the country's innovation and technological progress. While Social Sciences and Humanities and the Arts might seem less represented, as is common

[1] The National Open Access Monitor, Ireland is delivered as part of the National Open Access Monitor Project, managed by the Irish Research eLibrary (IReL) at Maynooth University. The project has received funding from Ireland's National Open Research Forum (NORF) under the Open Research Fund Call. *Corresponding Author: ioanna.grypari@openaire.eu*

[2] The Monitor from here on.

[3] The Monitor was deployed in early January 2024. This report utilizes data from the December 2023 version of the OpenAIRE Graph (https://graph.openaire.eu/) to ensure the baseline analysis provided herein reflects the latest available information at the moment of the Monitor's deployment.

due to differences in publication norms, they play a crucial role in Ireland's research landscape through diverse contributions across various disciplines.

Journals are the primary platform for peer-reviewed publications, with a growing engagement with repositories indicating an evolving research dissemination landscape.

Open Access and Licensing The share of licensed OA publications[4] increased from **21.6%** in 2007 to **66%** in 2022, showcasing a significant move towards open scholarly communication. However, **10%** of scholarly output remains unlicensed OA, and **16.6%** under Closed Access, in 2022, pinpointing areas for further OA advancement.

Medical and Health Sciences lead in licensed OA publications at **44.3%,** with notable disparities across fields, highlighting slower OA uptake in some areas due to cultural or infrastructural barriers.

OA Routes and Publishing Models Gold and Hybrid OA are the primary routes for licensed OA publications in Ireland, with Gold OA comprising **38.1%** and Hybrid OA **30.8%.** Repository-mediated OA accounts for only **6.5%** of the total, though it shows growth indicating its emerging significance. It represents **15%** of OA licenced publications by 2022.

There is also a discernible trend towards Hybrid OA suggesting evolving journal policies and the impact of initiatives like Plan S and Transformative Agreements.

Plan S Compliance Science Foundation Ireland (SFI), as the sole funder part of cOAlition S in Ireland, has seen a gradual increase in OA licensed peer-reviewed publications, reaching **85.1%** in 2022 and a decrease in unrealized OA to **12.9%**.

Overall, the majority of Plan S-compliant publications are published in Gold OA journals with Article Processing Charges (APCs), accounting for **39.6%**, pointing to a significant reliance on APCs that may raise concerns about inclusivity. There has been a notable increase in Plan S-compliant publications without APCs, largely due to Transformative Agreements, while Diamond OA journals have a minimal presence at **0.2%**, indicating possible underutilization or lack of support. Additionally, there are significant variations in Plan S compliance across different fields, highlighting the influence of discipline-specific factors on OA adoption.

Transformative Agreements Transformative Agreements (TAs) have seen an increase, covering **23.3%** of publications that would otherwise pay APCs in 2022, highlighting efforts to expand scholarly access and reduce author costs. Particularly, Social Sciences and

[4] Proper licensing is a requirement for Open Access, as defined by the Budapest Open Access Initiative https://www.budapestopenaccessinitiative.org/

Humanities benefit from TAs, with Social Sciences having the highest number of TA publications, showing active engagement in these fields.

Article Processing Charges (APCs) The availability of APC data is limited, with only **1.4%** coverage of the **83,614** peer-reviewed publications that incurred APCs, highlighting a significant gap in understanding the financial dynamics of OA publishing and its effects on the Irish scholarly landscape.

FAIR Principles[5] and Metadata Completeness The use of Creative Commons (CC) licenses in Ireland has significantly increased from **6%** in 2007 to **56.4%** in 2022, reflecting a trend towards greater openness and reusability of research. However, the continued issuance of restrictive licenses by major publishers limits the full potential for accessibility and utility of research. The assignment of publication PIDs like DOIs to **85.6%** of peer-reviewed publications enhances their discoverability and long-term accessibility, despite challenges in harmonizing these identifiers for seamless integration. The adoption of ORCID iDs has also grown, from **51.6%** in 2007 to **64.6%** in 2022, highlighting their importance in accurately attributing work and tracking contributions. These developments indicate both advancements and ongoing challenges in adhering to FAIR principles, aiming for open and accessible scholarly communication in Ireland while emphasizing the need for high-quality data management practices.

Overall Observations Ireland's scholarly landscape is actively adapting to new norms and expectations in OA. While some fields are advancing in licenced OA, others lag, often constrained by traditional publication models and data representation. The trends indicate a positive shift towards more open and transparent research practices. Still, the persistence of Closed Access publications and disparities in licensing practices call for continued efforts to foster a more open scholarly communication environment.

The rise of Hybrid OA, the sustained presence of Gold OA, and the growing importance of Transformative Agreements showcase a publishing sphere adjusting to evolving mandates and researcher needs. Monitoring these changes is essential to ensure that the benefits of OA are maximized for the entire scholarly community. The disparities across disciplines and the varied impact of Transformative Agreements highlight the need for nuanced and **field-specific strategies** to effectively transition to more open and accessible research models.

Main Data Challenges

Inconsistent metadata within Open Access monitoring presents a formidable challenge. The presence of inconsistencies across various metadata elements, including naming conventions and standardization, hinders the accurate identification of critical entities

[5] The FAIR principles have been adapted from their dataset-centric origins to apply to scholarly publications.

such as Research Performing Organizations (RPOs), publishers, journals, and licences. These discrepancies result in a lack of standardized practices, ultimately affecting the accuracy and reliability of the data.

Complex licence reporting adds to the complexity. This challenge has two facets: firstly, there is a significant gap in including licences in metadata, a crucial element for categorizing publications as OA. Secondly, the intricate understanding of licensing nuances, especially in mapping various licences and their degrees of openness, complicates the precise classification of publications.

Timeliness and indexing issues further complicate the data landscape. Delays in reporting, incomplete indexing of recent publications, and underutilization of ORCID iDs for authors hinder the comprehensive and accurate representation of research outputs and their impacts.

Access rights and version complexity pose additional hurdles. The dynamic nature of access rights, combined with challenges in distinguishing different publication versions, adds intricacy to metadata. This complexity makes it challenging to provide a clear view of publication statuses, impacting data accuracy.

Limited accessibility to full-text PDFs for a significant portion of OA peer-reviewed publications hinders researchers' ability to access and utilize scholarly content effectively. This constraint also limits the potential for advanced text mining and insights extraction.

In summary, these challenges collectively impede the accuracy, completeness, and reliability of OA monitoring data. Addressing these issues is critical to fostering a more robust scholarly communication environment and maximizing the benefits of Open Access.

Recommendations in Brief

In the pursuit of advancing OA in Ireland, the Monitor plays a crucial role. This report offers a comprehensive view, outlining direct and indirect improvement strategies, as well as long-term solutions and workflow recommendations, to enhance OA monitoring.

Direct Improvement Strategies emphasize leveraging existing tools and functionalities within the OpenAIRE ecosystem. These strategies are tailored to address challenges in metadata accuracy and completeness. By encouraging key stakeholders like policy makers, publishers, research funding and performing organizations (RFOs/RPOs), and researchers to adopt specific actions such as deduplication of organization names,

joining OpenAIRE for data enrichment, and engaging with the Monitor dashboard, the goal is to enhance the overall quality and user experience of the Monitor. This approach not only refines data management but also encourages active participation and feedback from various stakeholders in the research community.

Indirect Improvement Strategies focus on elevating the overall quality of data harvested into OpenAIRE. These strategies are designed to address timely reporting, accurate metadata entry, and clear licensing, involving a collective effort from all stakeholders. Key recommendations include the development of policies for the registration and timely deposition of research outputs, implementation of consistent metadata standards across publications, and regular updating of publication records in institutional repositories. These actions are essential for ensuring the precision and comprehensiveness of data, thereby facilitating a more robust and reliable OA monitoring system.

The Long-Term Solutions and Workflow Recommendations shift the focus to sustainable improvements and future readiness. Focusing on sustainable data management, long-term data preservation, routine data quality audits, and comprehensive training and capacity building, these recommendations are crafted to ensure that the OA ecosystem is well-equipped to adapt, grow, and maintain high standards of data integrity and usefulness. Emphasis is placed on developing frameworks for iterative improvement in metadata standards, promoting global policy initiatives that align with the FAIR principles, and establishing collaborative platforms for metadata correction and enrichment.

By aligning these strategies closely with the challenges and gaps identified in the Data Evaluation, the report ensures a data-driven and targeted approach. This strategy addresses current needs and prepares for future challenges in OA monitoring and research data management, laying the groundwork for a more open, accessible, and efficient scholarly communication environment in Ireland.

Contents

List of Figures

List of Tables

Abbreviations

AAM	Author Accepted Manuscript
APC	Article Processing Charge
CC	Creative Commons
DOI	Digital Object Identifier
EC	European Commission
EOSC	European Open Science Cloud
FAIR	Findable, Accessible, Interoperable, Reusable
FoS	Fields of Science
Graph	OpenAIRE Graph
GRID	Global Research Identifier Database
IReL	The consortium of Irish research libraries
ISNI	International Standard Name Identifier
ML	Machine Learning
Monitor	National Open Access Monitor, Ireland
MU	Maynooth University
NLP	Natural Language Processing
NORF	National Open Research Forum
OA	Open Access
OFR	Open Funder Registry
OS	Open Science[6]
ORCID	Open Researcher and Contributor ID
PID	Persistent Identifier
RFO	Research Funding Organisation
RPO	Research Performing Organisation
ROR	Research Organization Registry
SDG	Sustainable Development Goal
SFI	Science Foundation Ireland
TA	Transformative Agreement
VoR	Version of Record

[6] This term is interchangeable with Open Research.

Glossary

TERM	DEFINITION
ARTICLE PROCESSING CHARGE (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.*
PUBLICATION	*Research products intended for human reading (published articles, pre-prints, conference papers, presentations, technical reports, etc.)*
OPEN ACCESS	*We 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."*[7]
TRANSFORMATIVE AGREEMENTS	*Transformative Agreements are those contracts negotiated between institutions (libraries, national and regional consortia) and publishers that transform the business model underlying scholarly journal publishing, moving from one based on toll access (subscription) to one in which publishers are remunerated a fair price for their Open Access publishing services*[8]*.*
JOURNAL BUSINESS MODELS	
OA (GOLD)	*A journal that publishes only in OA.*
DIAMOND OA	*An OA (Gold) journal that does not charge article processing charges (APCs).*
SUBSCRIPTION	*A journal that charges for access to its articles.*
HYBRID	*A subscription journal where some of its articles are open access.*

[7] https://www.budapestopenaccessinitiative.org/read/

[8] https://www.coalition-s.org/Transformative-journals-faq/

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).*[9]
ROUTES TO OPEN ACCESS	
GREEN WITH LICENCE	*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 must be accompanied by a specified licence that outlines how the material can be used, shared, and distributed.*
HYBRID OA	*Hybrid articles are free to read at the time of publication, with an open licence. 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".*
UNREALISED OA	
BRONZE	*Bronze articles are free to read on the publisher's website, without a licence 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.*
CLOSED	*Articles with Closed Access rights can only be accessed for a fee.*
GREEN WITHOUT LICENCE	*Green articles deposited in a repository without any licence specified.*
ACCESSIBILITY - INTEROPERABILITY	
ACCESSIBLE	*A publication is accessible if the text file can be fetched via a valid URL in its metadata.*[10]

[9] https://www.coalition-s.org/Transformative-journals-faq/

[10] I.e.., if a publication does not include a valid URL, we cannot assess accessibility.

INTEROPERABLE	*A publication is considered interoperable if its full-text is in a machine-readable format, allowing machines to process and understand the content*[11].

[11] This is not the only prerequisite for interoperability but we adopt this definition here for exposition.

1 Introduction

As Ireland embarks on the next phase of the NORF National Action Plan for Open Research, it is crucial to understand the current state of Open Access (OA) in the country. The National Open Access Monitor, Ireland, represents a significant step in this direction. It marks an advancement in understanding and promoting Open Access research and scholarly publishing within Ireland. This report is an integral companion to the Monitor, a dynamic and innovative platform designed to guide Ireland's scholarly output towards 100% Open Access. Currently in its pilot phase until June 2024, the Monitor is accessible at https://oamonitor.ireland.openaire.eu/.

This report aims to establish a comprehensive baseline analysis of OA in Ireland, offering both a holistic and domain-specific perspective. It evaluates current Irish OA publishing practices and uptake, highlights challenges, and proposes solutions.

Open Access monitoring is a pivotal tool in decision-making, policy formulation, and the advancement of scholarly communication. By providing clear insights into the state of OA, this monitoring effort enables stakeholders to make informed decisions, strategize effectively, and foster a more open and collaborative research environment. It is a key driver in the transformation towards a more transparent, accessible, and equitable scholarly landscape.

The analysis of OA in Ireland, presented in Section 2, provides a detailed view of the current state and progression of scholarly communication, focusing on the Irish scholarly production, various Open and FAIR aspects and Plan S compliance, APCs, and Transformative Agreements. The section concludes by bringing together key observations and insights from our analysis, underscoring both the achievements and the challenges that need to be addressed to further advance OA in Ireland.

Section 3 outlines the methodological steps undertaken in the analysis. This transparent approach assures the integrity and reliability of our findings and paves the way for the data evaluation and challenges presented in Section 4. These findings are crucial for understanding the current landscape of data quality and integrity, which are fundamental in shaping effective Open Access monitoring and policy development.

Finally, Section 5 presents a series of strategies and recommendations aimed at enhancing OA monitoring in Ireland. Our approach spans three interconnected tiers of improvement, encompassing direct improvement strategies for the Monitor, indirect

improvement strategies for data enhancement, and forward-looking strategies for long-term OA monitoring enhancements.

This report marks a pivotal step in Ireland's progression towards a completely open scholarly environment. It lays the groundwork for continued analysis, policy formulation, and strategic initiatives in Open Access publishing.

1.1 Methodological Foundations for the Monitor and Report

In the development of both the National Open Access Monitor, Ireland, and this accompanying report, we have employed a methodological framework that is deeply rooted in the principles of Open Science. This approach is designed to provide thorough, accurate, and user-focused insights into Ireland's Open Access landscape. Central to our methodology is the OpenAIRE Graph (https://graph.openaire.eu/), which serves as the foundational data resource[12]. Our methodological principles encompass the following core areas.

Openness and Transparency: Our methods are grounded in transparency, using only publicly available data. We strictly adhere to the FAIR (Findable, Accessible, Interoperable, and Reusable) principles and international standards. This commitment ensures that our findings are not only trustworthy and replicable but also readily accessible for public scrutiny and engagement.

Comprehensive Coverage and Precision: By harnessing the extensive data capabilities of the OpenAIRE Graph and collaborating closely with various stakeholders, our aim is to capture the most complete and nuanced picture of OA in Ireland. This involves integrating data from a wide array of sources, ensuring that our indicators are both accurate and rich in detail.

Readiness and Timeliness: Our framework is built upon established open databases and proven technologies in natural language processing (NLP) and machine learning (ML). These tools are integrated into OpenAIRE's operational workflows, enabling us to provide timely and relevant results that are in step with the latest developments in the field.

Engagement and Inclusivity: At the core of our methodology is a commitment to serve a diverse array of users, from individual researchers to policy makers. We focus on creating

[12] See Section 3 for details.

an intuitive, user-friendly experience, underpinned by clear communication and an openness to feedback. This approach ensures that our platform is not only informative but also engaging and responsive to the varied needs of all stakeholders in the research community.

These guiding principles serve as the foundation for our work, directing both the National Open Access Monitor and this Report. They ensure our efforts are methodologically sound and comprehensive, while staying attuned to the evolving needs and expectations of Ireland's research community.

2 Baseline Analysis

In this chapter, we conduct a baseline analysis of Open Access (OA) in Ireland, examining the various facets that shape its landscape. Our analysis, visualizations, and detailed breakdowns cover a wide range of years without any restrictions. However, for visualizations that show yearly trends and recent changes, we specifically use data from the years 2007 to 2022 to provide clear insights.

Our exploration begins with an analysis of scholarly production, where we gage the volume and trends of peer-reviewed publications. We segment this data by year, scientific field, and publication types, providing insights into the growth and diversification of scholarly outputs.

We proceed to explore the different OA pathways, specifically contrasting repository-mediated and publisher-mediated OA. This examination includes assessing both immediate and embargoed OA, as well as instances of Unrealised OA. This part of our analysis seeks to unravel the complexities in the distribution and evolution of these OA models across the Irish academic landscape, elucidating how these approaches are adopted and their trajectories over time and various disciplines.

The chapter also delves into compliance with the FAIR principles, evaluating the prevalence of essential elements such as licences, abstracts, and ORCID IDs in publications. This assessment aims to reveal how closely Irish research aligns with these international standards, identifying areas of strong adherence and potential opportunities for enhancement.

We then turn our focus to Plan S compliance trends among Irish OA publications and the financial aspects of OA publishing in Ireland, with a specific focus on Article Processing

Charges (APCs) and the impact of Transformative Agreements. This section aims to shed light on the economic dynamics underpinning OA publishing practices.

The chapter concludes with a synthesis of our key findings, integrating the various analytical threads to present a comprehensive overview of the current state of OA in Ireland.

2.1 Scholarly Production

This section delves into an in-depth examination of Ireland's scholarly output, laying essential groundwork for our broader analysis. It assesses the distribution of publications by period, discipline, and type, offering a well-rounded view of the country's research environment. Such an analysis is fundamental for a deeper exploration of OA practices within the Irish scholarly realm.

Table 1: Outline of overall production of Irish publications

Irish publications
423,893
and % of Irish Peer-Reviewed Publications
333,404 *(78.7%)*

Of the **423,893** total Irish publications[13], **333,404** (78.7%) are peer-reviewed, illustrating the significant emphasis on academic integrity and quality control in Irish research. This high proportion of peer-reviewed content ensures that the majority of Ireland's scholarly output is validated through rigorous evaluation, contributing to the credibility and reliability of its research contributions on a global scale.

The analysis of Irish publications indicates a steady increase in peer-reviewed output from **8,717** in 2007 to **22,206** in 2020, suggesting a growing research landscape in Ireland. This increase may be partly due to the adoption of digital practices, such as the use of Digital Object Identifiers (DOIs) and digital repositories, which improve the visibility and accessibility of research[14].

[13] Refer to Section 3.2 for the methodology used to identify Irish publications, namely those authored by individuals affiliated with or funded by an Irish organization.

[14] The decline noted in 2022 is likely attributable to delays in indexing and reporting, reflecting common fluctuations in publication data.

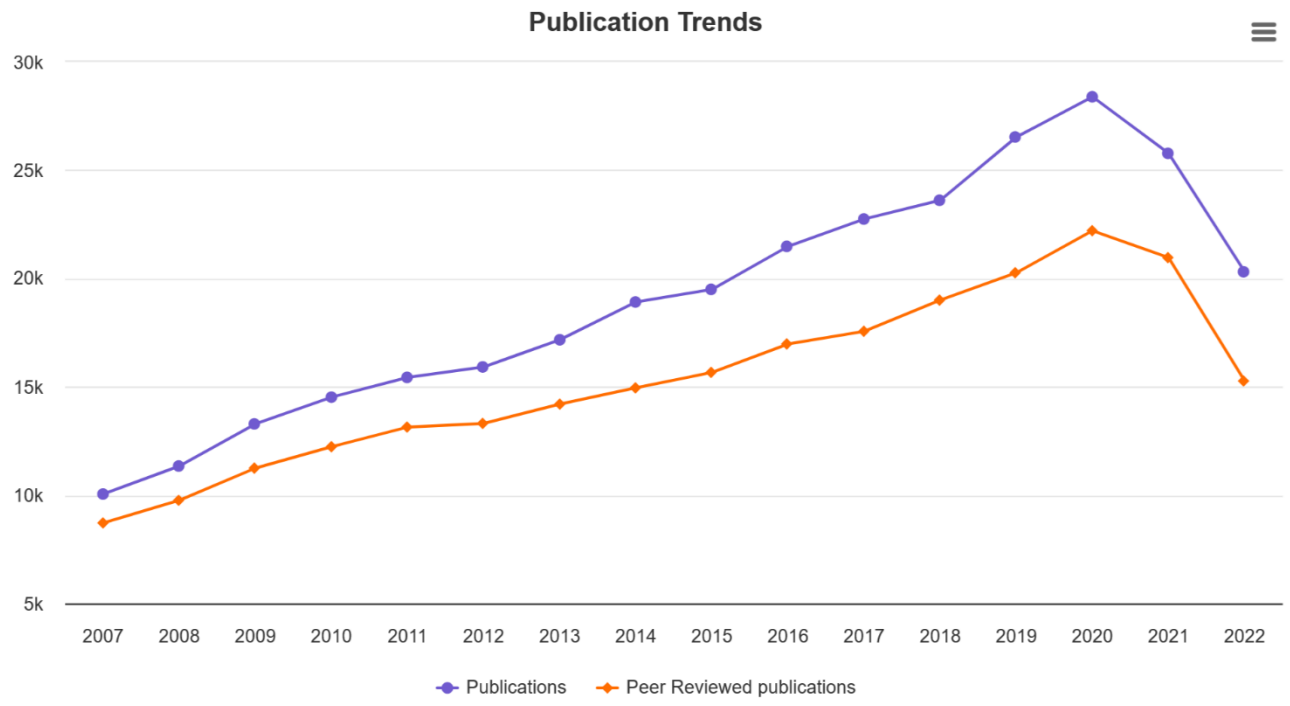

Figure 1: Publications Trends

Analysing the progression of peer-reviewed research in Ireland reveals distinct patterns across different Fields of Science (FoS), from broad level 1 to more specific level 2 categories. The data shows that Medical and Health Sciences lead in publication volume, which is consistent with global academic trends where health-related research often receives substantial focus and resources. The strong output in the Natural Sciences and Engineering & Technology suggests that these areas also receive significant attention, likely due to their direct impact on innovation and economic development—patterns that are observed worldwide.

When we examine the subfields within these broader categories (FoS level 2), interesting patterns emerge that may point to specific areas of research intensity or specialization within Ireland. For example, within the Natural Sciences, Physical and Chemical sciences show particularly high publication counts, possibly indicating niche strengths or centres of excellence within these areas. In Engineering & Technology, Electrical, Electronic, and Information Engineering stand out, reflecting a possible national strategic investment in these cutting-edge sectors.

While disparities in publication numbers could be indicative of unequal funding and resource distribution across disciplines, they also reflect the inherent nature of different fields. For example, Social Sciences and Humanities typically have different outputs, such

as books or comprehensive studies, which are integral to these fields but may not be represented in the same way as journal articles. Nevertheless, the significant number of publications in Social Sciences, particularly in economics, business, and education, suggests that these areas are well-established and actively contributing to Ireland's scholarly discourse. The prominence of history, archaeology, languages, and literature within Humanities and the Arts underscores the importance of these disciplines in contributing to a comprehensive understanding of both global and local cultural narratives.

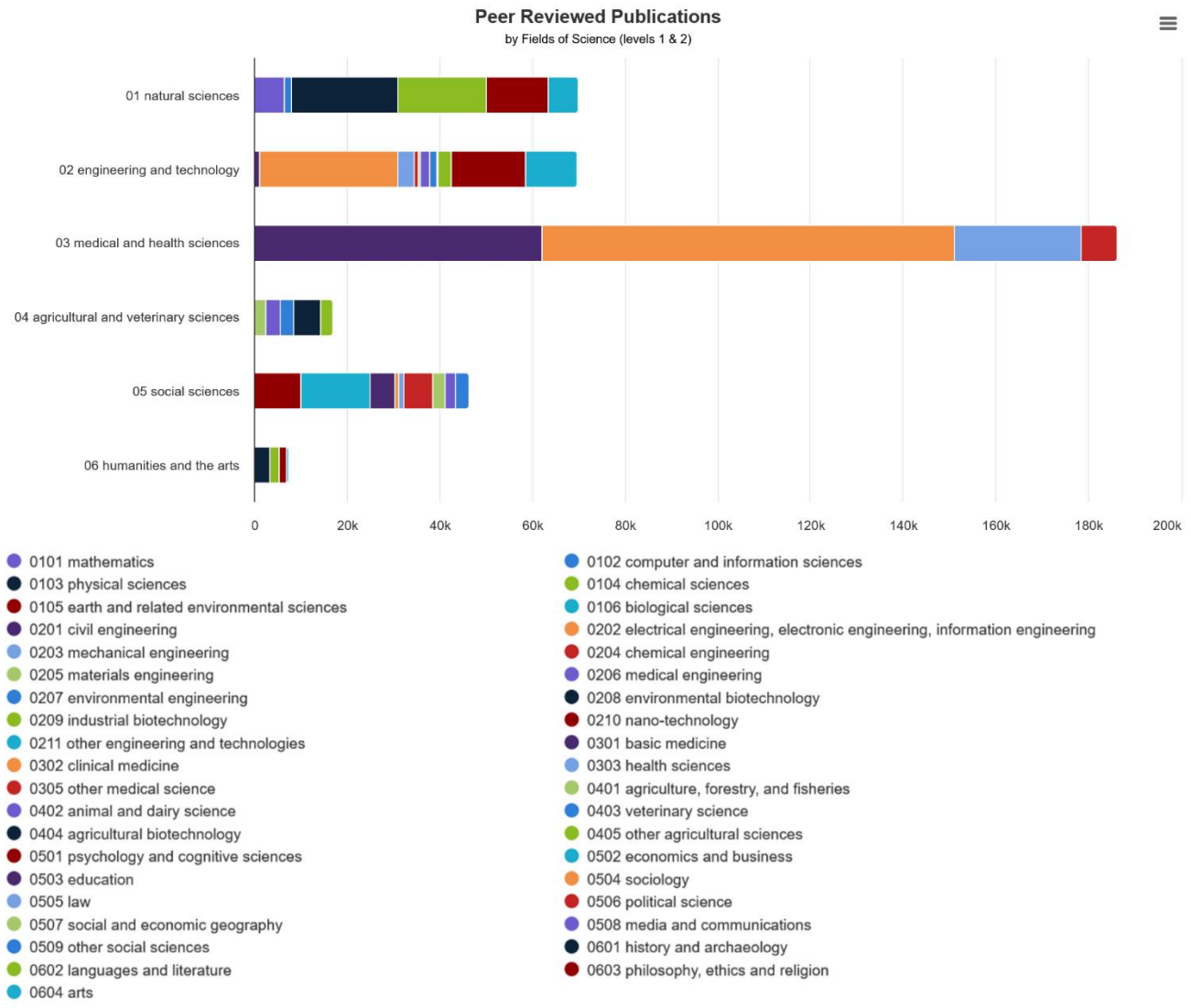

Figure 2: Peer-Reviewed Publications by FoS (levels 1 & 2)

An examination of the types of Irish peer-reviewed publications reveals a clear dominance of articles, which constitute **82.9%** of the total. This is followed by other

literature types (**11%**) and conference objects (**10.3%**), with book chapters also contributing to the scholarly mix[15].

