National Open Access Monitor, Ireland
Draft Report

First Insights and Path Forward

October 9, 2023

Executive Summary

This draft report provides an initial assessment of Ireland's Open Access (OA) landscape, with a primary focus on a baseline analysis and data evaluation. It offers insights into the current state of OA in the country, emphasizing data quality and coverage, while also explaining the construction and methodological steps taken to construct the baseline analysis. In addition to these assessments, this draft also outlines the next steps for the OpenAIRE team and the optional strategies to be followed by key stakeholders to ensure accurate monitoring.

In the final version of the report, we will enhance this foundation by comprehensively addressing high-level infrastructure challenges. Our primary goal will be to provide practical, actionable recommendations for achieving continuous and robust OA monitoring in Ireland. The final report will offer a more extensive analysis, drawing upon additional insights gathered through further investigations.

Contents

List of Figures

List of Tables

Abbreviations

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	Field of Science
Graph	OpenAIRE Graph
GRID	Global Research Identifier Database
IReL	The consortium of Irish research libraries
ISNI	International Standard Name Identifier
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
ORCID	Open Researcher and Contributor ID
OS	Open Science
PID	Persistent Identifier
RFO	Research Funding Organisation
ROR	Research Organization Registry
RPO	Research Performing Organisation
SDG	Sustainable Development Goal
SFI	Science Foundation Ireland

VoR	Version of Record

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.[1]
OPEN ACCESS	We use 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."*[2]
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.
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*

[1] https://graph.openaire.eu/docs/data-model/entities/result

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

	Access content; and offset subscription income from payments for publishing services (to avoid double payments).[3]
OA TYPES/COLOURS[4]	
GREEN OA	*Green articles are published in toll-access journals, but archived in an OA archive, or "repository". These repositories may be discipline-specific (like ArXiv) or institutional repositories operated by universities or other institutions. Green articles may be published versions or preprints and can have any license or no license.* *Green articles will be distinguished based on whether they are accompanied by an Open License or lack a license altogether.*
HYBRID OA	*Hybrid articles are free to read at the time of publication, with an open license. These are usually published in exchange for an article processing charge, or APC.*
GOLD OA	*Gold articles have all the same characteristics as Hybrid articles, but are published in all-Open Access journals, which are in turn called "Gold journals", or just "OA journals".*
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. Bronze articles will not be identified as OA as this does not align with the Budapest Open Access Initiatvie definition but will be captured and tracked.*
ACCESSIBILITY – INTEROPERABILITY	
ACCESSIBLE	*A publication is accessible if the text file can be fetched via a valid URL in its metadata.*[5]

[3]https://www.coalition-s.org/transformative-journals-faq/

[4]https://support.unpaywall.org/support/solutions/articles/44001777288-what-do-the-types-of-oa-status-green-gold-hybrid-and-bronze-mean-

[5] 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.*

1 Introduction

The National Open Access Monitor, Ireland (Monitor from now on) stands as a pivotal initiative to comprehensively assess the state of Open Access (OA) research and scholarly publishing within the country. This report has been developed specifically to accompany the launch of the Monitor, offering stakeholders an in-depth understanding of the methodologies employed, the data evaluated, and the preliminary findings gleaned.

Our primary aim is twofold: to present a thorough baseline analysis of the current OA landscape in Ireland, and to articulate the data evaluation processes we have instituted. While this report offers a comprehensive initial assessment, the subsequent final report will delve into more extensive evaluations, addressing infrastructural challenges and drawing from broader insights.

2 Baseline Analysis

This section present OpenAIRE's initial baseline analysis of OA in Ireland. The analysis encompasses an examination of Ireland's present publishing output, both from an overarching perspective and at a domain-specific level. The FoS classification system[6] of OpenAIRE, serves as the basis for this domain-level assessment, categorizing publications into distinct domains[7].

The "Scholarly production" section provides contextual information, detailing the volume and nature of research outputs. In the "Open and Fair" section, the accessibility and fairness of disseminated research are explored, assessing the breadth of availability of research outputs. The "Plan S" section is dedicated to the evaluation of how Irish research publications align with the guidelines established by cOAlition S[8].

While this analysis offers a foundational overview, subsequent refinements to the data are anticipated. Brief indications of the measures for these refinements are provided, with a comprehensive evaluation set forth in the Data Evaluation section.

The figures provided in this report are based on the September version of the OpenAIRE Graph. In the final report, these numbers will be updated to reflect the data from the December version of the Graph. ***As of November, there are more than 400,000 Irish publications, of which approximately 300,000 are peer-reviewed****.*

[6] https://explore.openaire.eu/fields-of-science

[7] See Section 3.4 for a description.