Table 2: Peer-Reviewed Publications by Type (top 7)

	# Irish peer-reviewed publications	% of Irish peer-reviewed publications
Article	*276,460*	*82.9%*
Other literature type	*36,708*	*11%*
Conference object	*34,438*	*10.3%*
Part of book or chapter of book	*22,292*	*6.7%*
Preprint	*12,251*	*3.7%*
Thesis	*6,156*	*1.8%*
Book	*1,156*	*0.3%*

Journals are the primary data source for peer-reviewed publications, followed by Publication and Institutional Repositories[16]. This distribution highlights the continued primacy of journals in academic publishing and a growing engagement with repositories as complementary research dissemination platforms.

[15] In the OpenAIRE Graph, numerous publications appear in multiple instances due to their life cycle and metadata variations. As a result, a single publication might be represented both as an article and as a pre-print, reflecting its different stages of publication. This dual representation can also arise from discrepancies or inaccuracies in metadata reporting. Moreover, while preprints alone are *not* classified as peer-reviewed, the instances we observe here pertain to peer-reviewed publications that are additionally catalogued as preprints.

[16] A publication repository is a digital archive or database designed to store, preserve, and provide access to academic and research publications. These repositories can include a variety of content types such as journal articles, preprints, conference papers, theses, dissertations, and other scholarly works. They serve as important platforms for the dissemination of knowledge, enabling researchers, scholars, and the public to access and share research findings freely. Repositories can be institution-specific (institutional), subject-specific (thematic), or cover a broad range of disciplines (publication repository).

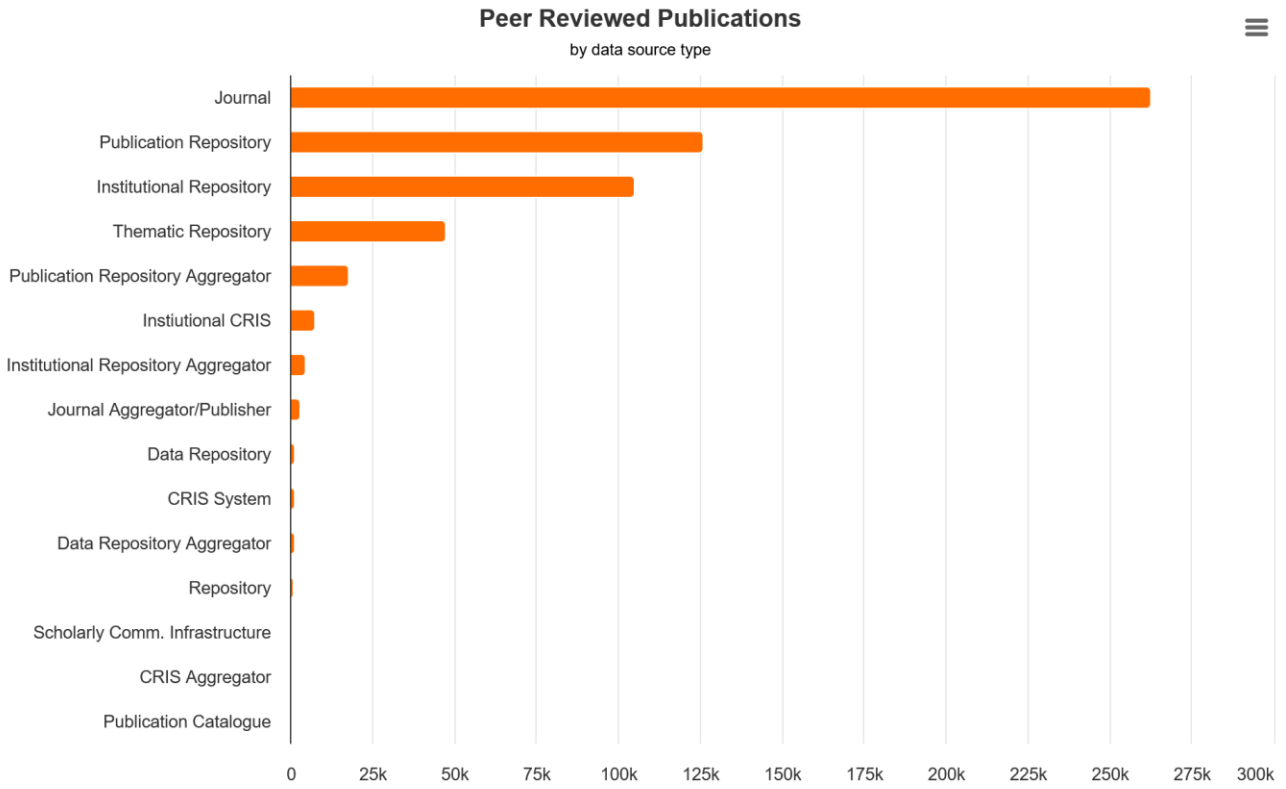

Figure 3: Peer-Reviewed Publications by Data Source Type

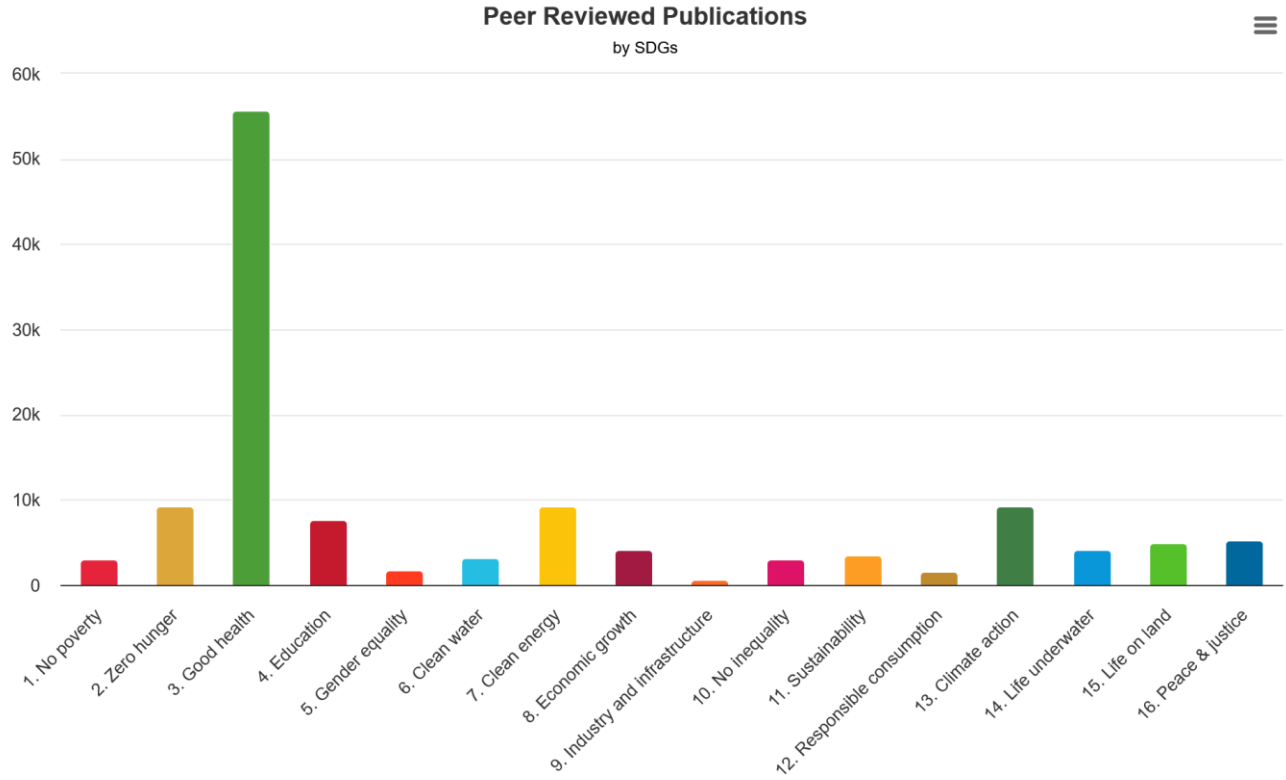

Figure 4: Peer-Reviewed Publications by SDGs

In summary, the analysis of scholarly production sets the stage for understanding the nuances of OA publishing in Ireland. It highlights the strengths and focuses of various research disciplines and sheds light on the evolving landscape of research dissemination, characterized by a blend of traditional and emerging platforms.

2.2 Open & FAIR

In this section, we delve into a set of essential indicators that illuminate the openness and FAIRness of publications within Ireland's scholarly landscape. This analysis forms a crucial foundation for understanding the current state of OA in Ireland and identifying areas where improvements and interventions may be necessary.

2.2.1 Access Rights

Here, we delve deeper into the nature of access rights associated with Irish peer-reviewed publications, a crucial factor in understanding the OA landscape in Ireland. Our analysis is guided by the Budapest Open Access Initiative's definition, which considers only those publications with an associated licence as truly 'open access'. This distinction is vital for the assessment and the overall objectives of the Monitor.

Table 3: Number and Share of Peer-Reviewed Publications by Access Rights

Access Rights[17]	# Irish peer-reviewed publications	% of Irish peer-reviewed publications
Open Access w/ Licence[18]	*129,947*	*39%*
Open Access w/o Licence	*57,933*	*17.4%*
Embargo	*84*	*0.03%*
Restricted	*1497*	*0.4%*
Closed Access	*87,619*	*26.3%*
Not Available[19]	*56,319*	*16.9%*

Examining the access rights data, we find that **39%** of Irish peer-reviewed publications are OA with a licence. This is a significant proportion, but there is room for improvement,

[17] The best available access rights of a publication.

[18] We consider *any* licence found in the metadata of the *Open Access* instance of a publication as open.

[19] "Not Available" refers to the access rights not being specified in the metadata record of the publication.

especially considering that **17.4%** are Open Access without a licence. These unlicenced publications, while accessible, lack the formal licensing that defines full openness and reusability under the Budapest standard. Notably, Closed Access publications still constitute **26.3%** of the total, indicating some persistence of traditional publishing models. The relatively high number of **16.9%** of peer-reviewed publications without any access rights present in their metadata requires further analysis and will be examined in the data evaluation section.

When we break down these access rights by data source type, the reliance on different platforms for disseminating research becomes evident.

- Journals, a primary data source, host a significant number of both licenced and unlicenced Open Access publications. However, the relatively higher number of Closed Access publications in journals suggests traditional models are still prevalent in this space.
- Publication and Institutional Repositories not surprisingly show a strong inclination towards hosting Open Access content, with a considerable number of publications available with licences.[20]
- Thematic Repositories and Publication Repository Aggregators also contribute to the Open Access ecosystem, albeit with lower overall numbers.

This analysis underscores the progress made towards achieving Open Access in Ireland, particularly in institutional and thematic repositories. It also highlights the ongoing challenges in fully transitioning to Open Access, especially within journal publishing. Understanding these dynamics is crucial for shaping future strategies and policies aimed at enhancing the openness and fairness of scholarly communication in Ireland.

[20] The high number of OA publications without a licence may be due to repositories not exposing the licence metadata via OAI-PMH protocol.

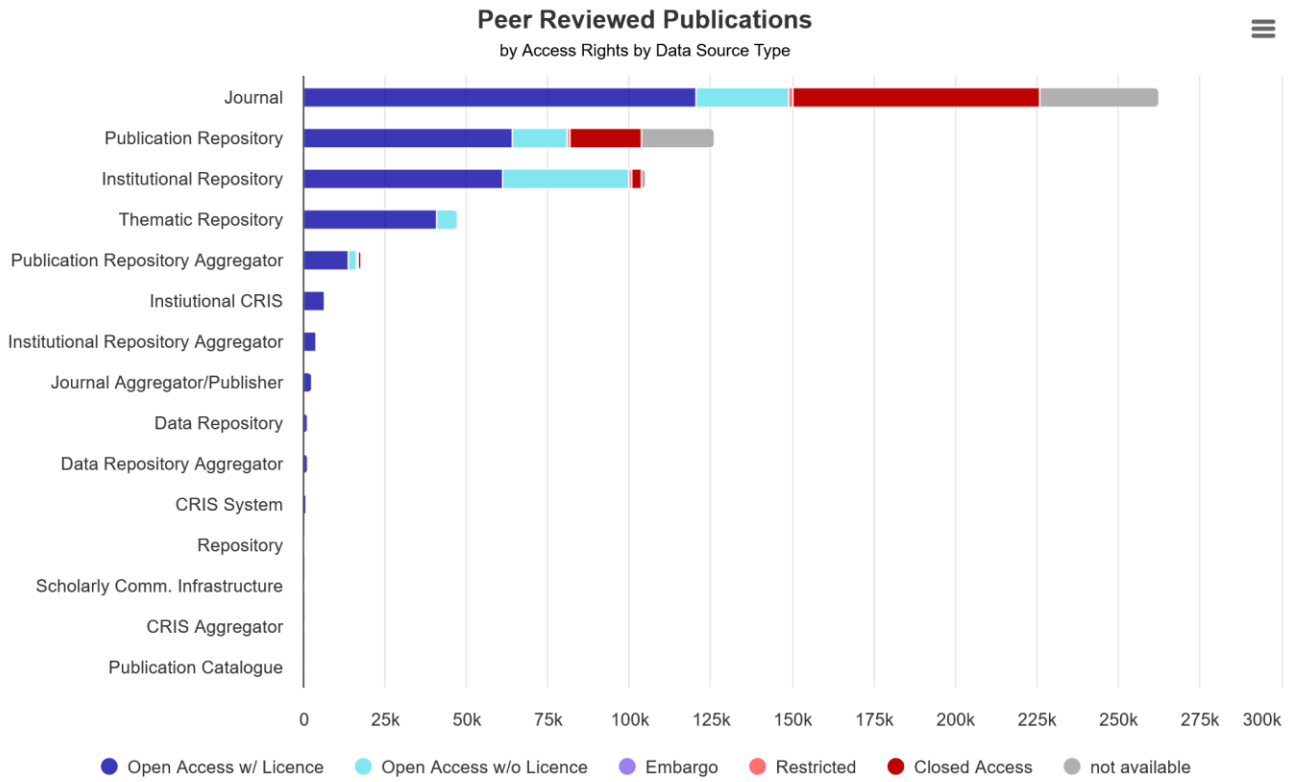

Figure 5: Peer-Reviewed Publications by Access Rights by Data Source Type

To delve deeper into this trend, the following chart provides a chronological breakdown. Over the years, there has been a clear trend towards increasing Open Access with licencing. Starting from a modest base in 2007 of **21.6%** of peer-reviewed publications, there has been a significant year-on-year rise, peaking in recent years, **60.5%** in 2020, **65.4%** in 2021 and **66%** in 2022. This growth trajectory underscores a progressive shift towards more open and transparent research practices in Ireland. Open Access without licencing is decreasing as a share of total Open Access indicating improvement in licensing practices over the years. The trend in Closed Access publications is decreasing steadily over time from **30.1%** in 2007 to **17.1%** in 2021 and **13.7%** in 2022, in absolute and relative terms while still representing a notable portion of the total output.

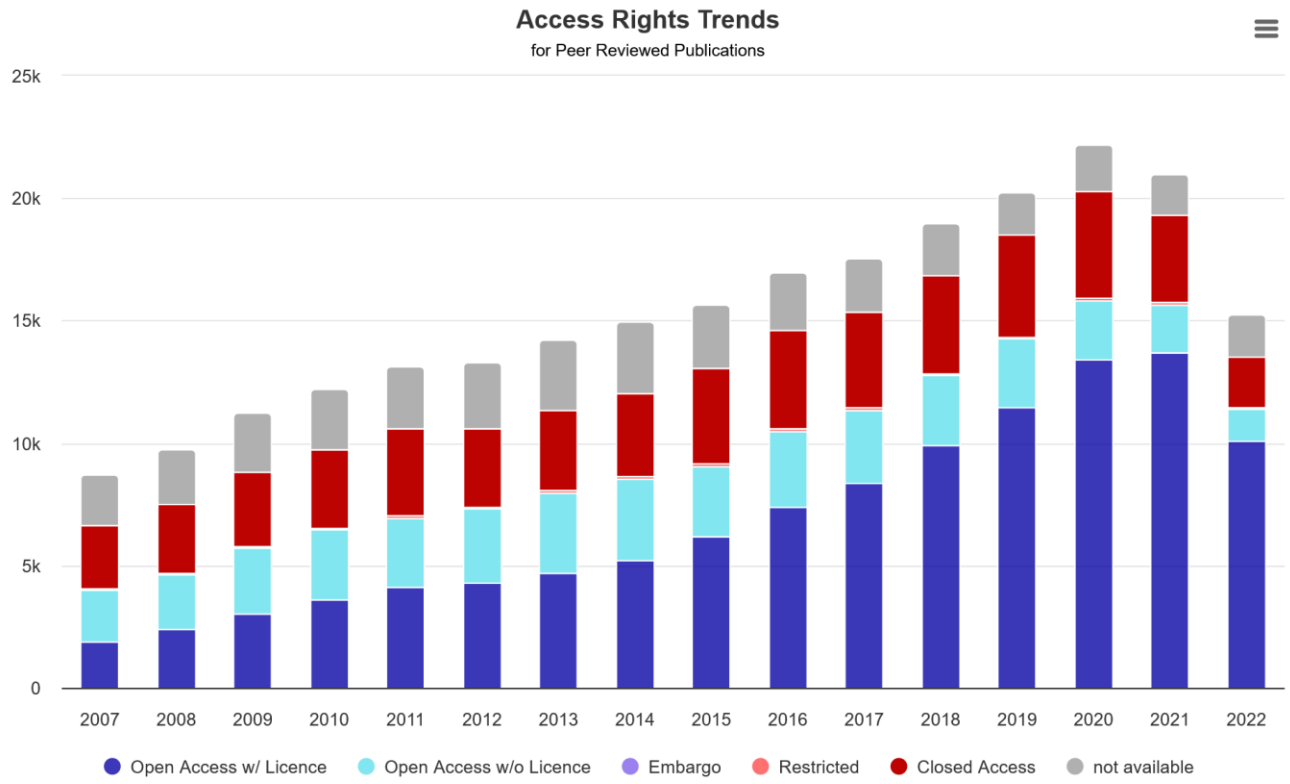

Figure 6: Peer-Reviewed Publications by Access Rights over time

A deeper analysis of access rights across different Fields of Science (FoS) at both the broad Level 1 and the more detailed Level 2 categories reveals insightful trends in Ireland's OA landscape.

Medical and Health Sciences, which have the highest volume of scholarly work, also lead in the proportion of licenced Open Access (OA) publications at **44.3%**, closely followed by the Natural Sciences at **42.7%**. When considering OA publications, both licenced and unlicenced, only Social Sciences, along with, Agricultural and Veterinary Sciences and Humanities and the Arts, fall below the **50%** mark for total OA availability. Notably, Agricultural Sciences and Humanities have the largest percentages of Closed Access publications, at **46.3%** and **36.2%** respectively. This disparity might reflect a more gradual adoption of Open Science principles within these fields or could be influenced by potential gaps in data representation, which might skew our perception of their OA engagement[21].

[21] We note that these patterns refer to the entire collection of peer-reviewed publications and are not limited to only recent years which may paint a different picture. The Monitor https://oamonitor.ireland.openaire.eu/ allows the user to refine the analysis by filtering for specific years as well as other parameters.

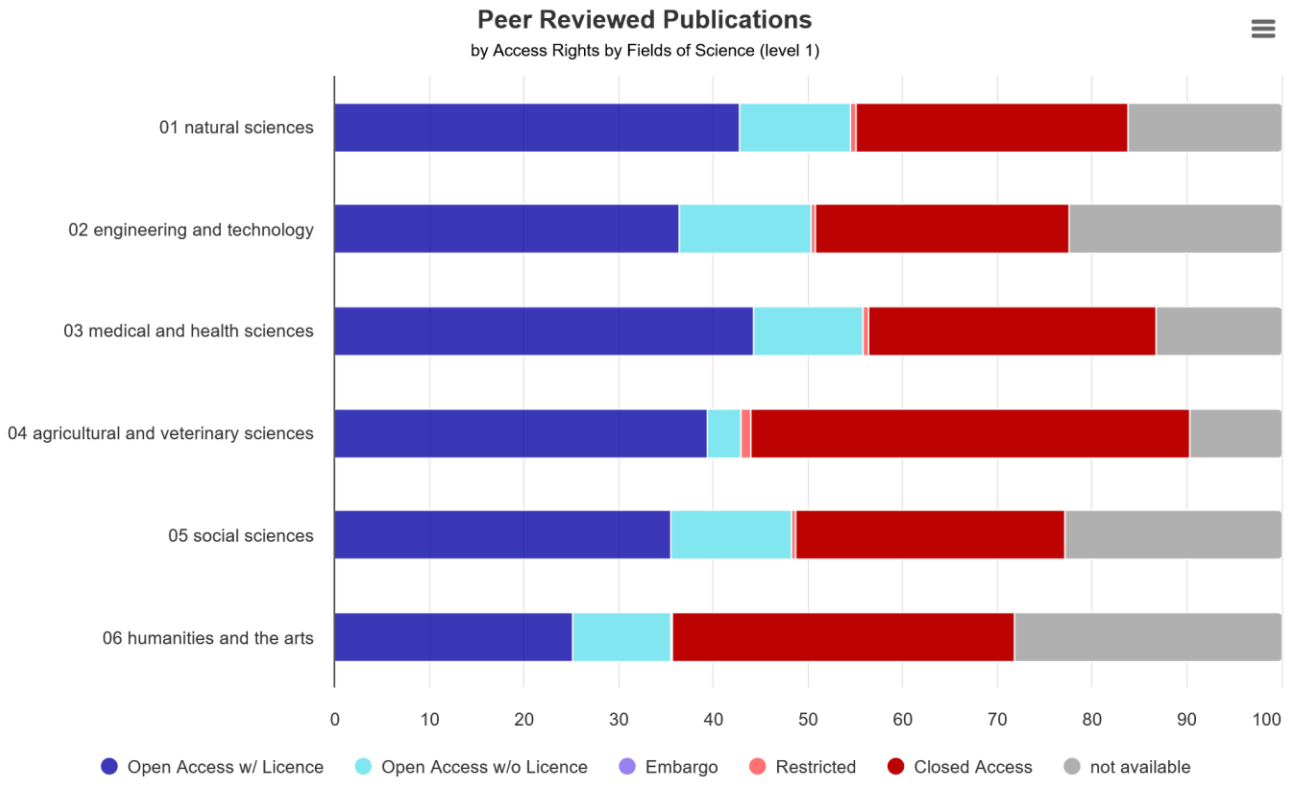

Figure 7: Peer-Reviewed Publications by Access Rights by FoS (level 1)

Drilling down to FoS level 2, we observe that Health Sciences stand out within Medical and Health Sciences with **55.3%** OA with licence, follow by Environmental Engineering in the Engineering and Technology sector with **53.6%** OA with licence, potentially due to its interdisciplinary nature. In Social Sciences, disciplines like Education (**31.2%** OA with licence), Law (**29.8%**), Political Science (**30.3%**) and Social and Economic Geography (**29.7%**) show Open Access footprint below the field average (**35.5%**), indicating diverse adoption rates of Open Science practices.

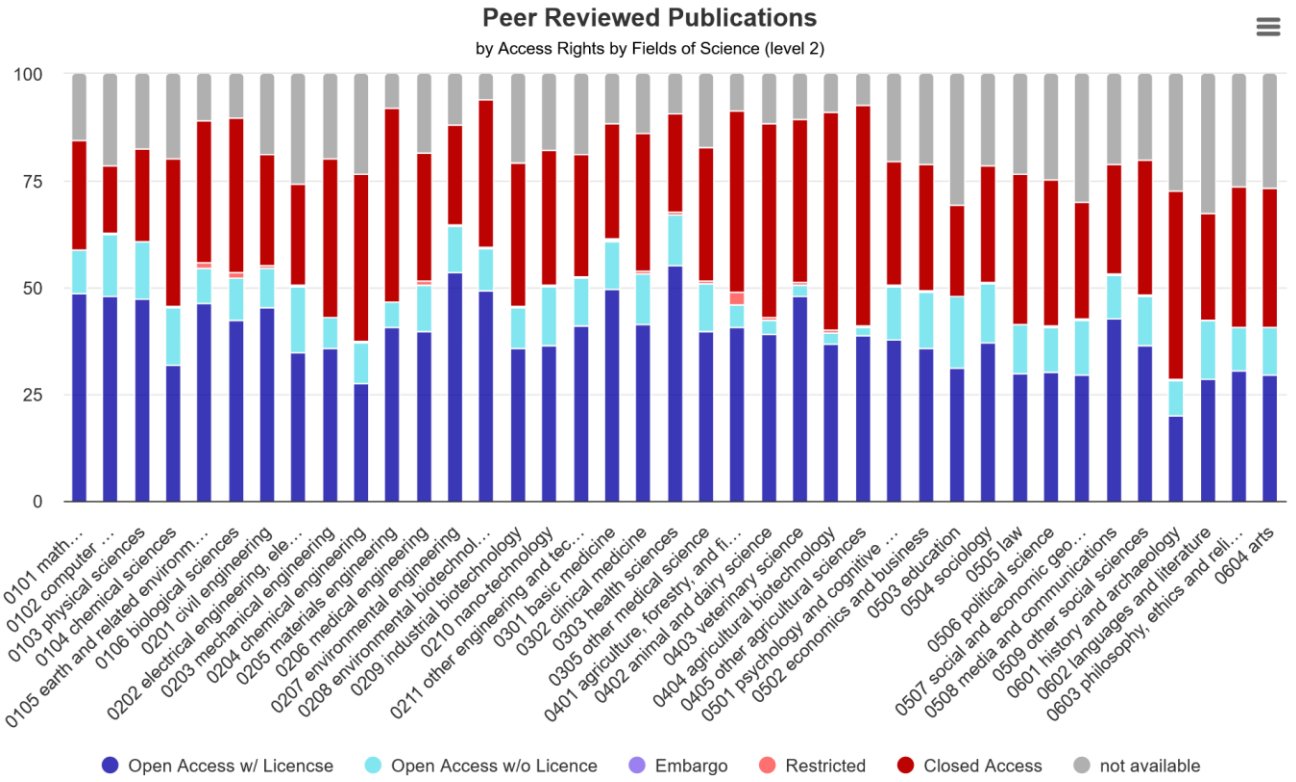

Figure 8: Peer-Reviewed Publications by Access Rights by FoS (level 2)

In terms of publication types, articles demonstrate the highest adoption of OA with a licence (**69.7%** of articles) and the most favourable ratio of licenced to unlicenced OA together with book chapters; although, the latter show the highest shared of Closed Access throughout at **49.6%**. Conference objects and books lag in licensing practices with **23.5%** and **37.4%** of OA without licence respectively. These patterns underscore the varying pace of Open Access integration across different publication formats, highlighting a complex landscape of access and rights management in scholarly communication.

Table 4: Peer-Reviewed Publications by Access Rights by Publication Type

	% OA with licence	% OA w/o licence	% Closed Access
Article	69.7%	6.5%	15%
Conference object	34.5%	23.5%	24.2%
Part of book or chapter of book	26.4%	0.2%	49.6%
Book	19%	37.4%	13.2%

As we have examined the various data sources, temporal changes, and disciplinary fields, the push towards Open Access is multifaceted and unevenly distributed across different sectors of the scholarly landscape. While some fields, such as Medical and Health Sciences, are leading the way in licenced Open Access, others, especially in the Humanities and Agricultural sciences, are lagging, potentially hampered by traditional publication models or a lack of representation in digital repositories.

The trends over time indicate a positive shift towards more open and transparent research practices, with Open Access with licence steadily increasing. However, the persistence of Closed Access publications—although decreasing—highlights the ongoing challenge of fully transitioning to Open Access models.

The discrepancies observed in licensing practices across different types of publications, with articles faring better than conference objects and book chapters, potentially point to systemic issues within the publishing industry that need addressing to foster a more open scholarly communication environment.

2.2.2 Open Access Routes

Repository-mediated vs publisher mediated OA

This section on OA routes provides a critical perspective on how Irish peer-reviewed publications with licences are made available. The distinction between repository-mediated OA (Green OA with a licence) and publisher-mediated OA, which includes both Gold and Hybrid OA, is significant in understanding the dynamics of OA dissemination.

From the data, we observe that publisher-mediated OA is the predominant route, with Gold OA accounting for **38.1%** and Hybrid OA making up **30.8%** of the total licenced OA publications. In comparison, repository-mediated OA represents a smaller share at **6.5%.** This suggests that while repositories play an essential role in OA dissemination, publishers currently have a more significant impact on the OA landscape in Ireland.

Table 5: Repository Mediated vs Publisher Mediated OA Peer-Reviewed Publications

OA route	# Irish OA licenced peer-reviewed publications	% of Irish OA licenced peer-reviewed publications (129,947)
Repository mediated OA (Green OA w/ licence)	*8,507*	*6.5%*
Publisher mediated OA (Gold OA)	*49,579*	*38.1%*
Publisher mediated OA (Hybrid OA)	*40,012*	*30.8%*[22]

Trend data from 2007 to 2022 reveals nuanced shifts in OA publishing routes. Growth is seen across all categories, with repository-mediated OA showing an encouraging rise in its proportion of total OA publications with a licence, from **7.9%** in 2007 to **15%** in 2022. This indicates a strengthened role for repositories in the OA landscape. However, the share of publications represented in both repository and publisher categories (the blue boxes in the figure below) has not seen a consistent increase, peaking in 2019 (**10.2%**). This could point to a missed opportunity for maximizing the reach and impact of scholarly work through dual availability in both repositories and through publishers.

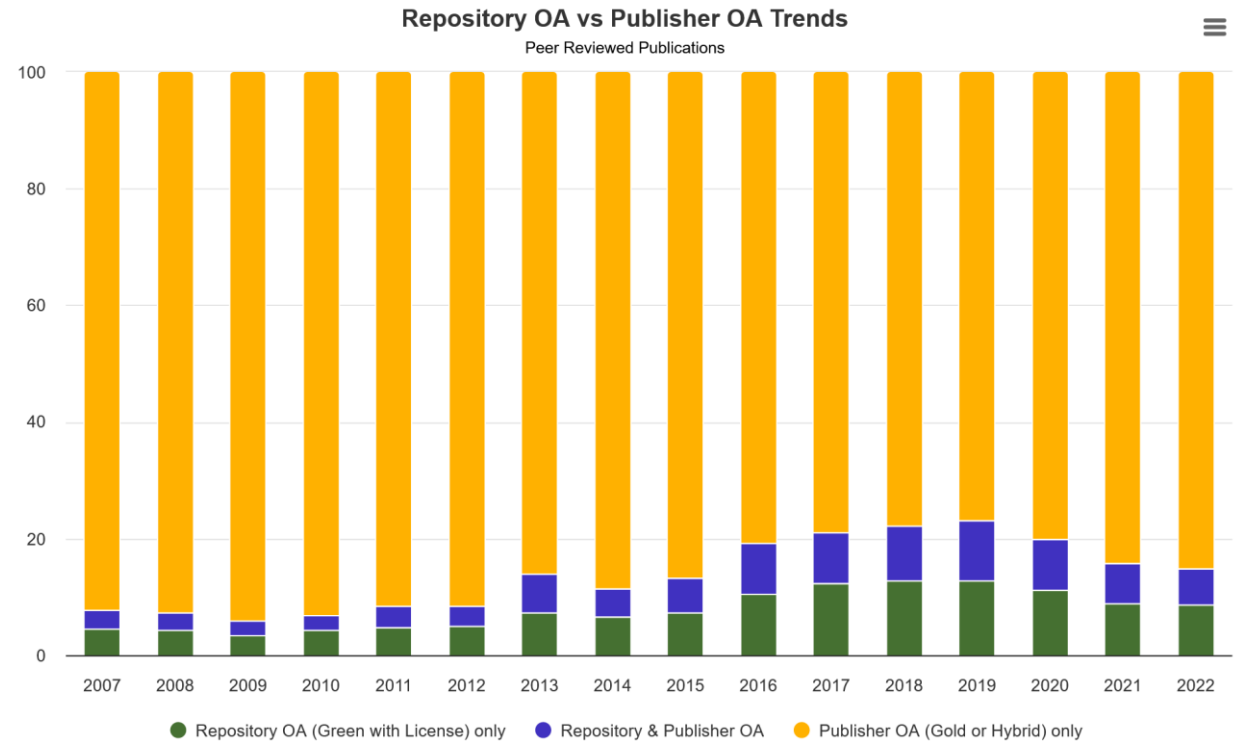

Figure 9: Repository-mediated vs Publisher-mediated Peer-Reviewed Publications Over Time

[22] The 24.6% of OA PR publications with licence missing from the table can *neither* be attributed to a journal *nor* a repository. One common scenario is that they are harvested from an aggregator which does not expose the type of the original data source.

Upon examining the data by Field of Science (FoS), we find that Medical and Health Sciences have the highest proportion of publisher-mediated-*only* OA at **84.7%**, a trend consistent with global priorities for immediate access to medical research. Natural Sciences, Engineering and Technology, and Agricultural and Veterinary Science also show significant engagement with publisher-mediated-*only* OA, with rates of **83.5%**, **80.6%**, and **83.8%** respectively. These rates reflect the disciplines' specific funding and publishing cultures, which favour publisher-mediated routes for OA dissemination. Although absolute numbers offer a different perspective due to variations in research output, these proportions underscore a clear preference for publisher-mediated OA in these fields.

In contrast, Social Sciences, along with Humanities and the Arts, exhibit a more balanced mix of publisher-mediated and repository-mediated OA. This distribution indicates a broader array of OA practices within these disciplines, with Humanities particularly leaning relatively more towards repository-mediated OA. This preference, despite a lower overall publication count, suggests a strategic adaptation to OA that aligns with the unique characteristics and dissemination needs of these fields. It highlights the diversity of OA adoption strategies across disciplines, underscoring the comprehensive nature of OA as it encompasses both publisher and repository pathways to maximize research accessibility and impact.

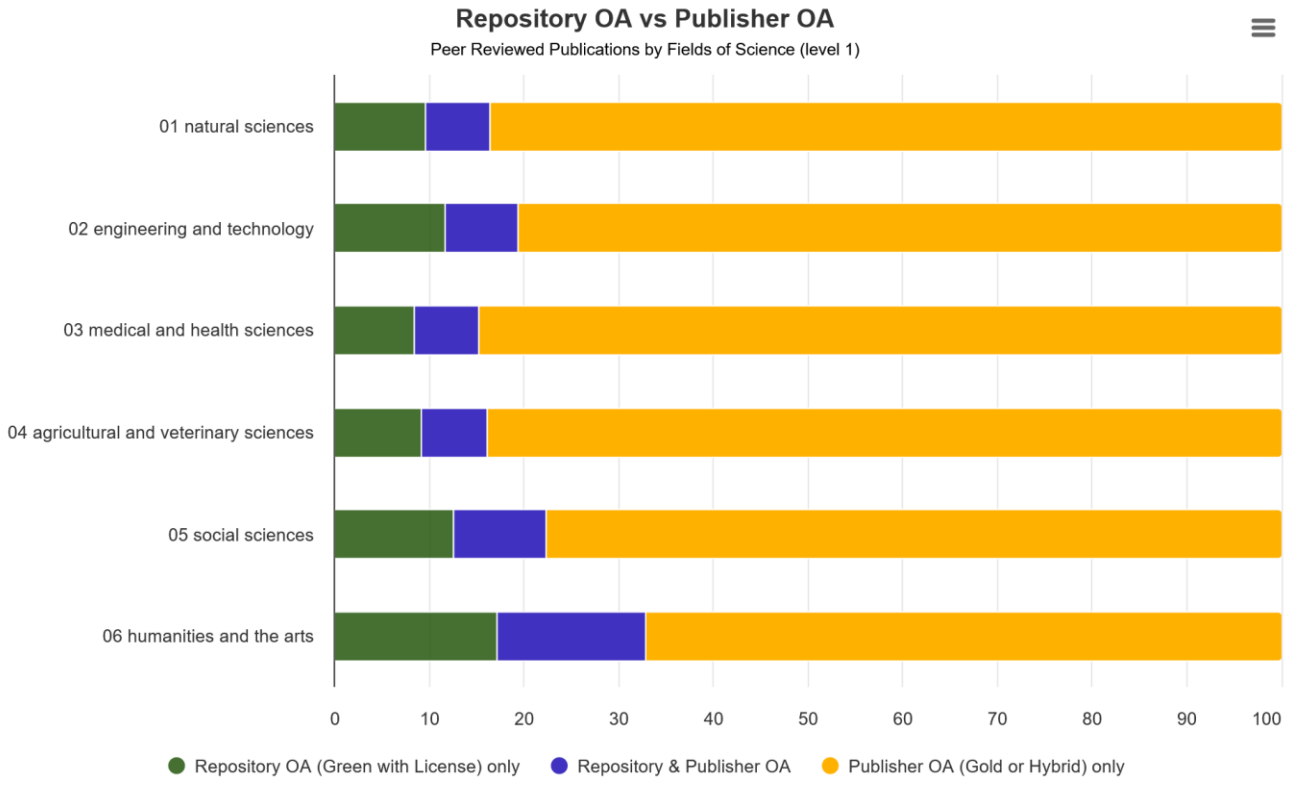

Figure 10: Repository-mediated vs Publisher-mediated OA Peer-Reviewed Publications by FoS (level 1)

Lastly, we observe a discernible trend in the distribution of Gold versus Hybrid OA publications. The data indicates a steady increase in the proportion of Hybrid OA, suggesting a shift in journal policies from traditional subscription models to hybrid models. This transition, potentially driven by Transformative Journals or Agreements, becomes particularly noticeable from 2020 onwards. By 2021 and 2022, Hybrid OA surpass for the first time Gold OA publications, reflecting a significant change in the publishing landscape.

This trend towards Hybrid OA may be a response to evolving OA policies and the growing adoption of Plan S principles. The subsequent section on Plan S and Transformative Agreements will delve deeper into this aspect, offering a clearer understanding of the forces shaping these trends.

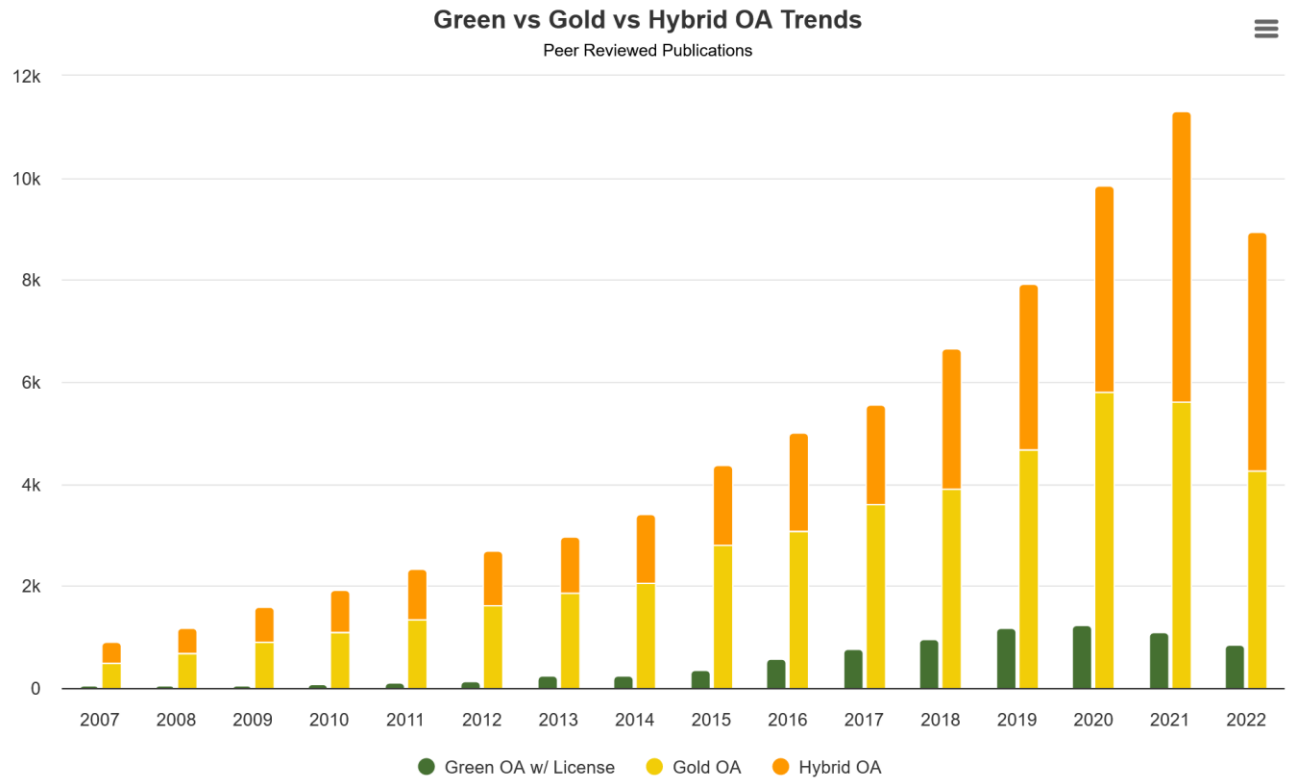

Figure 11: Gold OA vs Hybrid OA vs Green OA with Licence over Time

Table 6: Gold OA vs Hybrid OA vs Green OA with Licence over Time

Year	Green OA w/ Licence	Gold OA	Hybrid OA
2007	45	498	422
2008	55	676	501

2009	56	906	682
2010	87	1105	820
2011	116	1340	991
2012	142	1613	1090
2013	239	1876	1095
2014	242	2064	1361
2015	350	2815	1554
2016	589	3090	1920
2017	784	3608	1942
2018	974	3908	2748
2019	1178	4689	3231
2020	1235	5797	4049
2021	1102	5610	5709
2022	841	4273	4664

In summary, the shifting trends in OA publishing in Ireland reflect a landscape that is actively adapting to new norms. The rise of Hybrid OA, complemented by the steady prevalence of Gold OA, showcases a publishing sphere that is adjusting to evolving mandates and the diverse requirements of researchers. It is essential to keep an eye on how these changes influence the accessibility and utilization of research, with a focus on fully leveraging the advantages of OA for the wider scholarly community.

2.2.3 Immediate vs Embargo Open Access

In our exploration of OA practices, we focus on the timing of access — whether publications are made immediately available (immediate OA) or are initially embargoed. We note the following.

- Gold and Hybrid OA publications are available immediately, providing instant access upon publication.
- In assessing the immediacy of Green Open Access (OA) publications with a licence, a key challenge arises from the ambiguity surrounding the "publication date" in repository metadata. Some repositories use this term to denote the date an article was deposited, while others may use it to indicate the actual publication date. Without a standardized identifier for the deposition date, it is difficult to

ascertain whether access is truly immediate. This lack of clarity complicates our understanding of how swiftly research becomes available through Green OA[23].

- Embargoed OA articles present a unique challenge in tracking trends. Typically, they are only marked as 'embargoed' in the metadata until the embargo period expires. This categorization means that understanding the patterns of embargoed publications over time is complex, especially considering the currently low number of publications (only **84**) tagged explicitly as 'embargoed'.

These observations suggest a limitation in our ability to comprehensively assess the trends and variations between immediate and embargoed OA practices. The low tagging rate of embargoed articles hints at potential underreporting, which could skew the understanding of how often and effectively the embargoed route is being utilized in OA publishing[24].

Table 7: Immediate vs. Embargo Irish OA Peer-Reviewed Publications

	OA route	# Irish OA peer-reviewed publications w/ licence	% of Irish OA peer-reviewed publications w/ licence
Immediate OA	Green OA with licence	8,507	6.5%
	Gold OA	49,579	38.2%
	Hybrid OA	40,012	30.8%
Embargo OA		84	0.06%

2.2.4 Unrealised Open Access

Analysing Unrealized OA is crucial as it uncovers the gaps between potential and actual open accessibility of scholarly outputs. Unrealized OA encompasses three types of peer-reviewed publications:

1. **Closed Access publications:** Represent traditional publishing models.

[23] Many mandates require the deposit of either the Author Accepted Manuscript (AAM) or the Version of Record (VoR), yet this specification is not uniformly indicated in repository metadata, adding another layer of complexity to compliance and accessibility evaluations. In this report, when we consider Green OA with a licence, we do not impose restrictions on versioning due to the absence of detailed version information. However, it is important to bear in mind the potential variations in version availability when interpreting OA data.

[24] For more details on metadata coverage and quality, see Section 4.

2. **Green Publications without a licence:** Fail the Budapest OA Initiative's definition of OA due to the absence of a licence, reflecting a gap in metadata practices.
3. **Bronze publications:** These are accessible but unlicenced, again not meeting the full criteria for OA.

The large volume of Green publications without a licence (**72,320**, **21.7%** of Irish peer-reviewed publications) compared to those with a licence (**8,507**, **2.5%** of Irish peer-reviewed publications) suggests significant community practices impacting repository uptake. If licensing discrepancies are addressed, the landscape of repository vs. publisher-mediated OA discussed in Section 2.2.2 might appear considerably different. Ensuring automatic licensing in repositories, if not already in practice, could be a transformative step[25].

Table 8: Unrealised OA in Peer-Reviewed Publications

Unrealised OA	# Irish peer-reviewed publications	% of Irish peer-reviewed publications
Closed Access	*87,619*	*26.3%*
Green without Licence	*72,320*	*21.7%*
Bronze	*122,784*	*36.8%*

Trend analysis shows a steady increase in Bronze and Green publications without a licence, indicating a growing but Unrealised potential for OA. Closed Access publications, on the other hand, show a plateauing trend.

[25] The lack of licensing in repositories may be due to some repositories not exposing the licence metadata via OAI-PMH protocol although it is available.

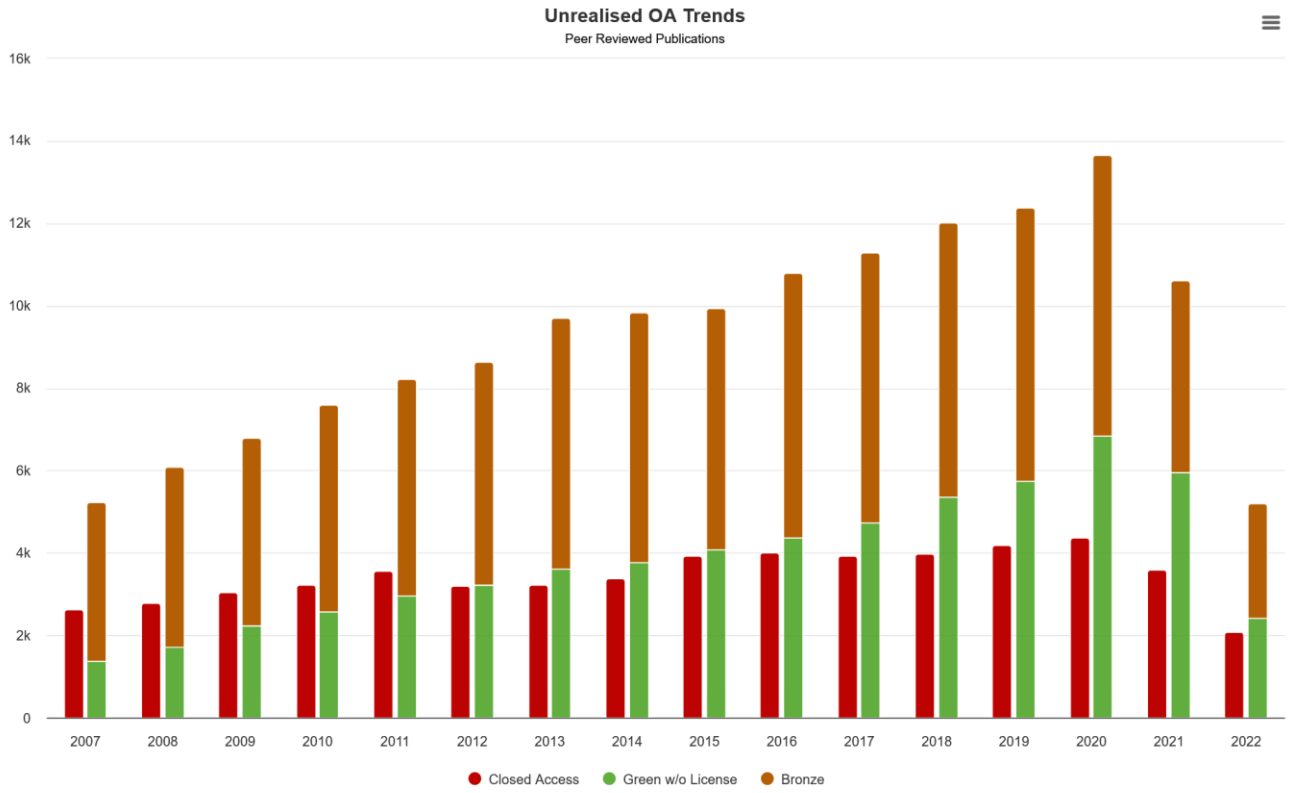

Figure 12: Unrealised OA Peer-Reviewed Publications Over Time

Turning to specific disciplines (FoS level 1), Humanities and the Arts, along with Social Sciences, exhibit a high ratio of Bronze to Green OA without licence. This prevalence might indicate a preference for accessible but unlicenced forms of publishing in these fields, perhaps due to specific disciplinary norms or publishing agreements.

The Medical and Health Sciences, despite their high research output, also have substantial publications in Closed Access (**40,967**), Green without licence (**34,869**), and Bronze (**67,943**) categories. This suggests the potential for significant shifts towards more openness. The lower ratio of Closed Access to Green without licence plus Bronze in these sciences indicates their closer alignment with OA principles, pending better licensing.

Within the Unrealised OA class of publications, in the FoS level 2 categories, various Engineering and Technology fields notably lean towards Closed Access, with Materials Engineering at the forefront, where **59.7%** of its output is not accessible. This reflects the sector's inclination towards protecting potentially commercially viable research. For Green OA without a licence, both Computer and Information Science and Physical

Sciences share a significant portion of Unrealised OA at **34.8%**, suggesting challenges in licensing and rapid technological advancement that hinder full openness.

Environmental Engineering emerges distinctly in Bronze OA, constituting **52.3%** of its publications in this category. This prominence may be due to the field's urgent need to disseminate research broadly to address environmental challenges, utilizing Bronze OA as a temporary solution for wider access within the constraints of traditional publishing models.

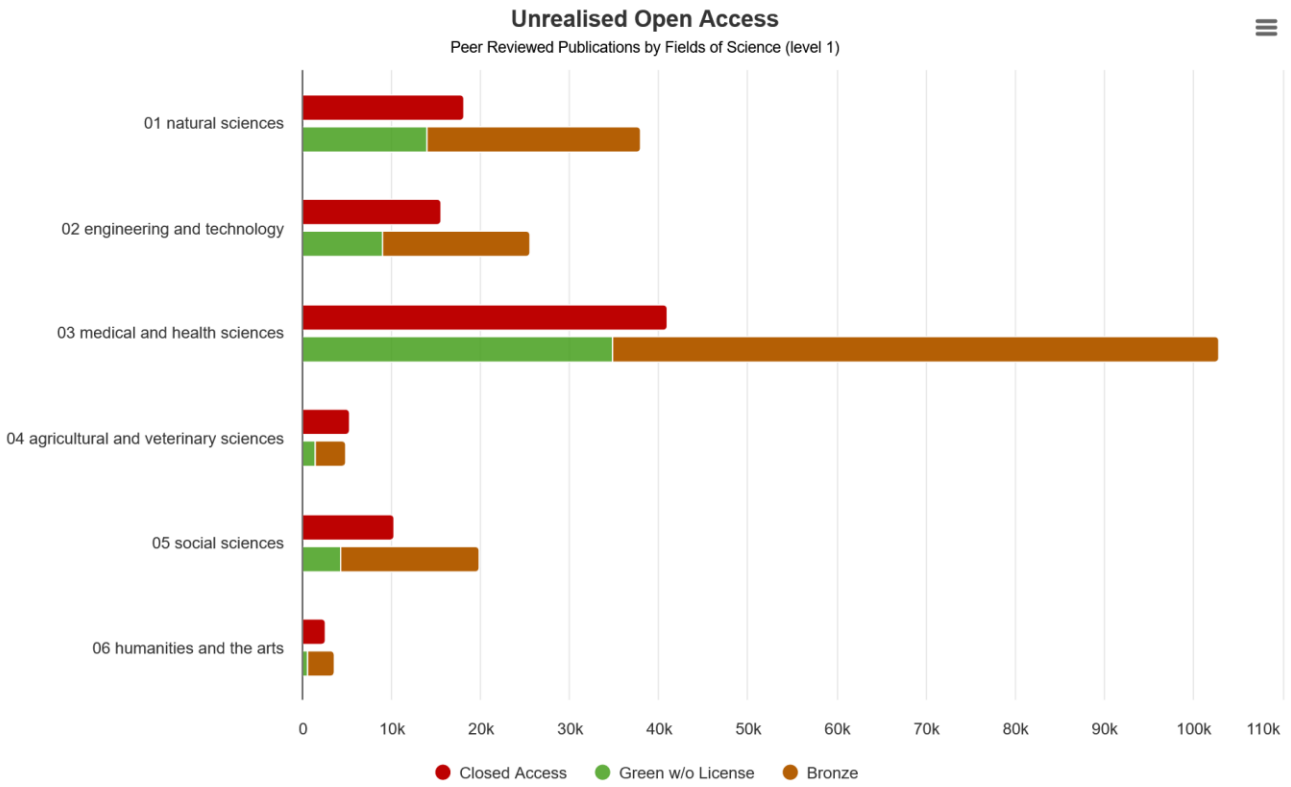

Figure 13: Unrealised Open Access Peer-Reviewed Publications by FoS (level 1)

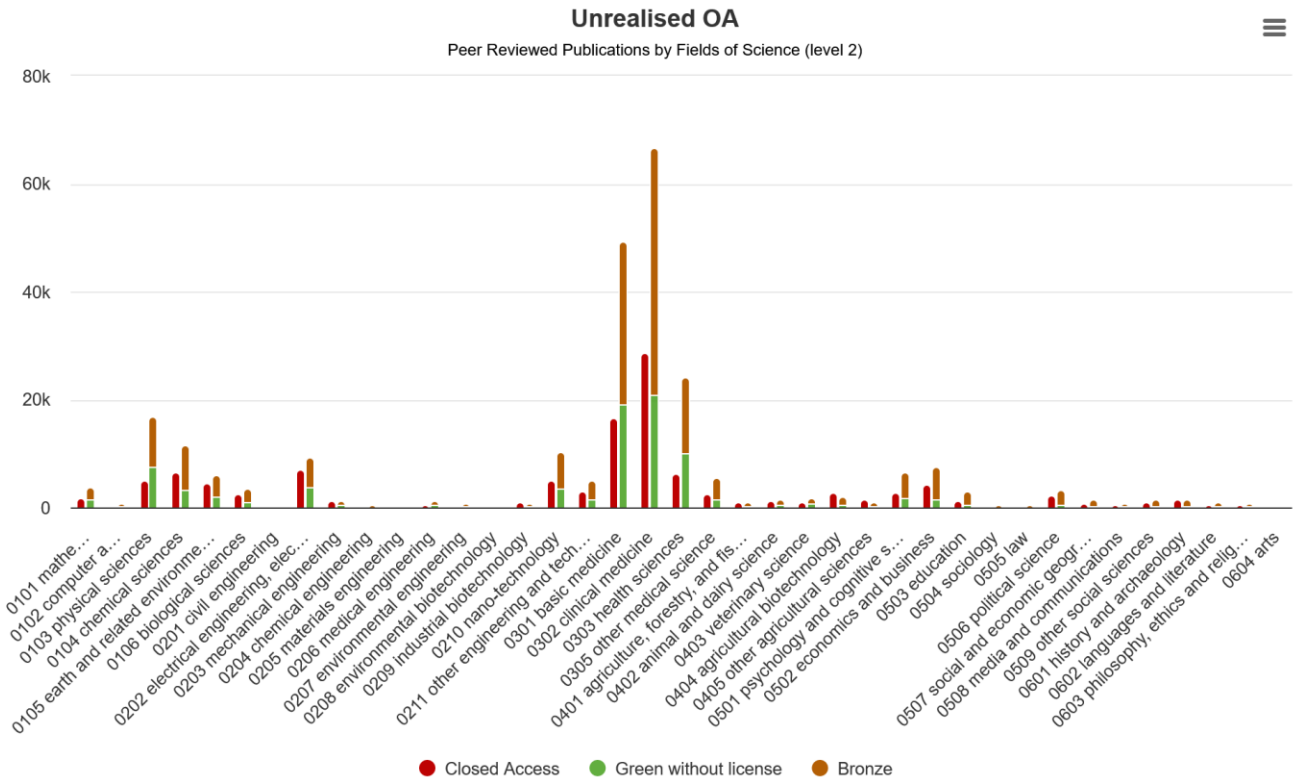

Figure 14: Unrealised Open Access Peer-Reviewed Publications by FoS (level 2)

Encouraging proper licensing in Green OA, can greatly enhance the accessibility and usability of these publications. While Bronze OA does not necessarily signify a weakness in practices, it reflects a semi-open state that could benefit from more explicit licensing to fully align with OA standards.

Given the disparities across disciplines, a tailored approach to OA policy and advocacy seems more effective than a universal strategy. Understanding the unique needs and characteristics of each field is key to successfully transitioning them towards more open and accessible research models.

2.2.5 FAIR

The FAIR principles, which stand for Findable, Accessible, Interoperable, and Reusable, originally designed to guide data management for enhanced usability and access, have been adapted from their dataset-centric origins to apply to scholarly publications. In this adaptation, we highlight key elements for each principle to illustrate how Irish peer-reviewed publications adhere to FAIR standards, enhancing their visibility, accessibility, and utility in academic research. Each principle correlates with specific aspects of the analysis we present below.

Findable

- **Persistent Identifiers (PIDs):** PIDs, such as DOIs, are crucial for making scholarly artifacts findable. They provide unique and persistent references to digital objects, making it easier for both humans and machines to locate the associated data or publications.
- **Abstracts:** The presence of abstracts enhances the findability of publications. Abstracts provide a concise summary of the content, aiding in the discovery process during searches.

Accessible

- **Licences:** The type of licence assigned to a publication affects its accessibility. Open licences like Creative Commons (CC) facilitate broader access by defining clear usage rights. This clarity helps users understand how they can legally access and use the data.
- **Accessibility of full-text:** This refers to the ability of users to retrieve the full text of a publication in a machine-readable format.

Interoperable

- **Use of PIDs:** Beyond findability, PIDs also contribute to interoperability. They help link data across different platforms and databases, facilitating the integration of data for comprehensive analysis and use.
- **File format of full-text:** This refers to the machine-readability of the full-text of the publication.

Reusable

- **Licences:** Again, licences play a critical role in the reusability of data. Licences that clearly articulate usage rights and restrictions enable users to understand how they can reuse and repurpose data.
- **ORCID iDs:** Associating ORCID iDs with publications increases their reusability[26]. These IDs unambiguously link researchers to their work, aiding in attribution and ensuring that the right contributors get credit for their work.

[26] Enhanced further by also including other PIDs such as ROARs and RAiDs.

Each of these elements contributes to making research outputs more FAIR - enhancing their value, ensuring their long-term utility, and promoting more efficient and effective science[27].

LICENCES

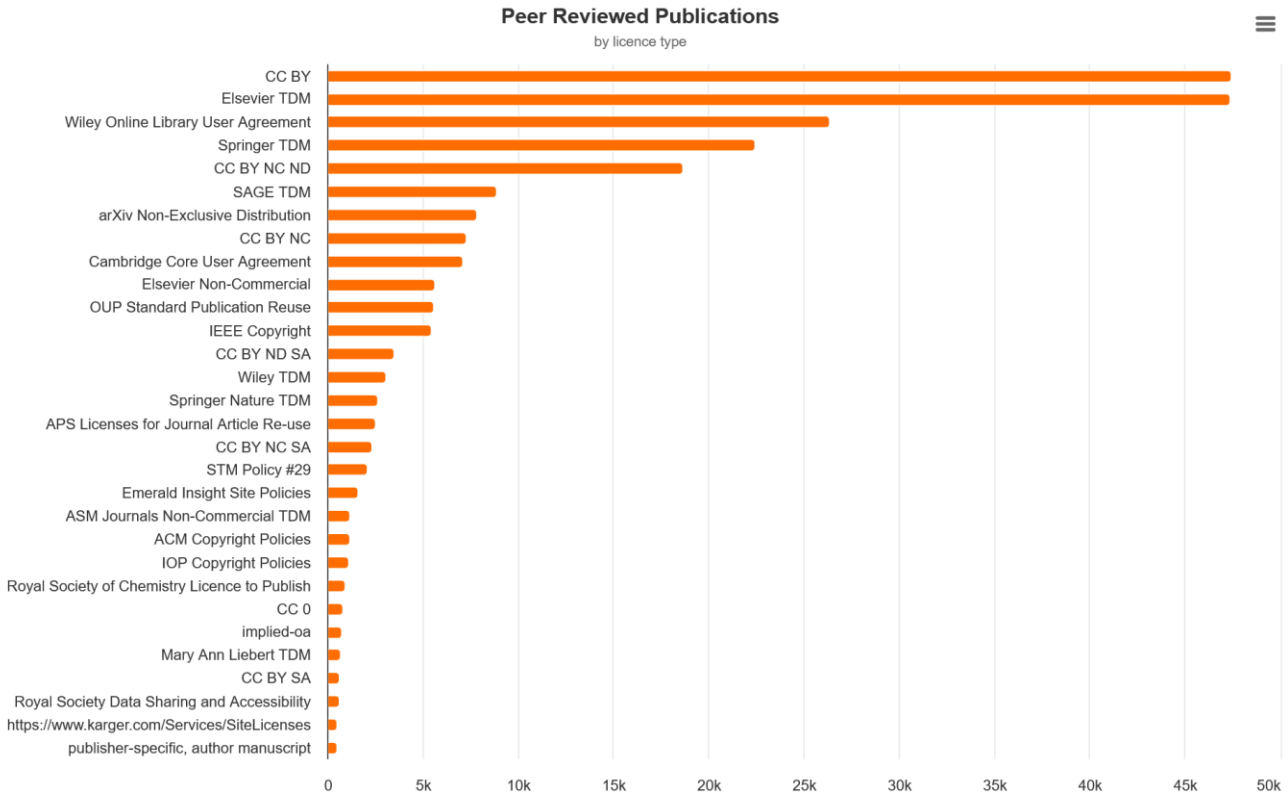

Figure 15: Peer-Reviewed Publications by Licence

The type of licence a publication carries significantly influences its reuse and distribution. Here, we focus on the prevalence and variety of licences used, which offer insights into how accessible and reusable these publications are under the FAIR Principles.

A substantial portion of publications is under restrictive licences from major publishers like Elsevier TDM (**47,384 peer-reviewed publications**), Wiley Online Library User Agreement (**26,330**), and Springer TDM (**22,432**). These licences often limit the use of the content,

[27] Overall metadata completeness of publications and other research products can be achieved by following the latest version of the OpenAIRE Guidelines (https://guidelines.openaire.eu/en/latest/). Institutional data sources registered in OpenAIRE PROVIDE (https://provide.openaire.eu/) can use the Metadata Validator (https://catalogue.openaire.eu/service/openaire.validator) to directly measure the metadata completeness of their institutional outputs.

especially regarding text and data mining (TDM), which can restrict the full exploitation of these publications for advanced research purposes[28].

In contrast, Creative Commons (CC) licences, which are more open and flexible, appear significantly. CC BY (**47,453**) is the most prevalent, followed by CC BY NC ND (**18,665**) and CC BY NC (**7,287**). The presence of these licences indicates a commitment to Open Science, enhancing the accessibility and reusability of research outputs[29].

Table 9: Irish Peer-Reviewed Publications with and without a CC licence

	# Irish peer-reviewed publications	***% of Irish peer-reviewed publications***
with CC licence	*76,739*	*22.5%*
without CC licence	*256,665*	*77.5%*

The licensing landscape in Irish scholarly publications thus presents a mixed picture. While restrictive licences from publishers are still common, the increasing adoption of CC licences, from **6%** in publications published in **2007** to **56.4%** in **2022**, points to a significant gradual shift towards more open and transparent research practices. Understanding these licensing trends is crucial in shaping future strategies and policies aimed at promoting more open and FAIR-compliant research outputs.

[28] The licences from major publishers like Elsevier, Wiley, and Springer, as well as those from other entities listed, generally tend to be more restrictive compared to Creative Commons licences. They often limit the use for text and data mining (TDM) and restrict commercial use and redistribution. They also vary significantly in permissiveness.

[29] Other licences can be considered Open upon examination of the legal agreement. Those identified by Unpaywall as Open were not discussed herein as they are not significantly present in the Monitor dataset https://support.unpaywall.org/support/solutions/articles/44002063718-what-is-an-oa-licence-.

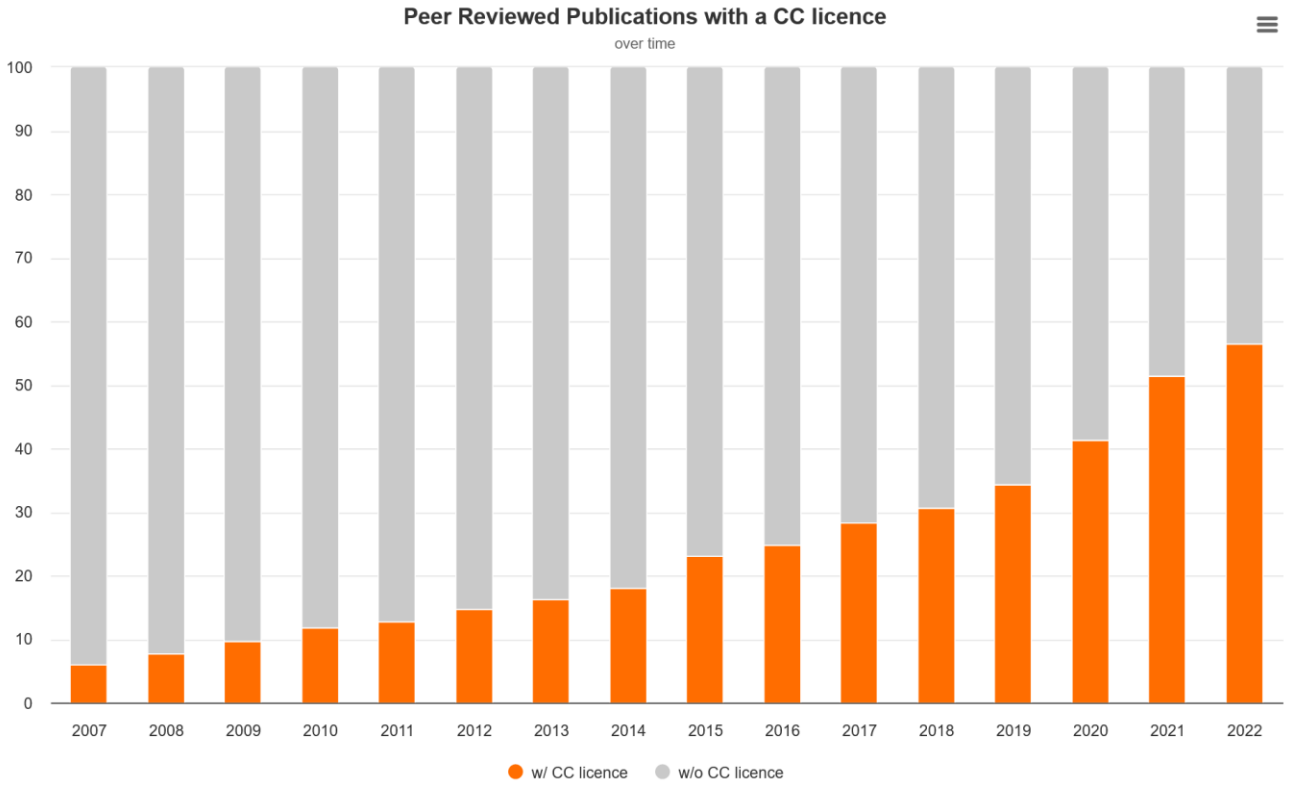

Figure 16: Peer-Reviewed Publications with a CC Licence over Time

In FoS level 1 categories, we note that the variations across CC licensing uptake are not as high as for other metadata elements, they are all around **20%**, with Medical and Health Sciences leading at **25%**, and Humanities and the Arts lagging with **14.3%**. This is not as surprising given the higher share of Closed Access publications in the latter. Looking into FoS level 2 however, provides a much more varied picture which helps identify areas of improvement within disciplines, such as Mathematics (**14.8%**) in Natural Sciences and Chemical Engineering (**11.4%**) in Engineering and Technology.

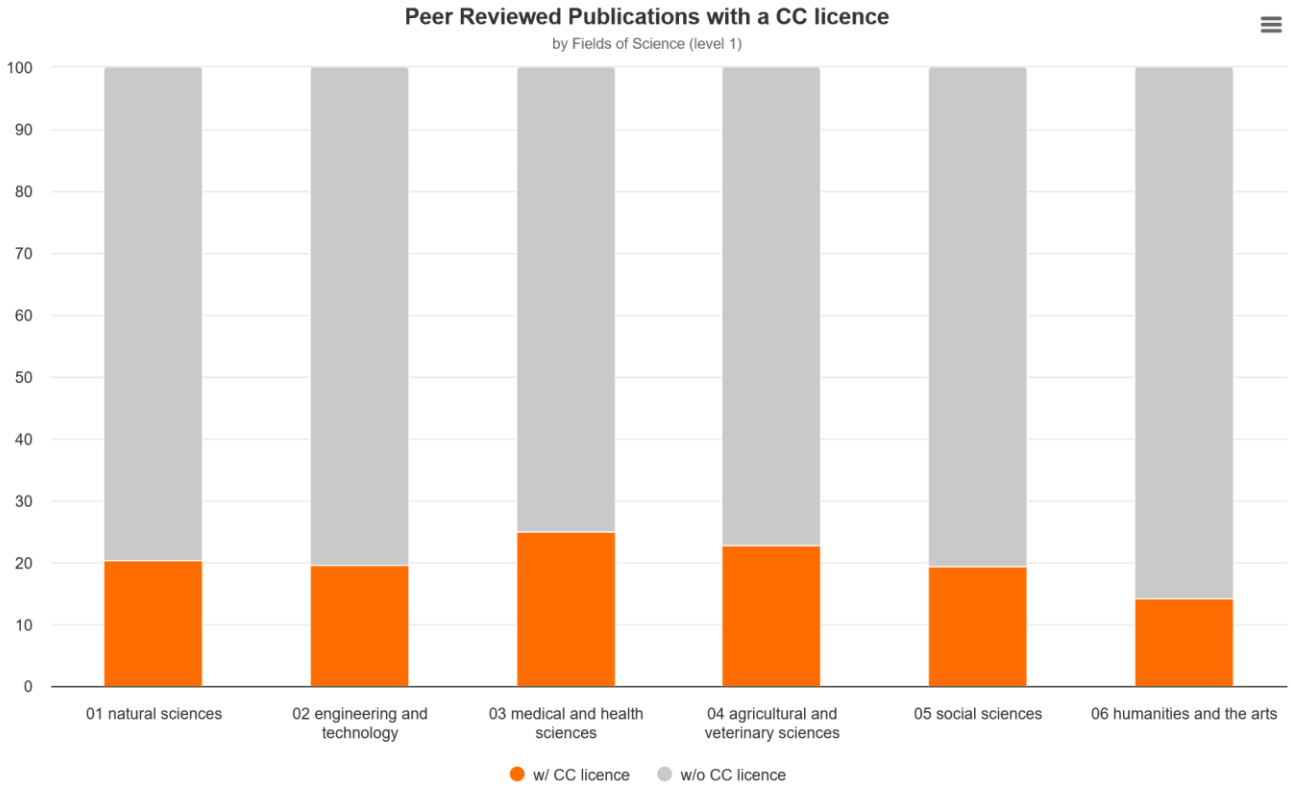

Figure 17: Peer-Reviewed Publications with a CC Licence by FoS (level 1)

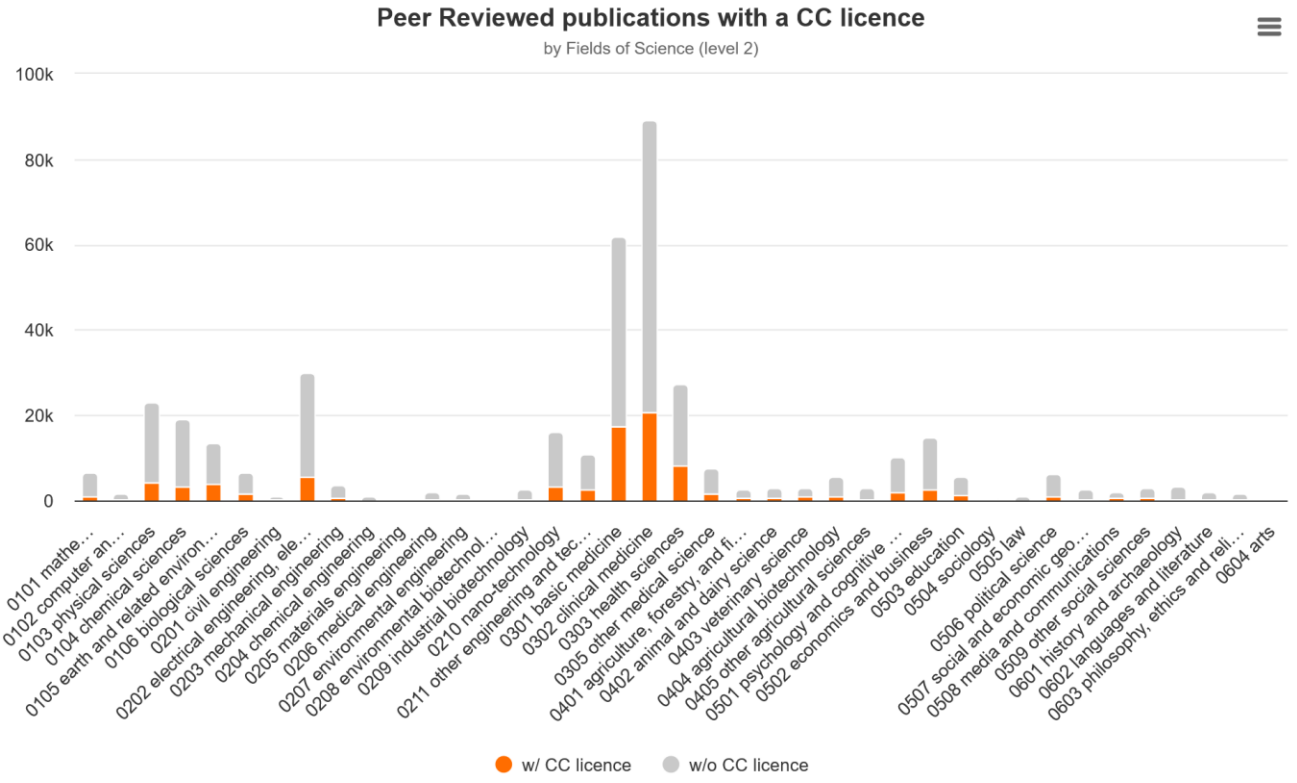

Figure 18: Irish Peer- Reviewed Publications with a CC Licence by FoS (level 2)

Turning to data sources, Thematic Repositories lead the way with a remarkable **56.9%** of their peer-reviewed content under CC licences. This high percentage underscores the commitment of thematic repositories to OA principles, ensuring that research is widely available and legally reusable. Institutional Repositories, with a **35%** rate of CC licence availability, demonstrate a significant portion of content adhering to OA, with potential for growth. Increasing the adoption of CC licences in these repositories could have a broad impact on the openness of scholarly communication across various fields.

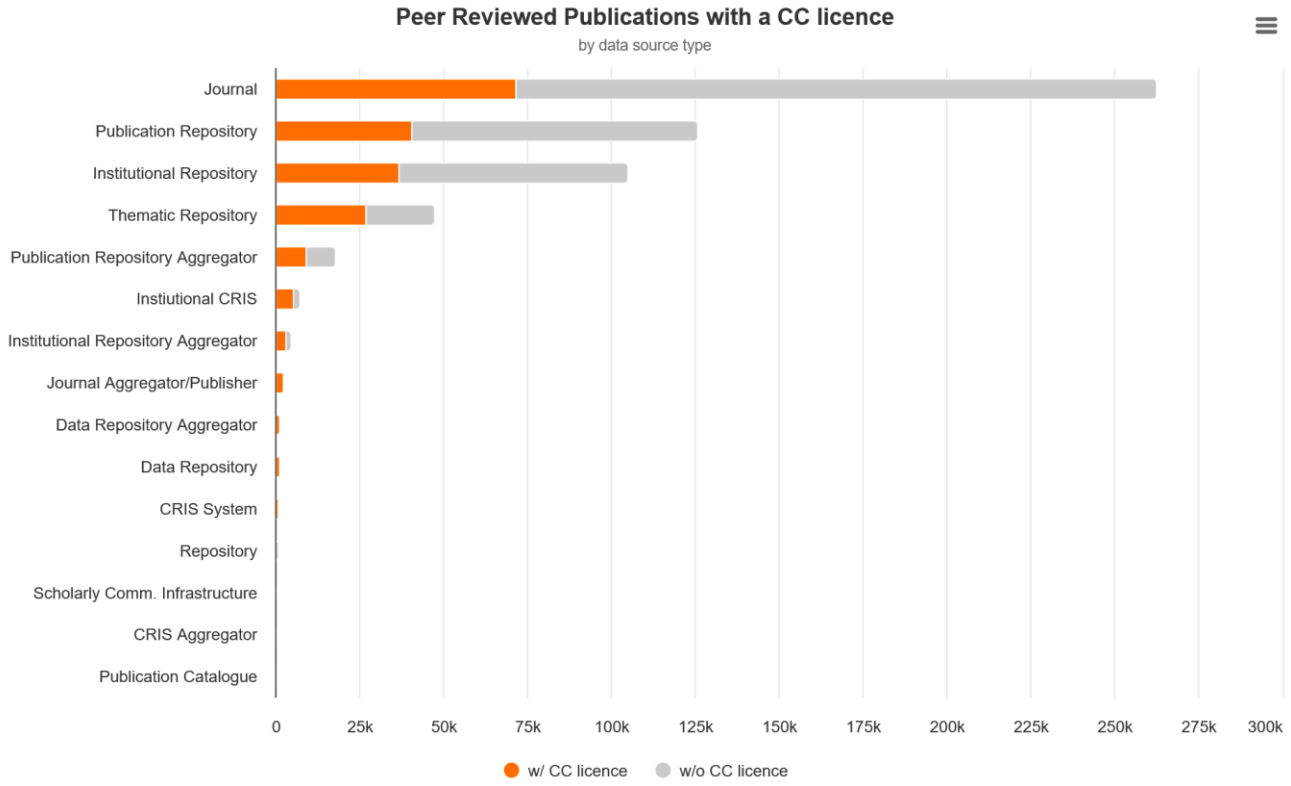

Figure 19: Peer-Reviewed Publications with a CC Licence by Data Source Type

PERSISTENT IDENTIFIERS (PIDs)

Persistent identifiers (PIDs) play a crucial role in the Open Access landscape as they enable the stable, long-term identification and accessibility of digital objects like publications. Their significance lies in their ability to facilitate the discovery, citation, and interlinking of research outputs, thus enhancing the integrity and efficiency of scholarly communication.

In the context of Irish peer-reviewed publications, there are **285,268** (**85.6%** of Irish peer-reviewed publications) with a PID. Further analysis of PID types reveals an interesting trend: alongside the use of Digital Object Identifiers (DOIs), there is an increasing employment of other identifier systems. The rising use of Handle IDs, growing from **968** in publications

published in 2007 to **2,372** in 2022, highlights their growing significance in complementing DOIs, especially in institutional repositories where they ensure persistent access to digital resources. Similarly, PubMed Central IDs and PubMed IDs have seen considerable growth, particularly in Medical and Health Sciences. This trend reflects the sector's movement towards more digital and easily accessible platforms, enhancing research visibility and interoperability within these disciplines.

The fact that there is a decreasing share of publications with only DOIs and an increase in those with multiple PIDs indicates a diversifying PID landscape. This can be viewed positively, as it suggests a richer, more interconnected digital ecosystem where research outputs are more easily discoverable and linked across various platforms and databases. However, it also poses challenges in terms of managing and harmonizing these identifiers to avoid confusion and ensure seamless integration. Therefore, while the trend towards multiple PIDs per publication enhances the robustness and reach of scholarly communication, it also underscores the need for coherent PID management strategies to maximize their benefits in the OA environment.

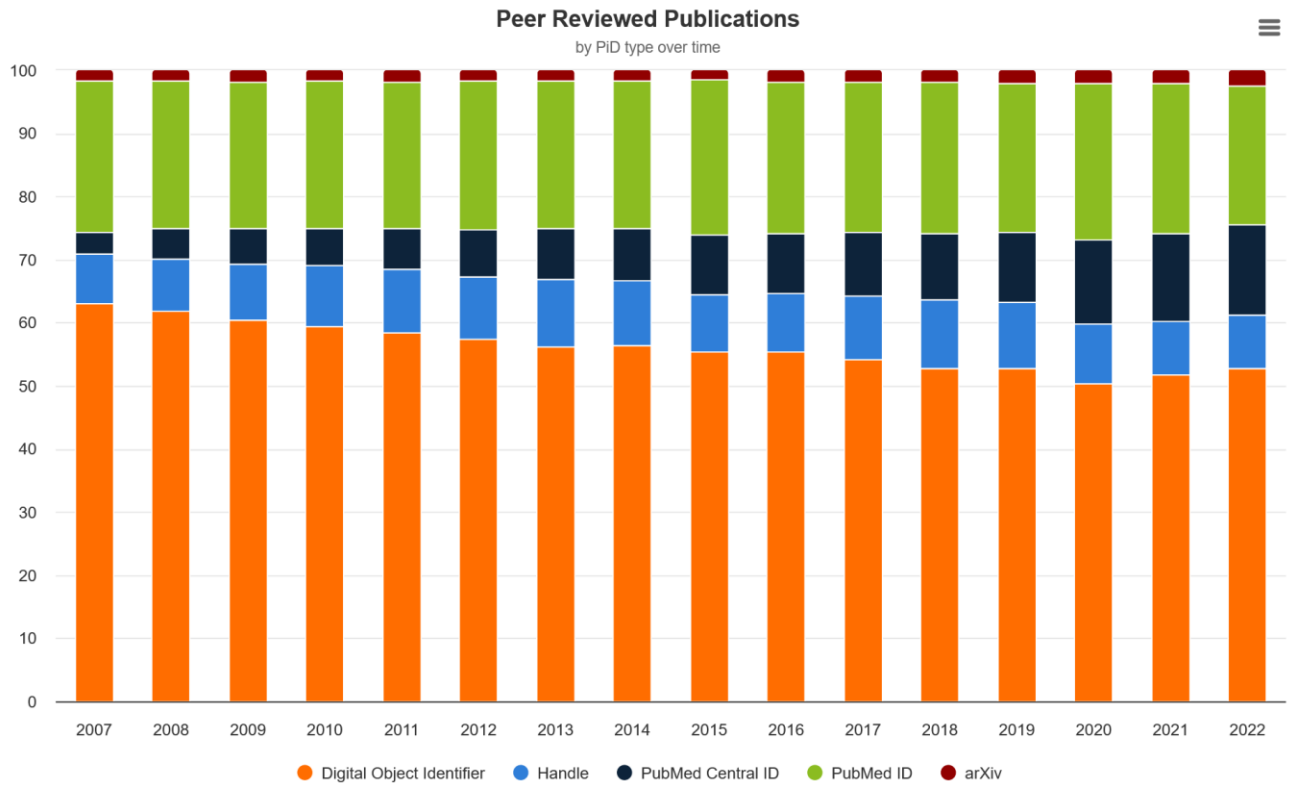

Figure 20: Peer-Reviewed Publications by PID Type over Time

ORCID iDs

ORCID iDs are unique identifiers assigned to individual researchers, playing a crucial role in distinguishing their work amidst the ever-expanding volume of scholarly publications. They are vital for several reasons: they resolve issues of name ambiguity, ensure accurate attribution of work, and facilitate the seamless linkage and tracking of an individual's research outputs across various platforms and affiliations over time.

An increasing trend in the use of ORCID iDs reflects a growing awareness and adoption of these identifiers among the research community. Starting from a usage rate of **51.6%** in publications from 2007, there has been a notable rise to **72.4%** in 2020, though with a slight dip to **64.6%** in 2022. This growth highlights a shift towards more efficient scholarly communication, with researchers increasingly managing their digital identity to secure proper recognition and visibility for their contributions. This trend benefits individual researchers by enhancing their profile and also supports the broader research ecosystem in improving data management and analysis of research outputs and collaborations.

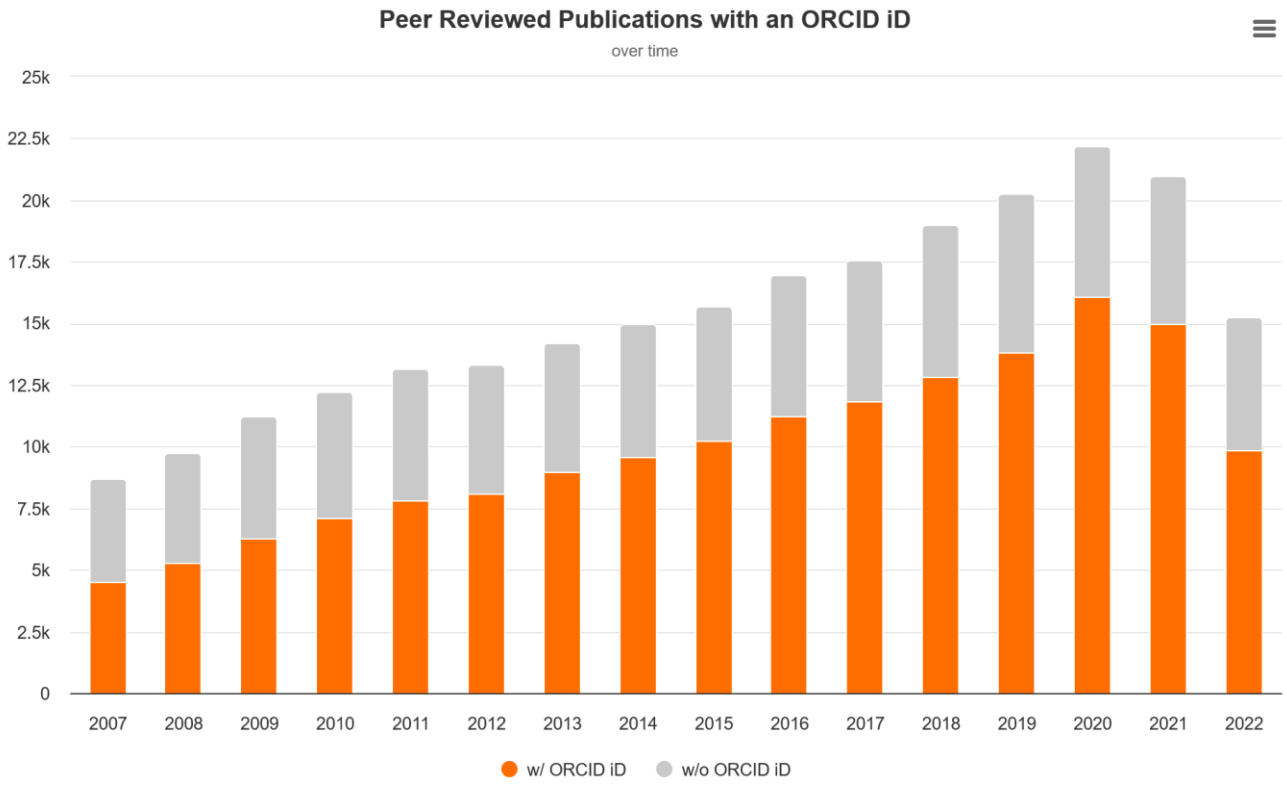

Figure 21: Peer-Reviewed Publications with an ORCID iD over Time

ABSTRACT

The presence of an abstract in research metadata is extremely important as it provides a concise summary of the research work, allowing readers to quickly assess the relevance

and scope of the publication. Abstracts serve as a critical tool for researchers and scholars, aiding in the efficient evaluation and selection of literature relevant to their field of study.

The data indicates an extremely high presence of abstracts in the metadata, consistently above **90%** up until 2021. This high percentage is indicative of robust metadata practices, suggesting that most publications are accompanied by summaries that can facilitate academic search and discovery.

However, it is important to approach these figures with caution. The mere presence of an abstract does not guarantee its quality or usefulness. In some cases, metadata may contain substandard abstracts, where the content could range from being irrelevant (such as a title or an acknowledgment being mistakenly labelled as an abstract) to incoherent or gibberish text. Therefore, while the high availability of abstracts is a good sign, it is equally crucial to ensure the quality and relevance of these abstracts to truly enhance the value and accessibility of the scholarly work.

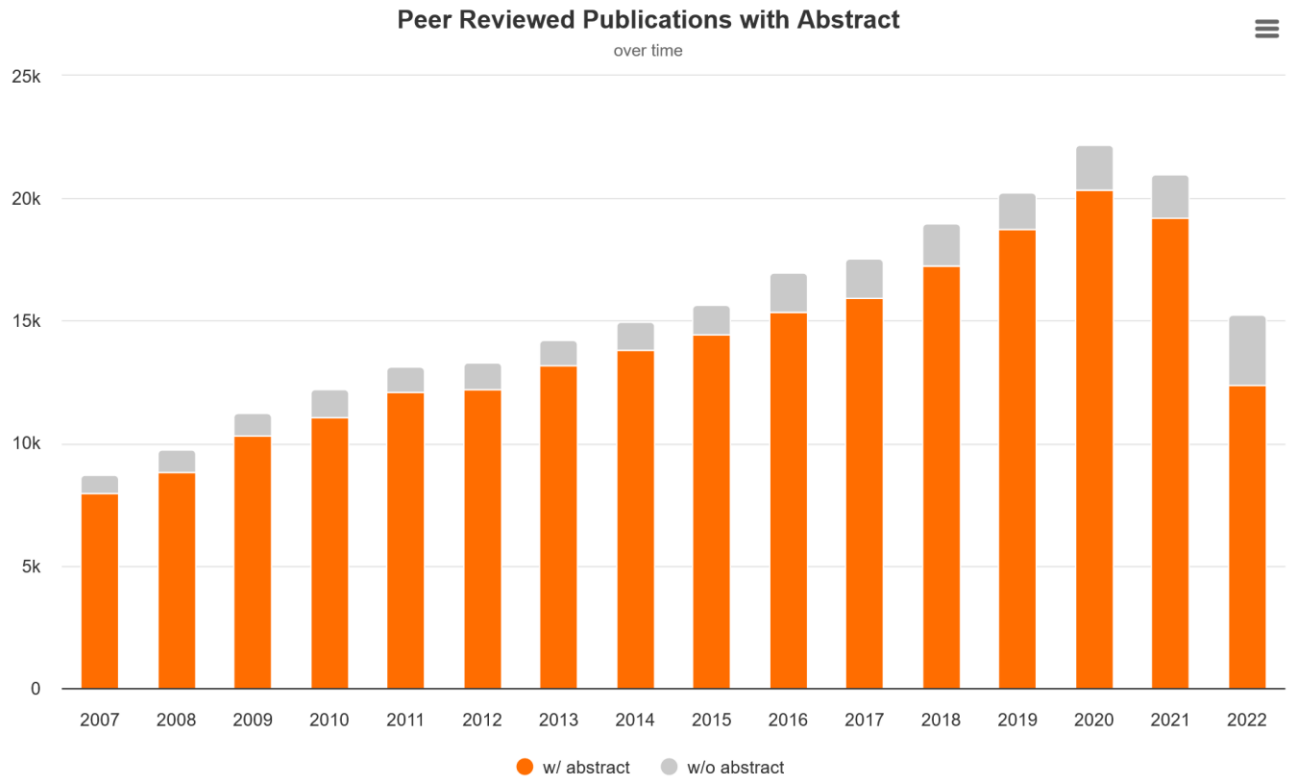

Figure 22: Peer-Reviewed Publications with an Abstract over Time

ACCESSIBILITY & INTEROPERABILITY

Of the **187,880** OA peer-reviewed publications in the Graph (with and without licence), **45%** have valid URLs in their metadata and only **33%** of all OA peer-reviewed publications have *URLs that link directly to the PDF of the full text*, rendering them accessible and interoperable[30].

Table 10: Accessibility & Interoperability of OA Peer-Reviewed Publications

% of Irish OA peer-reviewed publication with a valid URL in their metadata	*45%*
% of Irish OA peer-reviewed publications that are ACCESSIBLE (full-text can be fetched from URL link)	*33%*
% of Irish OA peer-reviewed publications that are INTEROPERABLE (full-text is in a machine-readable format)	*at least 33%*[31]

CONCLUDING FAIR ASPECTS

The analysis of FAIR principles in Irish peer-reviewed publications highlights a complex landscape of open science adoption, characterized by evolving licensing practices and the strategic use of PIDs. A noteworthy trend is the significant increase CC licences, from **6%** in 2007 to **56.4%** in 2022 publications, signalling a shift towards more open and reusable research outputs. Despite this progress, the prevalence of restrictive licences from major publishers points to ongoing challenges in maximizing the accessibility and utility of scholarly work.

The role of PIDs in enhancing the discoverability and long-term accessibility of research outputs is crucial. With **85.6%** of Irish peer-reviewed publications assigned a publication PID, the landscape is diversifying beyond traditional DOIs to include Handle IDs and PubMed IDs. This diversification aids in the seamless integration and interlinking of research outputs across various platforms. However, managing and harmonizing these different types of PIDs to avoid confusion and ensure coherent integration can be challenging.

[30] Interoperability is determined by a variety of other factors besides machine-readability but we use the term here for ease of exposition.

[31] We cannot ascertain the machine-readability of the rest of the publications as it was not possible to fetch the full text.

The adoption of ORCID iDs has seen significant growth, from **51.6%** in 2007 to **72.4%** in 2020 publications, highlighting their importance in accurately attributing work and tracking scholarly contributions. Unlike other PIDs, ORCID iDs focus on the author's identity, facilitating a clear connection between researchers and their outputs, thus enhancing the integrity of academic records.

The high availability of abstracts in scholarly metadata, exceeding **90%**, significantly aids in the efficient navigation of academic literature. However, ensuring the quality and relevance of these summaries is vital for their true utility in advancing scholarly communication.

In summarizing the exploration of certain key FAIR aspects, while strides have been made towards enhancing the openness and accessibility of scholarly communication, considerable challenges remain. The adoption of open licences and the implementation of PIDs, including notably ORCID iDs, indicate forward momentum. However, the prevalence of restrictive licences and the intricacies involved in PID management point to areas needing further attention and improvement. Additionally, the emphasis on data quality across all facets of the FAIR principles emerges as a critical factor in ensuring the reliability and usefulness of research outputs.

2.3 Plan S, APCs & Transformative Agreements

2.3.1 Plan S

Plan S[32] stands at the forefront of a major shift in OA publishing, especially since its requirements became effective post-2020. This initiative, led by cOAlition S, is a pivotal force reshaping the dissemination of publicly funded scholarly research. Its aim is to ensure the immediate and universal accessibility of research outputs, marking a new chapter in the narrative of scientific publishing.

Central to Plan S is the requirement that researchers funded by cOAlition S members must publish their research findings in OA journals or platforms. This includes the adoption of various compliant routes, such as publication in OA journals, deposition in recognized repositories, and engagement with Transformative Journals and agreements that progressively shift towards full OA.

A critical aspect of Plan S is its stance on transparency and accountability regarding publication fees. It endeavours to set a standard for fair and reasonable OA publication

[32] https://www.coalition-s.org/

fees, confronting the traditional financial models that have been dominating scholarly publishing. Furthermore, Plan S emphasizes the importance of licensing, particularly advocating for the use of the Creative Commons Attribution licence (CC BY). This approach not only enhances the openness of research but also enables the wider reuse of scholarly works with appropriate attribution to the original authors.

Currently there is only one funder in Ireland that is a member of cOAlition S, Science Foundation Ireland.

Table 11: Plan S Funders - Science Foundation Ireland (SFI)

Irish Plan S Funders: **Science Foundation Ireland**	***16,695 PR Publications***

The timeline below is meant to capture the pre- and post-Plan S (2021 onwards) access rights for SFI. There is a steady increase in the share of OA licenced peer-reviewed publications, from **68%** in 2018 to **85.1%** in 2022, and a gradual decrease in the share of Unrealised OA (OA without licence and Closed Access), from **27.7%** in 2018 to **12.9%** in 2022, rather than a sudden shift, potentially indicating preparation for Plan S compliance.

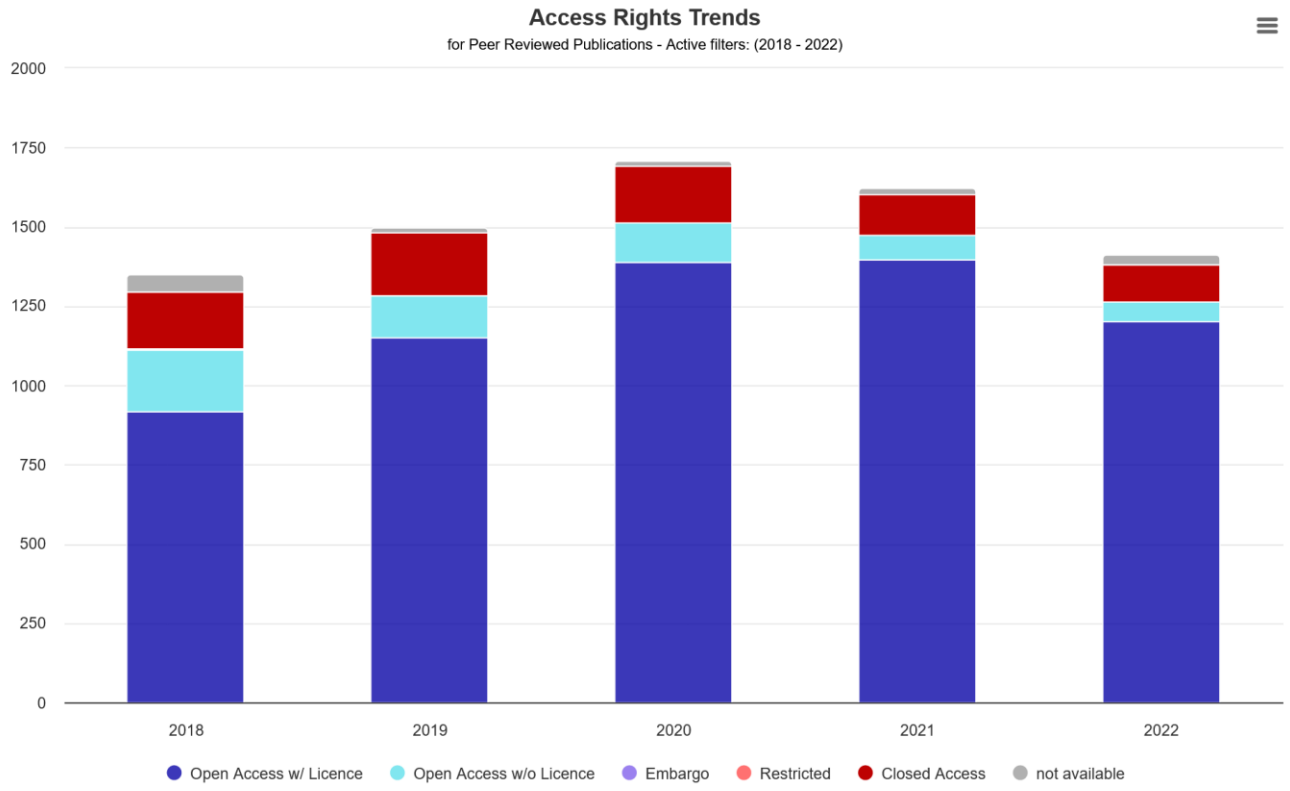

Figure 23: Peer-Reviewed Publications by Access Rights for SFI (Plan S Funder)

Followingly, we explore Plan S compliance indicators, particularly focusing on the period 2021 onwards and publications published

- in Diamond OA journals
- under Transformative Agreements
- in Transformative Journals
- following the Gold OA route.

These components, while not exhaustive of all Plan S requirements, provide insights into the evolving dynamics of OA publishing. They elucidate the way different OA models and agreements are being adopted and integrated in line with Plan S principles.

Table 12: Plan S-Compliant Peer-Reviewed publications

		# Irish OA peer-reviewed publications w/ licence post-2020	% of Irish OA peer-reviewed publications w/ licence post-2020 (31,801)
No APCs	In Diamond Journals	*47*	*0.2%*
	Under Transformative Agreements	*5,945*	*18.7%*
	In Transformative Journals	*2,617*	*8.2%*
Gold OA with APCs		*12,579*	*39.6%*

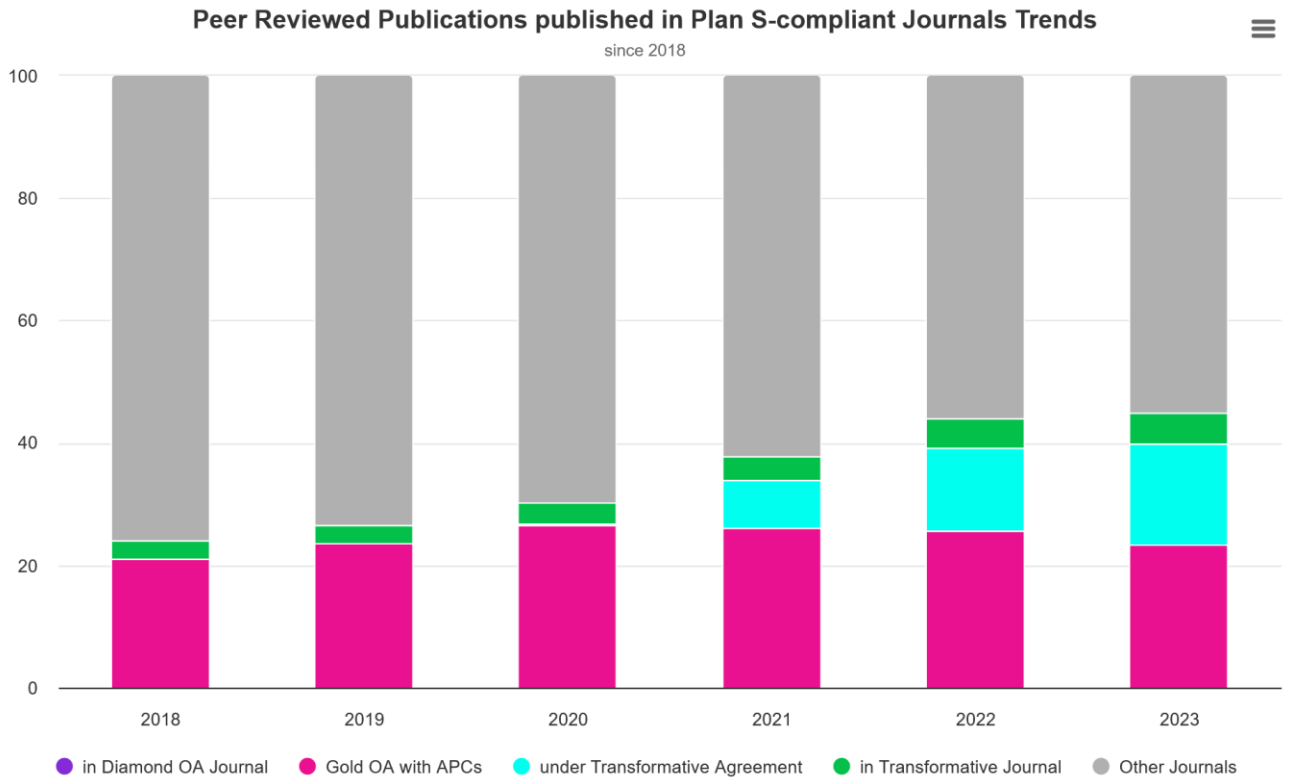

Figure 24: Peer-Reviewed Publications Published in Plan S-compliant Journals or under TA over Time

A significant majority of Plan S compliant publications are those published in Gold OA journals with APCs, accounting for **39,6%** of all OA peer-reviewed publications with a licence. This reliance on APCs might be sustainable in fields with adequate funding but raises concerns about inclusivity, particularly in less well-funded disciplines. Publications under Transformative Agreements and in Transformative Journals, represent **18,7%** and **8,2%** respectively. The data indicates a rise in the proportion of publications complying with Plan S mandates, driven primarily by the successful negotiation of Transformative Agreements, which are facilitated by IReL in Ireland. From 2020 to 2022, the share of publications published under Plan S compliance has increased by almost **70%**.

However, the number of publications in Diamond Journals, which offer OA without APCs, is markedly low at just **0.2%**. This small figure suggests that despite their potential as a cost-effective OA route, Diamond Journals may be underutilized or under-supported. This points to a potential area for policy intervention and increased advocacy to boost their visibility and viability in the OA landscape.

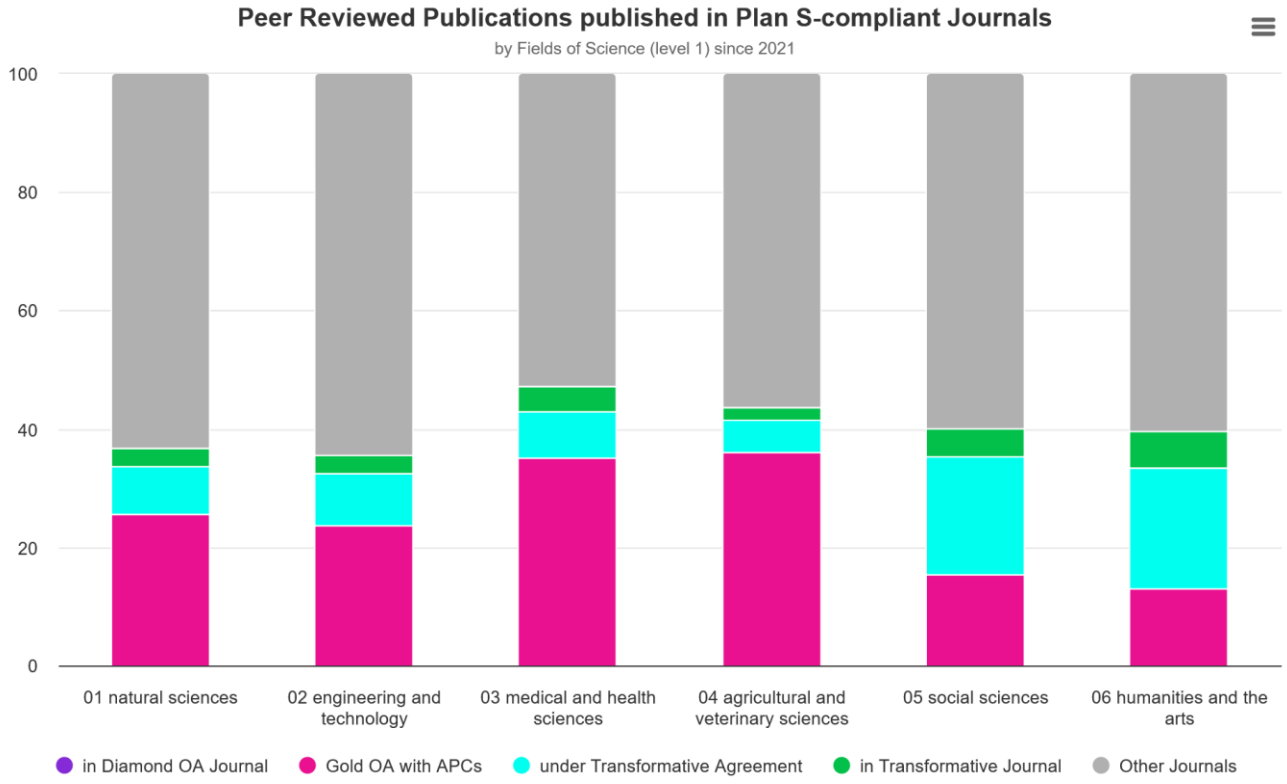

Figure 25: Peer-Reviewed Publications Published in Plan S-compliant Journals or under TA by FoS (level 1)

The landscape of scholarly publishing within Social Sciences, and the Humanities and the Arts offers a compelling narrative when dissecting the distribution of peer-reviewed publications published in Plan S-compliant journals or under TA. A striking observation is the relatively low representation of these fields in Gold OA with APCs (**15.3%** and **13.1%** respectively) juxtaposed against their prominent share of publications under Transformative Agreements (**20.1%** and **20.3%**). This trend may signal a concerted effort to negotiate new paths for OA that are better aligned with the financial realities and scholarly traditions of these disciplines.

For the Social Sciences, the move is underscored by the substantial volume of publications under transformative agreements (**644**), the highest among all scholarly fields. This also suggests a deliberate push towards redefining access in a domain where open dissemination can significantly enhance reach and engagement.

In the case of Humanities and the Arts, the shift towards transformative models, while also marked by higher proportions, may be partly attributed to the overall lower volume of publications. However, it is plausible to consider that this is not merely a numerical artefact but a conscious drive towards reshaping the OA landscape in these areas. Such

an approach could be seen as a response to the historical dominance of subscription models in these disciplines.

Within the rest of the fields, the distribution of publications across the various Plan-S compliant publishing models appears to align more consistently with the overall average production.

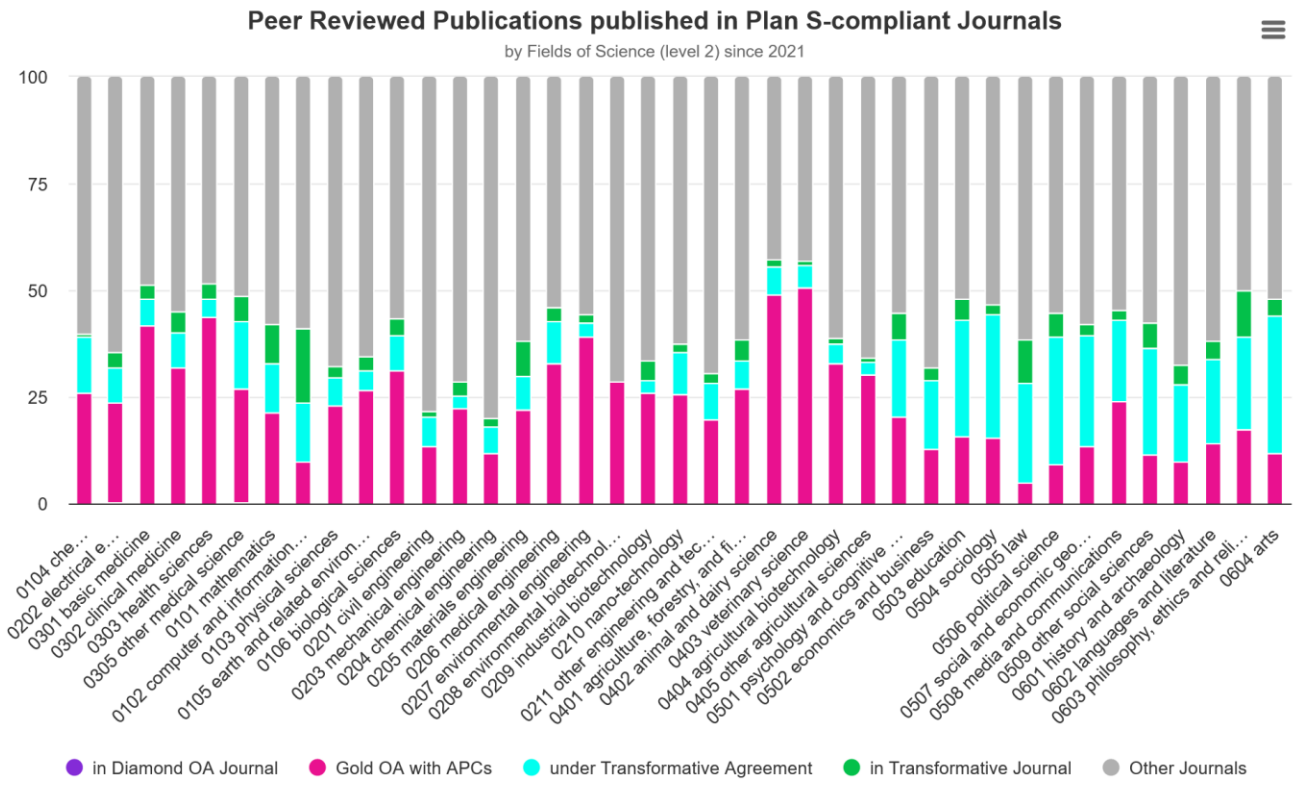

Figure 26: Peer-Reviewed Publications Published in Plan S-compliant Journals or under TA by FoS (level 2)

Analysing the data *within* FoS level 1 disciplines, we observe significant within-field variations in the distribution of publications published in a Plan S compliant way. Fields such as the Arts (within Humanities and the Arts), Media and Communications, Law and Psychology and Cognitive Sciences (Social Sciences), and Environmental and Chemical Engineering (Engineering and Technology) exhibit distributions that diverge from their field's average.

This variance suggests the presence of field-specific factors influencing the adoption of OA modalities. For instance, disciplines with a stronger tradition of subscription-based journals or those facing higher APCs might lean more towards Transformative Agreements, if available, or other Plan S-compliant pathways.

Understanding these variations is crucial for crafting effective and efficient open access policies. Tailoring strategies to address the unique challenges and incentives of each field can ensure that OA initiatives are not only broadly adopted but also maximize their potential benefits in enhancing the accessibility and visibility of research outputs. Such targeted policies could involve adjusting funding models, providing more support for transitioning journals to OA, or developing discipline-specific incentives for OA publication.

2.3.2 APCs & Transformative Agreements

The current landscape of APCs and Transformative Agreements within the realm of peer-reviewed publications presents a complex picture.

With APCs vs Under Transformative Agreements (TAs)

Transformative Agreements have become a crucial mechanism in the transition to OA, bypassing direct costs to authors. Since 2021, the proportion of publications benefiting from these agreements has been on the rise, reaching **23.3%** in 2022 and advancing to **29.2%** by 2023. This increase, despite potential gaps in the 2023 data due to publication and indexing delays, reflects the growing number of agreements signed. These agreements signify a commitment to expanding access to scholarly publications and eliminating financial obstacles for authors.

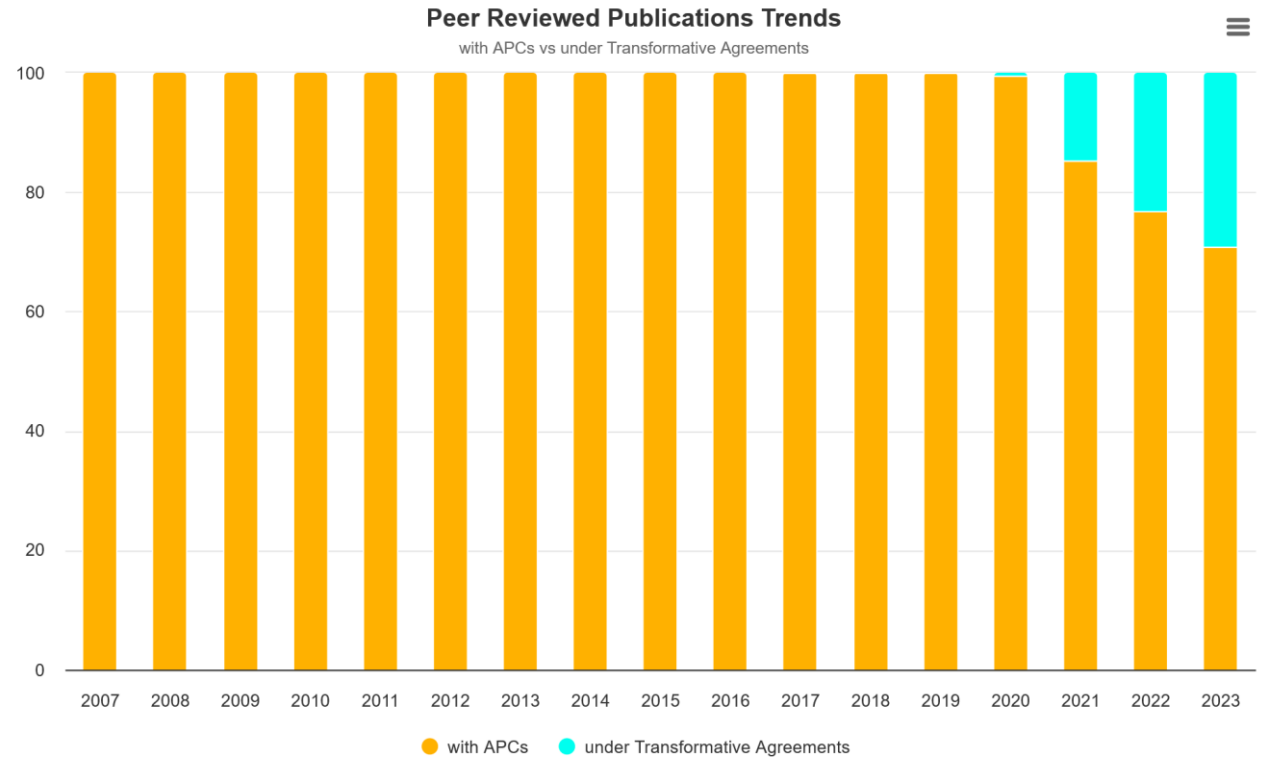

Figure 27: Peer-Reviewed Publications with APCs vs under Transformative Agreements over Time

The distribution of peer-reviewed publications under TA across different disciplines (FoS level 1) highlights a significant development in the scholarly publishing landscape. While the vast majority of publications are still subject to APCs, a discernible proportion across various disciplines benefits from TAs. Notably, the Social Sciences (**9.3%**) and Humanities and the Arts (**8.9%**) exhibit the highest engagement with TAs, suggesting these fields are at the forefront, mitigating direct publication costs for authors. In fact, the Social Sciences have the highest absolute number of publications under TAs overall. Conversely, disciplines such as Agricultural and Veterinary Sciences, Medical and Health Sciences, and Natural Sciences show lower engagement as a share of output, indicating a more gradual uptake of TAs.

These variations in the proportion of publications under Transformative Agreements across disciplines are influenced by the overall scholarly output within each field. For example, while Humanities and the Arts have 101 peer-reviewed publications benefiting from TAs, Medical and Health Sciences feature 879 publications under such agreements. Given the differing volumes of total research output, the relative impact of TAs is more pronounced for Humanities and the Arts.

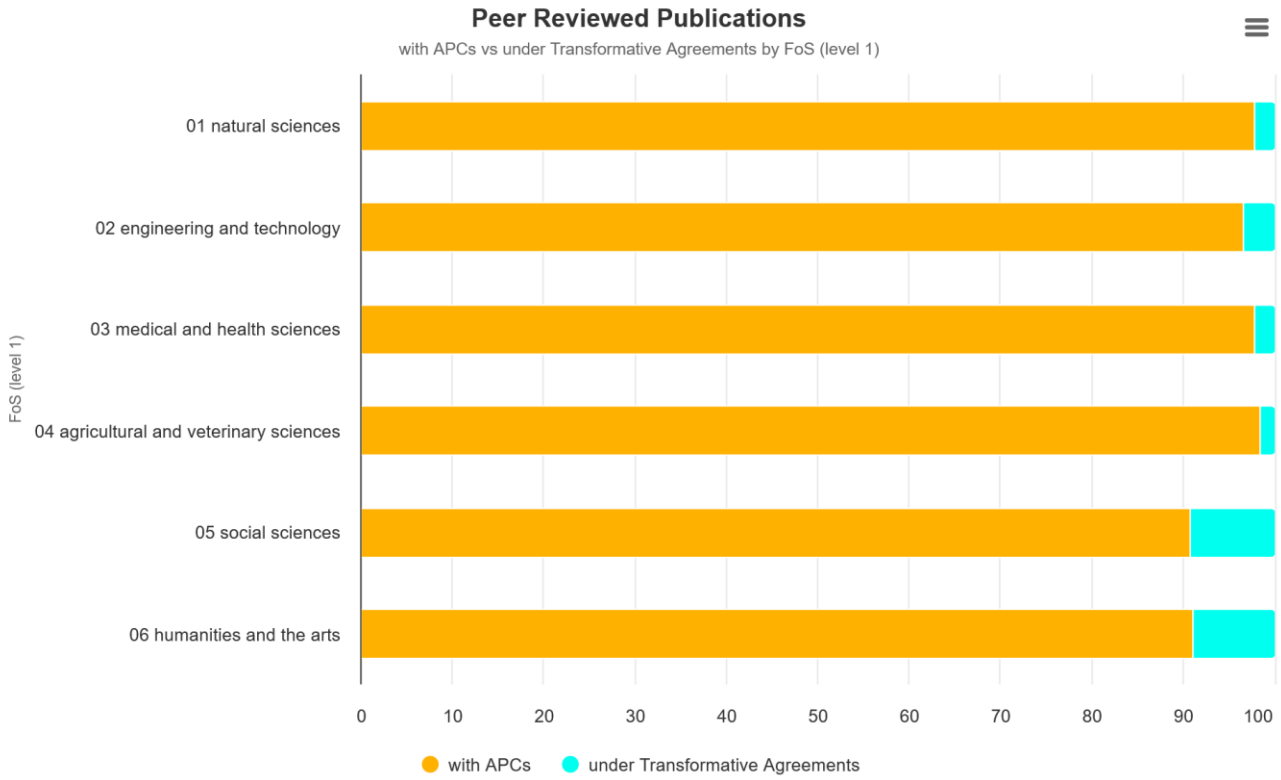

Figure 28: Peer-Reviewed Publications with APCs vs under Transformative Agreements by FoS (level 1)

The increasing adoption of TAs underscores a pivotal evolution in open access publishing. This trend, particularly pronounced in the Social Sciences and Humanities and the Arts,

signifies a strategic move towards mitigating publication costs and enhancing the accessibility of scholarly research across various disciplines.

APCs

Out of the **83,614** peer-reviewed publications that incurred an APC, only **1.4%** has APC data available (as reported by institutions to OpenAPC[33]), signalling a significant gap in data coverage. *This hampers our ability to fully understand and accurately extrapolate the financial aspects of OA publishing,* underscoring a pressing need for more robust reporting mechanisms.

Table 13: APCs & Transformative Agreements Overview

Peer-Reviewed publications that *incurred an APC*	*83,614*
Peer-Reviewed publications under Transformative Agreement *no APCs*	*6,006*
Peer-Reviewed publications with APC data available from OpenAPC	*1,194 (1.4% of all PR pubs with APCs)*
Total APCs incurred from OpenAPC – unclear who covered the APC	*€ 2.6Mi (1.4% coverage of APCs)*

An examination of the historical data on Article Processing Charges (APCs) reveals a consistent increase from 2008, peaking in 2017 with expenditures surpassing **€500,000**. After 2017, there is a notable reduction in APC spending. This decline is primarily due to

[33] https://openapc.net/

the analysis being based on only **1.4%** of the total data, but also partly due to the shift towards transformative publishing models.

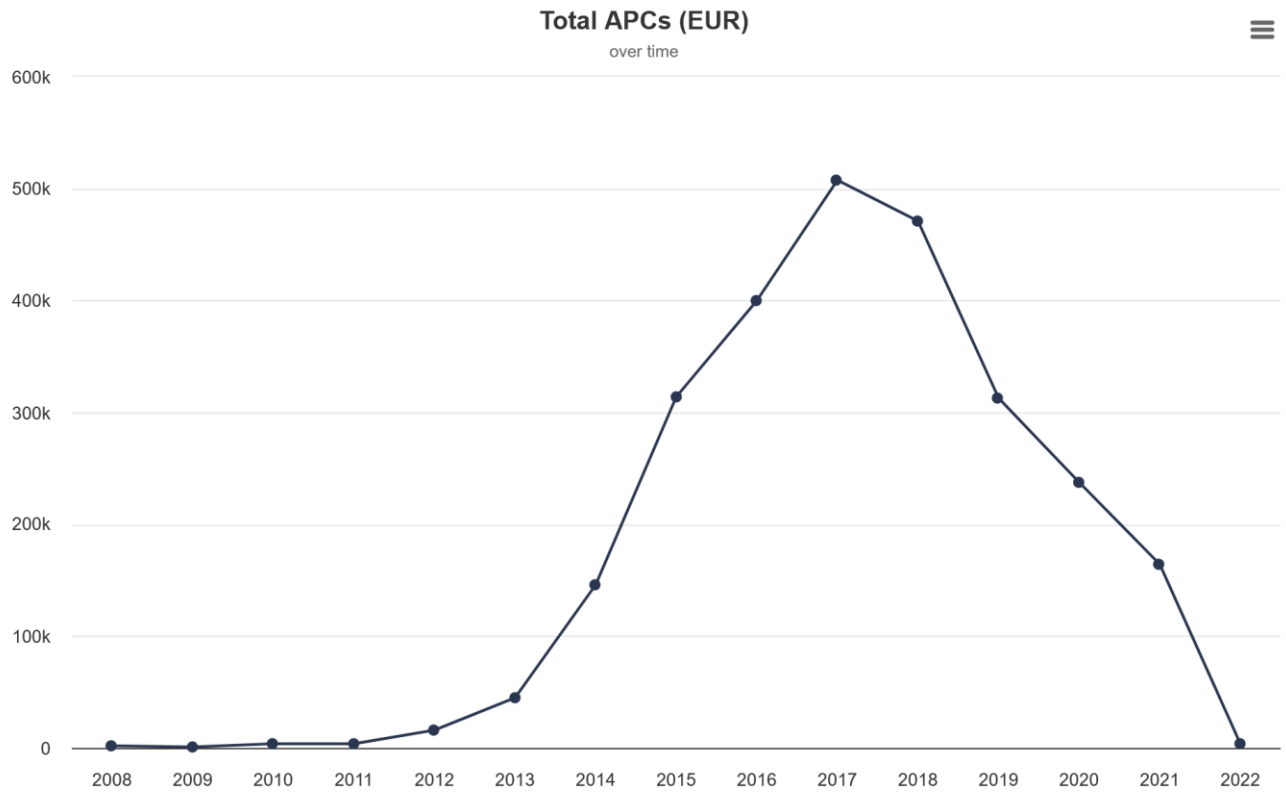

Figure 29: Total APCs (EUR) of OA Peer-Reviewed Publications over Time (1.4% data coverage)

Delving into journal-specific APC expenditures illustrates where the financial burdens for researchers and institutions are most concentrated. Notably, high-impact journals such as Nature Communications, PLOS ONE, and Scientific Reports lead the way in APCs, signifying a trend where prestigious publications command significant fees for OA services. This pattern underscores a broader dialogue within the academic community about the cost of OA and its implications for the dissemination of research findings. The prominence of journals with a medical or scientific focus suggests a high investment in disseminating medical research findings, possibly due to the urgency and societal value attributed to these fields.

These observations underscore a multifaceted scenario where high-profile or high-impact journals command significant APCs, potentially creating barriers to OA publishing for researchers with limited funding, such as those from low- and middle-income countries. As the scholarly community grapples with these financial aspects, the role of Transformative Agreements becomes increasingly pertinent.

Figure 30: Top 20 Journals by Total APCs (EUR) of Irish OA Peer-Reviewed Publications (1.4% data coverage)

The examination of average APCs across different scientific disciplines reveals distinct trends. Medical and Health Sciences exhibit the highest average APC at **€2,250**, reflecting the significant financial investment required for OA in this field, along with a moderate standard deviation. Humanities and the Arts stand out for having the lowest average APC at **€1,469**, coupled with the smallest standard deviation, that *could* suggest more uniform pricing within this discipline. A conclusion that can be verified once more data becomes available. Engineering and Technology shows lower average APCs compared to Natural and Social Sciences, highlighting the diversity in OA publication costs across disciplines and the potential impact on researchers' ability to publish. Again, it is crucial to approach these figures with caution due to the extremely limited data coverage, which means these observations may not fully represent the actual landscape.

Table 14: Average and Standard Deviation of APCs within FoS Level 1 (1.4% data coverage)

FoS Level 1	Average within FoS level 1	Standard Deviation within FoS level 1
Natural Sciences	€ 1,798	€ 421
Engineering and Technology	€ 1,583	€ 232

Medical and Health Sciences	€ 2,250	€ 261
Agricultural and Veterinary Sciences	€ 1,775	€ 144
Social Sciences	€ 1,809	€ 258
Humanities and the Arts	€ 1,469	€ 31

2.4 Main Findings

The analysis of OA in Ireland paints a multifaceted picture of the current state and progression of scholarly communication. This section synthesizes the key observations and insights derived from our extensive review, highlighting both the achievements and the areas needing attention to further advance OA in Ireland.

1. Scholarly Production and Research Landscape

- Among the total **423,893** Irish publications, **333,404** (**78.7%**) are peer-reviewed, reflecting a strong emphasis on academic rigor and quality. In addition, a consistent upward trend in peer-reviewed output, from **8,717** in 2007 to **20,964** in 2021, highlights a dynamic research environment and increased digital practices like DOIs and repository usage.

- Medical and Health Sciences lead Ireland's publication volume, in line with global health research priorities. Significant attention is also directed towards Natural Sciences, Engineering, and Technology, underlining their role in driving innovation and economic progress. Specialized strengths are evident in Physical and Chemical Sciences, as well as in Electrical, Electronic, and Information Engineering, highlighting Ireland's focused research areas. Although Social Sciences, and Humanities and the Arts might appear underrepresented due to differences in publication practices, digital presence and potentially funding, their contributions, particularly in economics, business, education, history, archaeology, languages, and literature, play a key role in the scholarly landscape.
- Journals remain the primary source for hosting peer-reviewed publications, but there is a growing engagement with repositories, indicating an evolving landscape of research dissemination.

2. Open Access and Licensing

- **Progressive Shift Toward Licensed OA.** The proportion of licensed OA publications has seen a significant increase, from **21.6%** in 2007 to **66%** in 2022, signifying a robust move towards open scholarly communication. There remains a significant portion

of the scholarly output that is either unlicensed OA (**17.4%**) or still under Closed Access (**26.3%**), highlighting areas where further efforts are necessary to achieve full OA. Nevertheless, the reduction of closed access publications to **13.7%** by 2022 indicates a gradual move away from subscription-based models towards more accessible research.

- Journals host a significant number of both licenced and unlicenced OA publications but traditional subscription models are still prevalent in this space. Publication and Institutional Repositories include the highest shares of licenced OA content.

- The Medical and Health Sciences lead in the proportion of licensed OA publications at **44.3%**, with the Natural Sciences following closely. Yet, the Social Sciences, Agricultural and Veterinary Sciences, and the Humanities and the Arts have less than **50%** of their outputs available as OA (licenced or unlicenced). This variance suggests a slower uptake of Open Science principles in certain fields, possibly due to cultural or infrastructural barriers. Further, detailed analysis reveals that Health Sciences within the Medical and Health Sciences sector, and Environmental Engineering within Engineering and Technology, exhibit high levels of licensed OA publications, underscoring their commitment to open dissemination of research. Conversely, certain disciplines within the Social Sciences show lower than average OA adoption rates, reflecting the heterogeneity in Open Science practices within fields.

3. OA Routes and Publishing Models

- **Publisher-mediated OA** (Gold and Hybrid OA), is the dominant route for making licenced OA publications available in Ireland. Gold OA accounts for **38.1%**, and Hybrid OA for **30.8%** of all licenced OA publications. Repository-mediated OA (Green OA with a licence) represents a smaller fraction, only **6.5%** of the total. Trend data shows growth across all OA categories, with a notable increase in repository-mediated OA, suggesting its strengthening role despite the predominant publisher-mediated OA.

- Disciplinary differences are evident, with Medical and Health Sciences, Natural Sciences, Engineering and Technology, and Agricultural and Veterinary Sciences showing a strong preference for publisher-mediated OA. In contrast, Social Sciences and Humanities show a more balanced mix.

- A trend towards Hybrid OA over Gold OA suggests evolving journal policies and the influence of initiatives like Plan S on publishing practices.

- Unrealised OA, which includes Closed Access, Bronze and unlicenced Green publications, highlights the gap between potential and actual open accessibility. A significant volume of Green publications without a licence (**72,320**) suggests a need for improved licensing practices to transform the OA landscape. Moreover, trends indicate a steady increase in Bronze and Green publications without a licence, suggesting growing potential for OA amidst a plateauing trend in Closed Access publications.

4. FAIR Aspects

- There is a marked increase in the use of Creative Commons (CC) licenses, from **6%** in 2007 publications to **56.4%** in 2022, indicating a move towards enhancing the openness and reusability of research outputs. Despite this positive trend, the prevalence of restrictive licenses issued by major publishers continues to pose challenges in maximizing research accessibility and utility.

- A high percentage (**85.6%**) of peer-reviewed publications are assigned publication PIDs, such as DOIs, Handle IDs, and PubMed IDs, facilitating improved discoverability and long-term accessibility. This diversity in PIDs which is increasing over time, supports the integration of research outputs across platforms, although harmonizing these identifiers to ensure seamless integration can present challenges.

- The use of ORCID iDs has significantly increased, from **51.6%** in 2007 to **71.4%** in 2021, underscoring their role in attributing work accurately and tracking scholarly contributions.

- Overall, the exploration into FAIR principles highlights both progress and obstacles towards open and accessible scholarly communication in Ireland. Moreover, the focus on ensuring **high-quality data** across all FAIR aspects (e.g. the quality of abstracts in metadata and the validity of URLs to the full-text of publications) is essential for maintaining the reliability and effectiveness of research outputs.

5. Plan S Compliance

- Currently, Science Foundation Ireland (SFI) stands as the sole member of cOAlition S in Ireland, indicating a limited RFO engagement with Plan S within the country. The analysis of SFI's access rights from 2018 to 2022 reveals a gradual increase in OA licensed peer-reviewed publications, reaching **85.1%** in 2022, alongside a steady decline in Unrealised OA from **27.7%** in 2018 to **12.9%** in 2022.

- Exploring Plan S compliance indicators, including publications in Diamond OA journals, under Transformative Agreements, in Transformative Journals, and following the Gold OA route, provides insights into the evolving dynamics of OA publishing. Gold OA journals with APCs constitute the majority (**39.6%**) of Plan S compliant publications, indicating a heavy reliance on APCs potentially posing inclusivity concerns. From 2020 to 2022, there has been a significant increase of almost **70%** in the share of publications published under Plan S compliance, primarily driven by successful Transformative Agreements. Lastly, despite being a cost-effective OA route, Diamond Journals witness a notably low representation (**0.2%**) among Plan S compliant publications, suggesting underutilization or lack of support, indicating a potential area for policy intervention and advocacy.

- As is the case throughout the analysis, significant within-field (FoS level 1) variations are observed in the distribution of publications compliant with Plan S, reflecting discipline-specific factors influencing the adoption of OA modalities.

6. Transformative Agreements

- The proportion of peer-reviewed publications benefiting from Transformative Agreements (TAs) has steadily increased, reaching **23.3%** in 2022 (out of all publications that would have incurred an APC) signifying a growing commitment to expanding access to scholarly publications and alleviating financial burdens for authors.

- Across various disciplines (FoS level 1), a discernible proportion of publications, particularly in the Social Sciences (**9.3% overall**) and Humanities and the Arts (**8.9%**), benefit from TAs, indicating proactive efforts to mitigate direct publication costs. Notably, the Social Sciences exhibit the highest absolute number of publications under TAs, suggesting significant engagement within this field.

7. Article Processing Charges (APCs)

- Out of **83,614** peer-reviewed publications incurring APCs, only **1.4%** have APC data available, indicating a significant gap in data coverage. This limitation hampers a comprehensive understanding of the financial aspects of OA publishing and underscores the need for more robust reporting mechanisms.

Overall Observations

The examination of Open Access (OA) in Ireland unveils a positive trajectory towards enhanced scholarly communication accessibility. There has been a noticeable surge in licensed OA publications, suggesting a growing commitment to openness. Nonetheless, hurdles remain, such as unlicensed OA and Closed Access articles, necessitating tailored

strategies for broader OA adoption. The analysis also reveals a preference for publisher-mediated OA, the growth of Hybrid OA, and the challenges and potential in realizing full OA across disciplines.

Strides towards FAIR principles and Plan S compliance similarly showcase progress and some challenges. Addressing gaps in data coverage and supporting underutilized OA avenues through policy interventions and advocacy are crucial for cultivating an inclusive scholarly communication ecosystem characterized by openness and accessibility.

3 Methodology

This section provides an outline of the methodological steps taken to conduct the analysis of OA in Ireland presented in the previous section. This transparent approach ensures the integrity and reliability of our findings.

3.1 OpenAIRE Graph: Foundation of the Monitor

The Monitor is built upon the OpenAIRE Graph (https://graph.openaire.eu, the Graph from here on). 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 includes around metadata records for about **240 million** research products from more than **129K** 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. 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. A new version of the OpenAIRE Graph is available every month. The OpenAIRE Graph is available for download and reuse under a CC-BY licence.

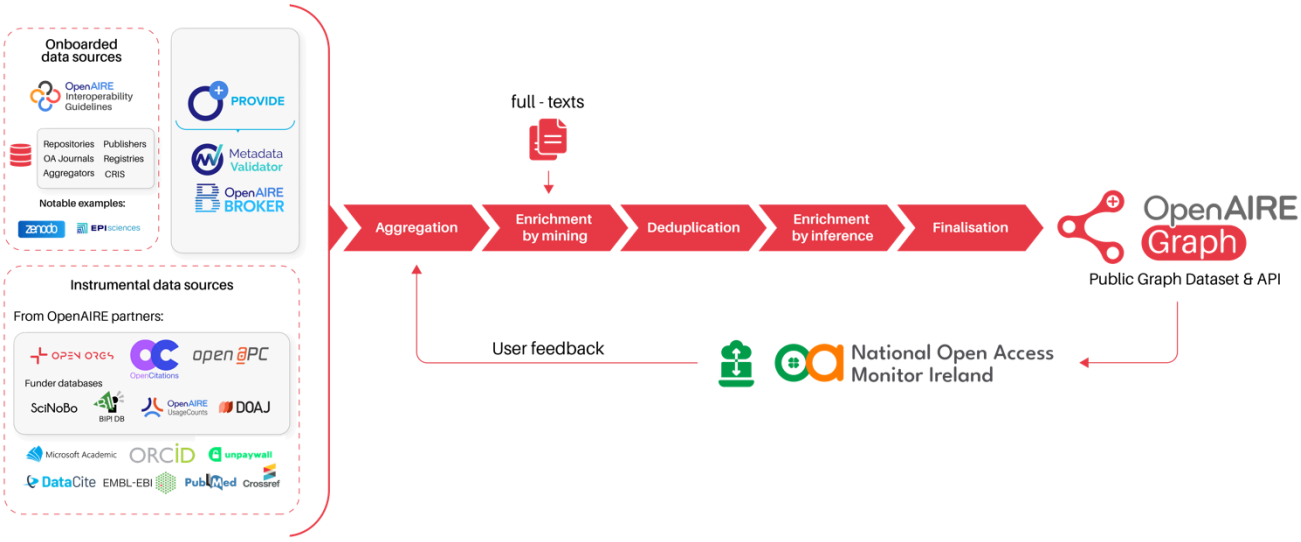

Figure 31: The Monitor and the OpenAIRE Graph Pipeline

3.2 The Publication Set of the Monitor

This section outlines the processes through which the Monitor and this Report compile and utilize data from Irish Research Performing Organizations (RPOs) and Research Funding Organizations (RFOs), ensuring comprehensive representation of Irish peer-reviewed publications.

We take the following steps:

1. Identify Irish RPOs and their publications.
2. Identify Irish RFOs and their publications.
3. Exclude non-peer-reviewed publications.
4. Make sure PIDs are used for Irish RPOs, RFOs and publications.

The final set for the Monitor currently includes **333,404** peer-reviewed publications.

Identification of Irish Research Performing Organisations' (RPOs) Publications

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

1. Institutional data sources registered in OpenAIRE (repositories, CRIS, Open Access Journals)
2. Metadata from harvested data sources such as Crossref.

3. Inferred links via text mining.
4. Links created via the claim and link functionalities in OpenAIRE EXPLORE[34].

By consolidating the RPO list provided by IReL through the "National Open Access Monitor Survey: Organisational Identity" and cross-referencing it with the associated RPOs within Ireland on the OpenAIRE Graph, we have a created a list of Irish RPOs for the Monitor. Currently, there are approximately **800** organisations tagged as RPOs that require further refinement[35].

Irish Institutional Data Sources Registered in OpenAIRE

Institutional data sources (1. above) provide direct affiliation information to the OpenAIRE Graph, i.e. a publication from an institutional source is immediately given the corresponding affiliation. We present the coverage of Irish institutional data sources in the Monitor below[36].

Table 15: Irish Institutional Data Sources & Alignment with OpenAIRE

Data Source Type	Registered in OpenAIRE		Harvested by OpenAIRE (directly or via a compatible aggregator)		Not registered but Harvested by OpenAIRE (for the purposes of the Monitor)	
	# sources	# publications	# sources	# publications	# sources	# publications
Institutional Repositories	*14*	*187,881*	*0*	*0*	*7*	*14,834*
Thematic/ Publication Repositories	*1*	*1,281*	*0*	*0*	*0*	*0*
CRIS Systems	*0*	*0*	*0*	*0*	*0*	*0*
OA Journals	*0*	*0*	*18*	*2,963*	*0*	*0*

[34] These functionalities, also available in the Monitor, allow users to claim research products as their own and add them to their synced ORCID record and to link research products with other research products, communities, or project. https://oamonitor.ireland.openaire.eu/how-it-works/user-actions#adding

[35] See Section 5.1.

[36] The Appendix describes steps taken by OpenAIRE to harvest repositories and subsequent steps to follow to ensure better coverage in the Monitor.

Among the registered Irish Institutional Repositories within OpenAIRE, a cumulative count of **187,881** publications is recorded. There are **3** repositories that are compliant with the version 3.0 of the OpenAIRE Guidelines. **10** (plus **1** Thematic Repository) adhere to the BASIC and to version 2.0 of the OpenAIRE Guidelines. Consequently, these repositories do not conform to the most recent IT and repository standards, which necessitate more contextually enriched content, including links and associations with various research outputs and entities. Furthermore, they do not accommodate diverse and enhanced vocabularies.

Identification of Irish Research Funding Organisations' (RFOs) Publications

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.
- Merging information from Crossref's Open Funder Registry (OFR)[37].
- Collecting links from users via the "link" functionality[38].
- 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.

Irish Funders in OpenAIRE

Irish funders are represented in the OpenAIRE graph through two primary avenues. The first is by directly joining OpenAIRE[39], a process that entails providing a comprehensive list of research projects, the creation of a tailored text mining algorithm for data extraction, and meticulous curation of project-publication links to ensure accuracy. Science Foundation Ireland (SFI) has successfully undergone this process.

Table 16: Irish Funders that have joined OpenAIRE

Irish RFOs that have joined OpenAIRE	Projects	Publications
Science Foundation Ireland (SFI)	*6,824*	*21,895*

[37] https://www.crossref.org/services/funder-registry/

[38] https://oamonitor.ireland.openaire.eu/how-it-works/user-actions#linking

[39] https://www.openaire.eu/funders-how-to-join-guide

The second avenue for representation is through the OFR using the funders' fundref IDs. While this allows funders to be associated with publications via valid DOIs in the OpenAIRE Graph, it does not offer the granularity of the direct integration, notably the curated project-publication links.

Table 17: General Representation of Irish Funders in OpenAIRE (excluding SFI)

# Irish RFOs integrated via OFR (excluding SFI)	# Irish RFOs' publications (excluding SFI)
142	*17,059*

For a funder to achieve the detailed representation observed with SFI, a direct integration with OpenAIRE is recommended. This not only ensures a comprehensive presence but also guarantees the precision of the data incorporated.

Peer-Reviewed Publications

We refine the set of Irish publications, as detailed in the preceding section, by focusing only on peer-reviewed artifacts, based on the following criteria. If either of them is true then a publication is considered peer-reviewed.

1. **Curated Peer-Review Assessment:** The OpenAIRE team has engaged in a curation process to determine peer-review status. This hand-curated assessment has been integrated into the Graph and is continuously under development.
2. **Exclusion of Grey Literature:** We filter out grey literature, which includes document types that typically bypass the peer review process, such as reports, and white papers. Given that the OpenAIRE Graph aggregates data from various sources, resulting in merged records, we specifically exclude entries where all instances are identified as grey literature.

   & Presence of DOI from Crossref: Since Crossref predominantly catalogues peer-reviewed content, its DOIs help maintain the scholarly credibility of our included publications.

The combination of these criteria gives the following number of Irish peer-reviewed publications of the Monitor.

Table 18: Irish Peer-Reviewed Publications

Irish peer-reviewed publications
333,404

Additional Criteria Under Examination

Beyond the core criteria, we are actively delving into additional parameters that might further refine our identification process.

- **Presence of DOI from DataCite:** We are ascertaining if such an inclusion can offer breadth to our dataset. The examples of peer-reviewed publications that we found with a DOI issued by DataCite were so far **all identified** as peer-reviewed in the Graph even without including this selection criterion. We will continue to examine this option and our system is flexible to incorporate it.
- **Reference Count by Field of Study (FoS):** We also investigated the potential for establishing a reference count threshold, beneath which a publication might be classified as non-peer-reviewed. However, the analysis of mean and standard deviation of reference counts within each FoS level revealed significant variability. This variability precludes the adoption of a uniform criterion based on reference count. We plan to continue our exploration in this area to lower levels of FoS to refine our assessment criteria.

Use of Persistent Identifiers (PIDs)

To achieve accurate and comprehensive monitoring of Irish scholarly publications, we place emphasis on the use of Persistent Identifiers (PIDs). PIDs serve as essential building blocks, allowing us to uniquely identify these publications, facilitating the discoverability, accessibility, and reusability of research outputs.

The Monitor specifically defines an Irish scholarly publication as one that contains a persistent identifier (PID) associated with an Irish organization. These PIDs can be found in various places within the publication's metadata, PID metadata, or even within the publication content itself. We seamlessly integrate a range of PIDs for both research outputs and organizations. The process of deduplication ensures that metadata records from different data sources are effectively merged, accompanied by publicly displayed provenance information. This comprehensive approach guarantees not only the widest possible coverage but also maintains the integrity and consistency of our data.

The table below provides an overview of the PIDs used for publications, organizations, and authors within the Monitor, along with the corresponding number of publications associated with each type of PID.

Table 19: Irish peer-reviewed publications by PID type

PID type	# Irish peer-reviewed publications
Publication PIDs	
Digital Object Identifier (DOI)	*309,897*
Handle	*48,124*
PubMed Central ID	*46,249*
PubMed ID	*126,957*
arXiv	*9,402*
Organisation PIDs	
Participant Identification Code	*307,904*
ISNI	*439,964*
OrgRef	*411,810*
Open Funder Registry	*390,591*
Wikidata	*442,293*
GRID	*448,838*
RingGold	*22,453*
ROR	*449,951*
OrgReg	*414,289*
ORCID iDs	
ORCID iD	*191,845*

3.3 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. On the Monitor, managers

overseeing the National, RPO, and RFO dashboards will have access to OpenOrgs, empowering them to deduplicate Irish RPO records.

Journals, Publishers, and Licences: To ensure precision and reliability in its data, the Monitor disambiguates journals using their ISSN numbers and publishers through the utilization of Crossref metadata, including ROR IDs and DOI prefixes, among other identifiers. This effort will be bolstered by custom text similarity algorithms.

Additionally, OpenAIRE is systematically working on normalizing licences. Currently, about **98%** of licences in the December version of the Graph have been successfully grouped and normalized. However, accurately comparing and categorizing these licences, especially non-Creative Commons (non-CC) ones, remains a challenge.

Authors: Researcher dashboards[40] are fully integrated with ORCID profiles, streamlining the identification process by linking directly to researchers' ORCID IDs. This integration accounts for all name variations associated with a researcher's ORCID ID, ensuring accurate and comprehensive representation. When researchers log in using their ORCID ID, they can easily claim additional research outputs. These claims, once made, synchronize with their ORCID profile, and are updated in both the ORCID system and the Monitor with the monthly update of the OpenAIRE Graph. This seamless connection between the two platforms guarantees that researchers have consistent and visible access to their entire body of work.

3.4 Enrichment via Text Mining

To enrich metadata and enhance the comprehensiveness of scholarly records, OpenAIRE employs several text-mining methods. These methods include:

- **Affiliation Matching:** This process involves matching affiliations extracted from PDF and XML documents with organizations listed in the OpenAIRE organization database.
- **Funding Classifiers:** Utilizing a document classification algorithm, OpenAIRE analyses free text from abstracts of publications to categorize scientific text into one or more predefined content classes, such as funders and projects.
- **Extraction of Acknowledged Concepts:** OpenAIRE scans plaintexts of publications to identify acknowledged concepts. These may include grant identifiers (projects) from funders, accession numbers of bioentities, mentions of EPO (European Patent Office) patents, and custom concepts that link research objects to specific research communities and initiatives within OpenAIRE.

[40] https://oamonitor.ireland.openaire.eu/researcher

- **Metadata Extraction:** OpenAIRE employs the CERMINE project to extract plaintext and metadata from PDF documents. This extraction process covers various aspects, including titles, authors, affiliations, abstracts, keywords, journal names, volume, and issue information, parsed bibliographic references, as well as the structure of document sections, section titles, and paragraphs.

The OpenAIRE PDF aggregation system was strategically redirected to prioritize the collection of PDFs from Irish OA publications. Consequently, we were able to validate **84K** URLs of OA peer-reviewed publications and successfully retrieved **61K** PDFs from these. As the OpenAIRE PDF aggregation system operates continuously, we expect the number of successfully retrieved PDFs to grow steadily over time. These documents were then processed through the Graph pipeline, facilitating the implementation of the text-mining methods mentioned earlier. Additionally, we mined these documents for classification into Fields of Science (FoS) and Sustainable Development Goals (SDGs).

FoS Classification System: To categorize into distinct levels FoS[41], we have integrated an advanced classification system (Kotitsas, et al. 2023[42]). This system utilizes Natural Language Processing (NLP) to analyse various components of the OpenAIRE Graph, including abstracts, citations, references, and venues. As a result, each publication is systematically classified into FoS classes down to level 3, adding precision to its scientific domain. This hierarchical categorization not only provides a structured framework but also bolsters our ability to pinpoint multidisciplinary overlaps within the research.

SDG Classification System: To contextualize the impact of research on addressing paramount global challenges, we have incorporated a classification mechanism aligned with the UN Sustainable Development Goals (SDGs). This schema is engineered to elucidate the alignment of research endeavours with critical issues, ranging from climate adaptation, biodiversity preservation, mitigation of environmental contaminants, to socioeconomic upliftment.

[41] The taxonomy is presented here: https://explore.openaire.eu/fields-of-science

[42] Kotitsas, S., Pappas, D., Manola, N., & Papageorgiou, H. (2023). SCINOBO: a novel system classifying scholarly communication in a dynamically constructed hierarchical Field-of-Science taxonomy. Frontiers in Research Metrics and Analytics, 8, 1149834.

3.5 Indicators

The table below presents the construction methodology of indicators included in this Report, offering a detailed look at how each indicator is derived and calculated. The definitions of these indicators are given in the Glossary[43].

Table 20: Construction Methodology of Indicators

Attribute	Construction Methodology
Under Transformative Agreements	We have identified and retrieved from OpenAPC (IReL OpenAPC Dataset[44]) the set of articles with metadata published under Transformative Agreements for Ireland.
Journal Business Models	
OA (Gold)	A journal is fully Open Access if one or more of the following occur: It is in the Directory of Open Access Journals (DOAJ) It has a known fully OA Publisher (OpenAIRE's curated list). It only publishes OA articles.
Diamond OA	We obtain APC data from DOAJ using DOAJ's Public Data Dump[45] (an exportable version of the journal metadata). We used it to determine whether a particular fully OA journal charges APCs.
Subscription	Journals without any open access articles.
Hybrid	Journals that are not Gold and publish at least one OA peer-reviewed article.
Transformative	We identify Transformative Journals by ISSN matching with the publicly available Transformative Journals data[46] from the Plan S initiative.
OA Routes/Colours	
Green OA with Licence	A scholarly publication with an OA instance in a repository that includes a licence specified in its metadata.
Gold OA	A scholarly publication hosted by an (Gold) OA journal (as defined above) and integrated in the Graph.

[43] The indicators on the Monitor are described here: https://oamonitor.ireland.openaire.eu/methodology/terminology#constructed-attributes

[44] https://github.com/OpenAPC/openapc-de/tree/master/data/transformative_agreements/IReL

[45] https://doaj.org/docs/public-data-dump/

[46] https://journalcheckertool.org/Transformative-journals/

Attribute	Construction Methodology
Hybrid OA	A scholarly publication with an OA instance hosted by a Hybrid Journal (see above) that includes a licence specified in its metadata.
Unrealised OA	
Bronze	A scholarly publication with an OA instance hosted by a Hybrid Journal (see above) that does not include a licence in its metadata.
Green without licence	A scholarly publication with an OA instance in a repository that does not include a licence in its metadata.
Accessibility – Interoperability[47]	
Accessible	We construct the dataset of accessible publications by leveraging OpenAIRE's collection of full-text OA publications, which includes PDFs of more than **20M** documents. This involves an examination of URL links within each publication's metadata to locate and retrieve the corresponding PDF document, if it is retrievable that publication is accessible. Recognizing that a single publication may be associated with multiple links, our method entails navigating each one to ensure comprehensive coverage.
Interoperable	The construction of interoperable publications within the Monitor is intrinsically tied to their accessibility. Since we systematically fetch PDFs, any publication that is accessible through this process is also considered interoperable. In essence, the minimum threshold for interoperability is met when a publication's full text is accessible in a machine-readable format through the PDF aggregation system.

3.6 Additional Aspects

In this section, we outline the construction methodologies for additional metadata elements requested. While the subsequent Data Evaluation section delves into a comprehensive analysis of all key elements, here we focus solely on those requiring construction methodologies that have not been previously addressed.

Corresponding author affiliation: In identifying the corresponding author's affiliation within the Graph, we encounter limitations due to the lack of explicit tagging for this role in metadata from integrated data sources. We have implemented the following methodologies to address this gap.

[47] We recognize that additional criteria can be incorporated to construct indicators for these FAIR principles.

- Contributor Rank Analysis: This method involves identifying the corresponding author based on their position in the author list. Typically, we consider the first author as the corresponding author if the list is not alphabetical. However, sequencing of authors is rarely represented in metadata, limiting the effectiveness of this approach. Using this method, we have identified the first author in only **0.04%** of Irish peer-reviewed publications.
- Text Mining: (ongoing) Specifically for Irish publications, text mining is employed on PDFs to discern the corresponding author's affiliation. This method is applicable only for OA publications where PDFs are available. Of the PDFs received, a fraction could not be processed. Among the processed PDFs, only about **8%** mentioned "corresponding" or "correspondence" on the first page, that is **1.5%** of all OA peer-reviewed publications. Future steps include employing advanced tools for PDF processing that utilize machine learning and computer vision. Additionally, we plan to search for alternative tags that might indicate the corresponding author, such as "mailto" or publisher-specific tags.
- IReL OpenAPC Dataset[48]: The dataset, which encompasses publications published under Transformative Agreements, includes valuable information on the institutions of corresponding authors. It contributes to our analysis with circa **6K** publications, representing **1.8%** of Irish peer-reviewed publications.

Upon achieving adequate data representation, we will incorporate corresponding author information into the Monitor.

Publicly-funded: To identify Irish scholarly publications that align with the definition of "publicly funded research as research undertaken in whole or in part via publicly funded resourcing or remuneration, e.g., salaries, grants, contracts, etc.,"[49] several steps are underway:

1. We used the responses for the publicly funded RPOs/RFOs from the National Open Access Monitor Survey conducted by IReL.
2. We also utilised OFR's metadata to identify publicly funded RFOs with the "Government" type. OFR's metadata offers pertinent information on funder type which is being integrated in the Graph.

Table 21: # and % of Irish Publicly-Funded Publications

Irish publications
423,893

[48] https://github.com/OpenAPC/openapcde/tree/master/data/Transformative_agreements/IReL

[49] National Action Plan, page 12 https://doi.org/10.7486/DRI.ff36jz222

and % of Irish publicly-funded Publications
254,489 (60%)

Implications of Definition: According to the tender's criteria, publications from RPOs with any government funding are considered publicly-funded. This definition influences the Monitor's data by marking these publications as publicly-funded across all related RPO profiles, irrespective of direct public funding received by each institution.

4 Data Evaluation & Challenges

This section presents a critical assessment of various metadata elements and their associated challenges. The findings outlined here are instrumental in understanding the current landscape of data quality and integrity, which are pivotal in shaping effective Open Access monitoring and policy development.

Key Highlights

The following key issues, areas for improvement have been identified.

Affiliation and Author Identification:

- Variability in institutional naming conventions which necessitates improvements in the use of PIDs to mitigate duplicates and inaccuracies.
- Requirement for broader adoption of ORCID iDs among researchers to facilitate more reliable author identification and connect scholars with their work effectively.

Funding Transparency and Publication Links

- Lack of data linking publications to their respective funding sources and projects and need for standardized reporting practices to improve the reliability of funding information and project affiliations within metadata.

Open Access Compliance and Publication Versions

- Clearly distinguishing between different publication versions (e.g., pre-print, post-print) to comply with OA mandates.
- Tracking changes in the OA status of journals over time to accurately categorize publications and support compliance efforts.

Accessibility and Licensing

- Regular need for licence normalization (different versioning) and for understanding the permissiveness of publisher licence and other non-CC licences.
- Including a valid URL to a full text of an OA publication.

Quality Assurance and Timeliness

- Verifying the peer-reviewed status of publications to uphold scholarly integrity.
- Tackling the indexing delay for recent publications to maintain the relevance and timeliness of the research monitoring system, including exposing the information of when a publication was first deposited in a repository.

Equity and Financial Considerations in Publishing

- Addressing the need for equitable representation across all research disciplines in metadata records.
- The extreme lack of data on APCs which is key for understanding the financial aspects of OA publishing.

Our examination underscores the critical areas for improvement in the metadata of Irish peer-reviewed publications. Collaborative efforts across the research community, including standardization of metadata practices and embracing technological solutions, will be key to achieving these goals. Through such enhancements, we can ensure that Irish research contributions are accurately captured, easily accessible, and fully leveraged to advance the frontiers of knowledge and Open Science.

In the following table, we will delve deeper into each metadata element, exploring the specific challenges and implications in detail, setting the stage for the subsequent recommendations and solutions.

Table 22: Metadata Analysis

Metadata element & Issue	Quality	Relevance & Challenges
Affiliations: Identification and association of research publications to Irish RPOs	The OpenAIRE Graph currently includes **330,174** peer-reviewed publications affiliated with Irish RPOs. Efforts to ensure high-quality data include advanced source	Accurate affiliations are essential for tracking Ireland's research output. **48.4%** of Irish publications are from OpenAIRE's harvested institutional data sources (repositories, CRIS, OA journals), emphasizing the role of repositories and journals in data integrity. The rest, derived from external metadata (such as Crossref), text mining,

Metadata element & Issue	Quality	Relevance & Challenges
	ingestion, deduplication processes, and continuous data validation[50].	and OpenAIRE EXPLORE's claim and link functions[51] highlight the need for diverse data inputs to capture the full scope of Irish research activities. The primary challenge in identifying affiliation links is that many data sources do not provide this information in their metadata or provide strings in place of PIDs describing the affiliations.
RPOs Name: Accurate Identification of RPO names and all their variants	The OpenAIRE Graph currently lists **8001** Irish RPOs, suggesting duplicates due to naming variations.	Accurately identifying RPOs is vital for correctly linking research outputs to specific Monitor profiles. The challenge lies in the diversity of naming conventions, which can lead to duplicates and misidentification. OpenAIRE addresses this by employing PIDs and other deduplication methods. Additionally, the OpenOrgs tool, available to primary RPO dashboard managers, plays a crucial role in refining RPO profiles, ensuring accurate labelling of departments and affiliated entities. The complexity of managing and standardizing these variations represents an ongoing challenge in maintaining data integrity.
Funded publications: Establishing comprehensive funder-publication links.	While SFI (Science Foundation Ireland) is curated due to having joined OpenAIRE, for other RFOs, we rely on OFR harvested metadata. Its quality is difficult to assess externally.	Establishing clear links between funders and publications is crucial for accurate research tracking and measuring of OA compliance to various funder mandates. Funders that join OpenAIRE provide a list of projects, enabling us to develop dedicated text mining algorithms. These algorithms achieve near-perfect precision in identifying project-publication links and improve the quality of harvested metadata, enhancing the accuracy of funder-publication connections in the OpenAIRE Graph[52].
Grant award ID: Ensuring availability & establishing project-publication links.	For SFI-funded projects, there is full coverage of Grant Award IDs, enabling the inference of project-publication links. Other funders in OFR that have not joined OpenAIRE currently lack these specific links, offering	Project-publication links are crucial for assessing compliance and providing detailed insights for funders. The main challenge is the lack of uniform reporting standards among funders, complicating text mining and metadata harvesting for precise project-publication connections. OpenAIRE is developing dedicated algorithms for the Irish Research Council, leveraging their publicly available project metadata.[53]

[50] The latest version of the Graph includes over 79 million affiliations (>140Mi in total when counting multiple affiliations) out of 242 million research outputs (publications, datasets, software, other research products).

[51] Both available in the Monitor as well.

[52] https://www.openaire.eu/funders-how-to-join-guide

[53] We anticipate this enhancement to be reflected in the Monitor by March 2024.

Metadata element & Issue	Quality	Relevance & Challenges
	only funder-publication connections in the harvested data.	
Version of the Publication: Differentiating between the various publications versions such as pre-print, Version of Record (VoR), Author Accepted Manuscript (AAM) or post-print.	The versioning of a publication is rarely included as a distinct metadata element exposed by repositories, and the sources that do provide it offer limited coverage.	Accurately distinguishing between publication versions is crucial for compliance with RFO mandates that usually require deposition of the AAM, VoR, or post-print. OpenAIRE Guidelines (https://openaire-guidelines-for-literature-repository-managers.readthedocs.io/en/v4.0.0/field_resourceversion.html) incorporate this metadata element with a separate recommended rule "Resource Version" utilising the COAR Version Types Vocabulary (http://vocabularies.coar-repositories.org/documentation/version_types/).
PIDs: Identifying and ensuring consistent coverage of Persistent Identifiers (PIDs) for publications, organisations, and authors.	**85.6%** of Irish peer-reviewed publications Irish peer-reviewed publications currently have a PID. ORCID iDs are present in almost **57,6%** of these publications.	PIDs are critical for clearly identifying research outputs and contributors, enhancing data findability and interoperability. A key challenge, lies in the adoption ORCID iDs. Although there is significant progress, around **30%** of Irish peer-reviewed publication still lack an ORCID iD. Additionally, the integration of PIDs across various platforms and databases often faces hurdles due to different standards and systems. Actively using ORCID's sync and claim functionalities of the Monitor can help authors improve PID coverage, but broader adoption and standardization across the research landscape are necessary for achieving comprehensive PID integration[54].
Peer-Reviewed publications: Verifying peer-review status of Irish publications	**78.7%** of publications from Irish RPOs and RFOs are identified as peer-reviewed. OpenAIRE is assessing this quality through continued data examination, including manually curated entries for peer-reviewed journals and	The verification of peer-reviewed status is crucial for maintaining the integrity and scholarly value of the research outputs within the Monitor. This ensures that vetted and quality-assured publications contribute to Ireland's research profile. A major challenge is the lack of a standardized 'peer-reviewed' identifier in metadata, complicating the verification of publications' scholarly status[55].

[54] More information on PIDs in the OpenAIRE Graph is available at https://graph.openaire.eu/docs/data-model/pids-and-identifiers

[55] For detailed information on the methodology used for determining peer-reviewed status, please refer to the Section 3.5.

Metadata element & Issue	Quality	Relevance & Challenges
	conference proceedings.	
Year of publication: Low(er) coverage for 2022	For 2022, the OpenAIRE Graph includes **15K** peer-reviewed Irish publications, compared to **22K** in 2020 and **21K** in 2021.	Timely inclusion of the latest publications is crucial for accurately reflecting current research trends. A key challenge is the inherent delay in publication and indexing. Research often experiences a lag between completion, publication, and eventual indexing in databases. This delay typically results in lower coverage for the most recent year, affecting the immediacy and comprehensiveness of a monitoring system.
Publisher & Journal: Challenges in deduplication	Due to variations in naming conventions, abbreviations, and metadata inconsistencies, the same publisher or journal may appear as distinct multiple entries.	Accurate and consistent representation of publishers and journals is essential for reliable bibliometric analysis. Inconsistencies in naming can lead to duplicated entries, impacting the accuracy of data and research metrics.[56]
Licence I: Challenges in deduplication and interpretation of licence agreements	Approximately **98%** of licences found in metadata of Irish peer-reviewed publications have been successfully grouped and normalized, but categorizing non-CC (Creative Commons) licences remains a challenge.	Accurate categorization and interpretation of licences are vital for understanding the usage rights and restrictions of publications, which directly impacts researchers' ability to reuse and disseminate knowledge. The legal complexities and variety of non-CC licences pose significant challenges in standardizing this metadata. Efforts to clarify and correctly categorize these licences are crucial for ensuring that users have clear and correct information regarding the accessibility and reuse conditions of scholarly works.
Licence II: Incomplete Licensing in OA Metadata	**30.8%** of OA peer-reviewed publications in the Monitor are not associated with a specific licence.	The challenge of missing licences significantly impacts the ability to determine the true extent of Open Access in Ireland and assess its adoption among various stakeholders. Without clear licensing, it is not possible to confirm a publication's Open Access status, which is essential for comprehensive OA monitoring and for guiding data-driven decisions within the OA landscape.
FoS: Coverage and potential under-representation	The FoS coverage for peer-reviewed publications in the OpenAIRE Graph is	Comprehensive FoS coverage is vital for equitable representation of all research disciplines. In fields like Humanities and the Arts, and Social Sciences, a key challenge is discerning whether the lower

[56] OpenAIRE's ongoing enhancements, including the refinement of algorithms using PIDs, aim to address these duplication (see Section 3.3)

Metadata element & Issue	Quality	Relevance & Challenges
of specific disciplines	**77.8%,** with **12.9%** from Humanities and the Arts, and Social Sciences. Differentiating community practices from actual research output is a key challenge in accurately assessing disciplinary representation.	representation is due to variation in publication norms, to variation in funding support or to digital practices uptake.
Access rights I: Missing data	**16.9%** of Irish peer-reviewed publications do not include access rights	Comprehensive access right coverage is key for monitoring Open Science uptake and progress.
Access rights II: changing over time, such as an Embargoed publication becoming OA.	To the best of our knowledge, the original access rights of a publication deposited in a repository is not a metadata element that repositories expose.	Monitoring the changes in access rights, such as an embargoed publication transitioning to OA, is crucial for compliance tracking and understanding the evolving accessibility of research outputs. A primary challenge is the absence of initial access right details in repository metadata, making it difficult to track the evolution of access status. The Monitor's approach of maintaining historical snapshots, including monthly deposits in Zenodo and visible indicator aggregates in the portal, tries to address this challenge by enabling the observation of changes in access rights over time. However, capturing the complete trajectory of each publication's accessibility remains complex due to the initial data gaps.
URLs to PDF full texts: Coverage and validity of URLs linking to PDFs in publication metadata.	Of the **188K** OA peer-reviewed publications in the Graph, **45%** have valid URLs in their metadata, with **33%** of all OA peer-reviewed publications) linking directly to full-text PDFs.	Accurate URLs linking to PDF full texts are essential for ensuring metadata integrity and the true accessibility of OA publications. A key challenge here is the inconsistent or incorrect reporting of these URLs. Many publications either lack direct links to full-text PDFs or have URLs that lead to outdated or inaccessible pages. This inconsistency hampers the effectiveness of metadata in providing direct access to research content.
Corresponding author affiliation: Determining the institution of the corresponding author.	Currently, the reporting of corresponding author affiliations in the metadata is not widespread. Efforts involving text mining of full-text PDFs and the use of the	Accurately identifying the corresponding author's affiliation is crucial for bibliometric analyses and understanding research collaboration networks. The major challenge lies in the inconsistent or incomplete reporting of corresponding author details in publications. This inconsistency hampers efforts to automate identification and validation processes, affecting the overall data quality.

Metadata element & Issue	Quality	Relevance & Challenges
	OpenAPC IReL dataset[57] have resulted in coverage of approximately **2%** of all peer-reviewed publications[58].	
APCs: Lack of data availability	Of the **83,614** peer-reviewed publications that incurred an APC, only **1194** have APC data from OpenAPC (**1.4%**).	Understanding the financial aspects of OA publishing is crucial for policy makers and institutions. The lack of shared APC data by publishers limits the ability to conduct comprehensive economic analyses of OA publishing, impacting budgeting and policy decisions.
Historical OA status of journals: journals changing OA status	Historical journal data is typically not available.	Accurately tracking the evolution of journals from subscription-based to hybrid or fully Open Access is vital for correctly categorizing articles. The current lack of historical data leads to potential misclassification of articles based on the journal's present status (e.g. diamond OA), affecting the accuracy of OA categorization and skewing analyses of OA trends and policy development.

This detailed analysis of metadata elements and their associated challenges provides a comprehensive snapshot of the current state of data quality and integrity for Irish peer-reviewed publications, and establishes a **valuable benchmark** for future assessments. By identifying specific areas of concern and highlighting key issues, we have laid a solid foundation for measuring progress and guiding improvements. This benchmark serves as a critical reference point, enabling us to monitor advancements in metadata management and the effectiveness of OA monitoring over time. As we move forward, the insights gained from this analysis will inform the next section of the report, focusing on targeted recommendations and solutions. These will address the challenges identified, paving the way for enhanced data quality, improved standardization, and more effective Open Access practices.

[57] The dataset includes data on publications published under Transformative Agreements including the institution of the corresponding author.
https://github.com/OpenAPC/openapcde/tree/master/data/Transformative_agreements/IReL

[58] Refer to Section 3.6 for more details.

5 Recommendations & Solutions

In this section, we delve into a series of strategies and recommendations aimed at enhancing OA monitoring in Ireland. Our approach unfolds across three interconnected tiers of improvement.

First, we explore direct improvement strategies, focusing on actions and tools **within the Monitor** itself that leverage its functionalities for immediate enhancements in data quality and user engagement. Following this, we address **indirect improvement strategies which centre on elevating the quality of data being harvested into the OpenAIRE Graph.** These recommendations target stakeholders contributing data, with the goal of ensuring that the information feeding into the Monitor is precise, comprehensive, and up-to-date.

Concluding our approach, we examine **long-term monitoring solutions** and suggested workflows. These encompass broader, sustainable improvements that extend beyond immediate solutions, involving policy development, collaborative efforts among stakeholders, and embracing technological advancements for future challenges.

By adopting these strategies, stakeholders from various sectors of the Open Access ecosystem can collaboratively foster a more robust, accurate, and efficient monitoring system. These recommendations are designed not just to tackle current challenges but also to establish a groundwork for ongoing enhancement and adaptability in the ever-evolving domain of research and data management.

5.1 Direct Improvement Strategies for the Monitor

To immediately and effectively enhance the quality and functionality of the Monitor, a range of direct strategies can be implemented. **These strategies leverage existing tools and functionalities within the OpenAIRE ecosystem, facilitating tangible improvements in data management and user interaction.** The following table outlines specific actions that various stakeholders can undertake. These actions are designed to directly address the challenges identified in our metadata analysis, improving data accuracy, completeness, and the overall user experience. By engaging with these strategies, stakeholders can contribute to the refinement and advancement of the Monitor, enhancing its value as a tool for Open Access monitoring and research analytics.

Table 23: Direct Improvement Strategies for the Monitor

Recommendation	Publishers	RPOs	RFOs	Researcher	Institutional Repository Managers
Organisation Deduplication via OpenOrgs: regularly deduplicate RPO names and manage department/affiliated entity relationships[59].		X	X		
Join OpenAIRE: RFOs can join OpenAIRE by providing their project data for customized text mining and metadata enrichment[60].			X		
Registration to OpenAIRE PROVIDE: Register institutional data sources to OpenAIRE PROVIDE, follow the latest version of the OpenAIRE Guidelines, and use the Metadata Validator and Broker services for enhanced metadata quality[61].					X
Register with ORCID and Sync to Monitor: Register with ORCID and synchronize your ORCID and with Monitor accounts to keep both records current[62].				X	
Use Linking Functionality in OpenAIRE: Use Linking functionality to enrich connections within the Graph by linking research products to other researcher products, research communities or projects[63].	X	X	X	X	X
Engage with Monitor Dashboard & Provide Feedback: Interact with the Monitor dashboards and provide data quality feedback for continuous improvement.	X	X	X	X	X
Attend Engagement and Training Events: Participate in engagement and training events to stay informed about best practices and updates[64].	X	X	X	X	X
Reach Out for Help & Guidance: Contact the Monitor team for any issues or requests[65].	X	X	X	X	X

[59] Users who wish to perform this section need to apply to be the primary dashboard manager of their organization by completing this form https://app.onlinesurveys.jisc.ac.uk/s/maynoothuniversity/national-open-access-monitor-dashboard-manager-application-form.

[60] https://www.openaire.eu/funders-how-to-join-guide

[61] https://www.openaire.eu/validator-registration-guide

[62] For details select "Researchers" in this page https://oamonitor.ireland.openaire.eu/how-it-works/the-5-monitors#tabs-content

[63] https://oamonitor.ireland.openaire.eu/how-it-works/user-actions#linking

[64] https://oamonitor.ireland.openaire.eu/engagement-training

[65] https://oamonitor.ireland.openaire.eu/contact-us

5.2 Indirect Improvement Strategies

The effective management and improvement of metadata within the OpenAIRE Graph require concerted efforts from all stakeholders in the research ecosystem. Recognizing the challenges identified in our metadata analysis, the following table presents a series of indirect strategies. **These are aimed at enhancing the overall quality and accuracy of data harvested into OpenAIRE.** By addressing key areas such as timely reporting, accurate metadata entry, and clear licensing, each stakeholder group plays a critical role in this process, and the recommended actions outlined below are tailored to leverage their unique positions and capabilities.

Table 24: Indirect Improvement Strategies for the Monitor

Stakeholder	Recommendation
Policy Makers	Develop and enforce policies that mandate the registration and timely deposition of research outputs in national and international repositories, emphasizing the importance of accurate and standardized metadata. Develop and advocate for discipline-specific OA policies and support mechanisms.
Publishers	Implement consistent metadata standards across publications, particularly for identifying peer-review status, version of record, corresponding authors, and licensing information. Encourage open sharing of publication metadata to aid in more accurate data harvesting and indexing. Proactively share publication data, including APCs, BPCs, historical access rights of journals, and full-texts where possible to enhance the comprehensiveness and accuracy of the data available. Foster a culture of transparency and openness in licensing, ensuring clear communication of usage rights.
RFOs	Mandate and guide beneficiaries on accurate and timely reporting of project outcomes and publications, emphasizing the use of PIDs and standard metadata formats in their submissions. Advocate for and implement policies that require transparency in APC and BPC data from funded publications.
RPOs	Regularly review and update publication records in institutional repositories, focusing on the timeliness of deposits. Encourage researchers to use standard naming conventions and persistent identifiers (PIDs) for affiliations. Report APCs/BPCs to OpenAPC.

Stakeholder	Recommendation
Researchers	Consistently use ORCID iDs in all research outputs, ensure timely updates of personal profiles in research databases, and accurately report affiliations and funding sources in publications.
Institutional Repository Managers	Adopt and adhere to the latest OpenAIRE guidelines for metadata.
Repository Managers (non-institutional)	Develop and maintain robust systems for metadata standardization and quality control. Ensure efficient and accurate handling of diverse research outputs, promoting consistency and completeness in metadata. Collaborate with primary data providers and other repositories to streamline data integration processes and jointly address metadata challenges. Adapt repository systems to accommodate and reflect historical changes in journal OA status. Implement, if not already present, automatic licensing functionalities in repositories.
All Stakeholders	Participate in training and engagement events to stay informed about best practices in metadata management and Open Science policies. Actively engage in community discussions to address common challenges and share solutions.

5.3 Long-Term Solutions

Shifting the focus to the future, this section outlines long-term solutions and workflow recommendations essential for the sustained advancement of OA monitoring. Drawing from the insights gained in our detailed metadata analysis, we propose strategies that address current challenges and try to anticipate evolving trends in research and data management. These recommendations are crafted to ensure the Open Access ecosystem is well-equipped to adapt, grow, and maintain high standards of data integrity and usefulness in the years to come.

Sustainable Data Management and Policy Enhancement (Policy Makers, RFOs, RPOs)

- Develop/Adopt a framework for iterative improvement in metadata standards, building on identified metadata inconsistencies, to ensure relevance and effectiveness in capturing evolving research outputs.
- Advocate for policy harmonization across international platforms, focusing on standardizing metadata formats and PIDs, as highlighted by the analysis of ORCID iD coverage and publication data discrepancies.

- Promote global policy initiatives that resonate with the FAIR principles, facilitating a unified approach to OA data management.

Long-Term Data Preservation and Accessibility (Repository Managers, RPOs, and Data Curators)

- Implement strategies that adhere to the FAIR principles while considering future research artifact accessibility needs, accounting for the rapid evolution of research dissemination methods.
- Establish or enhance national and international repositories to comply with preservation standards and adapt to changing formats.

Routine Data Quality Audits and Collaborative Correction Processes (All Stakeholders):

- Institutionalize regular data audits to correct inconsistencies and gaps in metadata, such as affiliation, funding information, and licensing details.
- Create collaborative platforms for stakeholders to participate in metadata correction and enrichment, sharing expertise and resources.

Training, Capacity Building, and Feedback Loops (Policy Makers, Repository Managers RPOs/Librarians, Research Community Leaders):

- Conduct targeted training programs on metadata management areas identified as lacking, like licence reporting.
- Develop adaptive learning resources and feedback mechanisms to keep stakeholders updated with best practices and allow them to contribute to continuous metadata improvement.

By aligning these long-term solutions and workflow recommendations closely with the specific challenges and gaps identified in our metadata analysis, we ensure a data-driven and targeted approach addressing current needs in a long-term, sustainable manner.

6 Conclusion

The baseline analysis of the OA landscape in Ireland has revealed a complex and evolving picture. We observe significant strides toward embracing Open Research practices, alongside areas that necessitate further attention and development. This conclusion synthesizes our key findings and reflections, offering a cohesive understanding of where Ireland stands with respect to Open Access uptake.

The scholarly production in Ireland, with a high proportion of peer-reviewed publications, underscores a commitment to academic rigor and quality. The growth in scholarly output over the years is indicative of a vibrant research environment, increasingly adopting digital practices such as DOIs and digital repositories.

In the realm of OA, while a considerable number of Irish publications are Open Access with licensing, a significant portion remains either without a licence or under Closed Access. This situation reflects the ongoing transition in the Irish academic community toward fully Open Research practices. The disparity in licensing across different publication types underscores the need for a more consistent approach to OA, particularly in publishing models like books and conference proceedings.

Publisher-mediated OA, encompassing Gold and Hybrid models, dominates the OA landscape in Ireland. This trend highlights the significant role of publishers in shaping OA practices. However, the growing presence of repository-mediated OA points to an evolving landscape where institutional and thematic repositories are gaining importance as complementary platforms for research dissemination.

The analysis of Plan S compliance has shown a predominant reliance on Gold OA with APCs. However, the increasing uptake of Transformative Agreements and Journals, particularly in Social Sciences and Humanities and the Arts, suggests a strategic shift towards different OA models in these fields. This shift indicates a nuanced approach to OA adoption, reflecting the diverse needs and dynamics of various research disciplines.

Regarding APCs, the average cost per publication provides a baseline understanding of the financial aspect of OA publishing. However, the limited coverage of APC data underscores a need for greater transparency and more comprehensive reporting in this area. The variability in APCs across different scientific disciplines suggests differing financial pressures and publishing cultures in these fields.

Finally, the examination of FAIR principles and metadata completeness has brought to light the mixed landscape of licensing practices, with a blend of restrictive and open licences. The widespread use of PIDs like DOIs and the increasing adoption of ORCID iDs are positive indicators of the Irish research community's move towards more streamlined and efficient scholarly communication. However, the quality and completeness of metadata remain areas for improvement to fully realize the benefits of open and accessible research.

The recommendations and solutions outlined in this report, spanning direct improvement strategies (via the Monitor functionalities available), indirect approaches, and long-term solutions, are aimed at fortifying OA monitoring and fostering a more robust, accurate, and efficient system. These strategies, derived from our analysis, offer practical steps for various stakeholders in the research ecosystem to collaboratively enhance the OA landscape in Ireland.

As we conclude this report, we reiterate the importance of ongoing efforts to refine and advance OA practices. The collective commitment of policymakers, publishers, researchers, RPOs, RFOs, repository managers and other stakeholders will be crucial in navigating the evolving challenges and opportunities in the world of scholarly communication.

7 Appendix

Institutional Repository Harvesting for the Monitor

(continued from Section 3.2)

In the first phase of the Monitor's development, attention was directed towards Institutional Repositories (IRs). These repositories were characterized using information from OpenDOAR, FAIRSharing registries, and those identified in the National Open Access Monitor Survey: Organisational Identity. **8** institutional repositories were identified that had not been registered with OpenAIRE. Subsequently, we initiated the registration process and harvested their metadata records. Additionally, transformation rules were adjusted for each repository to custom transform metadata records, ensuring alignment with OpenAIRE Guidelines. The primary focus of these efforts was on crucial fields for the harvesting process and, where applicable, the identification of publications.

- PID
- Title
- Author
- Publication date
- Resource Type
- Access Rights

We successfully retrieved metadata from **7** of these repositories. However, one repository, identified as Obsolete in the gap analysis conducted by OpenAIRE in collaboration with the Irish NOAD in 2020 (Irish Health Publications Archive), was not harvested. Research@Thea, reported by two institutions in the survey (Technological University of the Shannon: Midlands Midwest and Atlantic Technological University), had **927** publications in its metadata records. Despite this, it is categorized as a Thematic Repository in OpenDOAR, providing metadata for multiple institutions. Additional refinement is necessary to register and harvest it into OpenAIRE. Affiliation matching for Research@Thea will be examined in the next phase, along with any other applicable repository types (such as ThematicPublication) Repositories.

The subsequent actions, in conjunction with the support of the "National Open Access Monitor Survey: Organisational Identity" and in collaboration with the Irish network of repositories, will involve the identification and registration or selective harvesting of repositories/CRIS that are currently not registered. Additionally, an ensuing phase will focus on enhancing compatibility with the OpenAIRE Guidelines, upgrading to version 3.0 or, preferably, version 4.0. Moreover, we will concentrate on outreach activities to communicate with institutions whose metadata records we have aggregated. The aim is

to provide them with information about the process and encourage them to align with the OpenAIRE Guidelines. Concurrently, based on the survey findings, several CRIS are currently in ongoing development. We plan to engage with these institutions, assessing their development progress, and establishing collaborations to ensure their inclusion in OpenAIRE and compliance with the OpenAIRE CRIS Guidelines.

Further, institutions can either directly register their OA journals in OpenAIRE or facilitate our own registration and harvesting process by supporting the OAI-PMH harvesting protocol. We note that institutional journals do not automatically imply that publications published in those journals are affiliated with the corresponding institutions.