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

2.1 Scholarly production

This section serves as a foundational exploration of scholarly production in Ireland, providing crucial context for the subsequent analysis. It offers an insightful overview of the country's peer-reviewed publications, including both their absolute numbers and a granular breakdown across various dimensions. By examining the distribution of publications over time, scientific disciplines, publication types, and data source types, we gain a comprehensive understanding of Ireland's research landscape. This comprehensive analysis lays the groundwork for a more in-depth investigation into the state of OA within the Irish scholarly community, ensuring that the subsequent sections are firmly rooted in the broader scientific context.

Table 1: Outline of overall production of Irish publications

Irish publications
345,316
 # and share of Irish Peer Reviewed Publications
244,156 (70.7%)
 # and share of Irish Open Access publications
226,025 (65.45%)
 # and share of FoS coverage
78,436 (22.71%)

We note that a substantial 70% of Ireland's research output consists of peer-reviewed publications, indicating a high level of academic and scientific quality.

We are currently working to improve the FoS coverage by applying our FoS classifiers to Irish publications.

Figure 1: Irish Peer Reviewed Publications over time

In this chart, we observe a consistent, gradual upward trend in peer-reviewed Irish publications over time. To delve deeper into this trend, the following two charts provide a detailed breakdown by FoS domains (FoS level 1) and topics (FoS level 2). This finer-grained perspective can shed light on the disciplines that boast larger numbers of peer-reviewed publications.

The delay in the availability of 2022 Irish peer-reviewed publications, as of October 2023, can be attributed to the natural timeline associated with academic publishing. OpenAIRE is actively addressing this delay by focusing on two main strategies: refining data harvesting from Irish institutional sources, enhancing text mining for affiliations specifically for Ireland and reviewing the methodology for establishing peer-review status.[9]

[9] By November 2023, we had doubled the affiliations for 2022 for the Irish scholarly production.

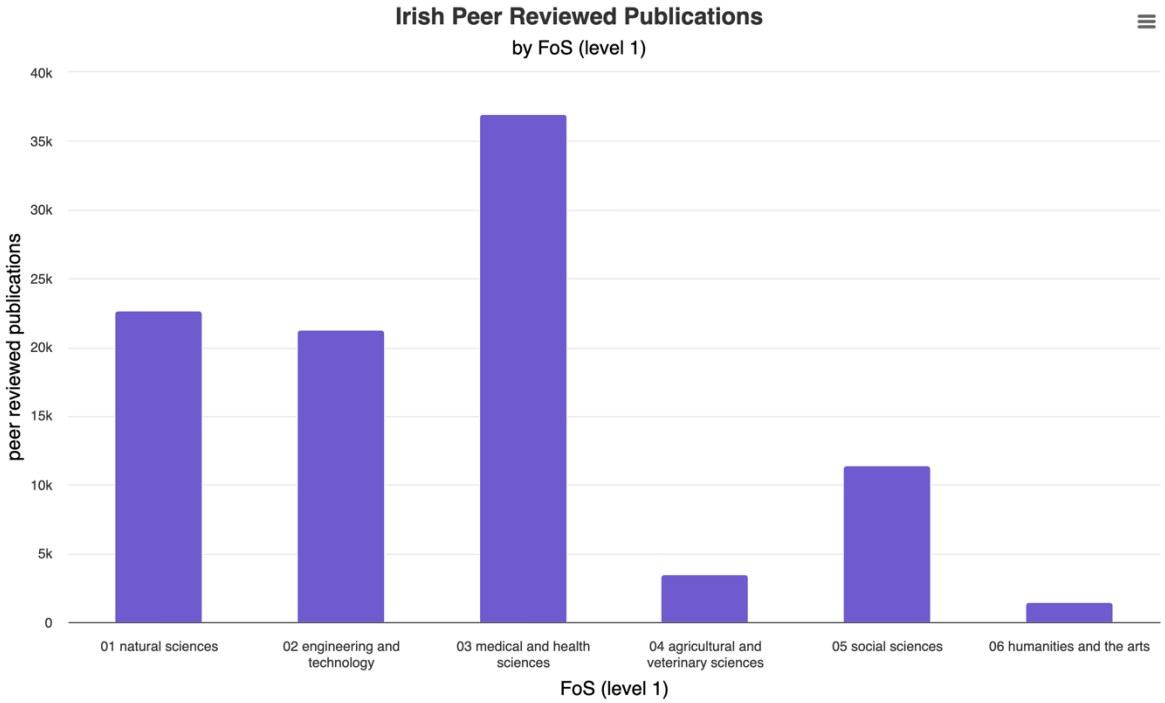

Figure 2: Irish Peer Reviewed Publications by FoS (level 1)

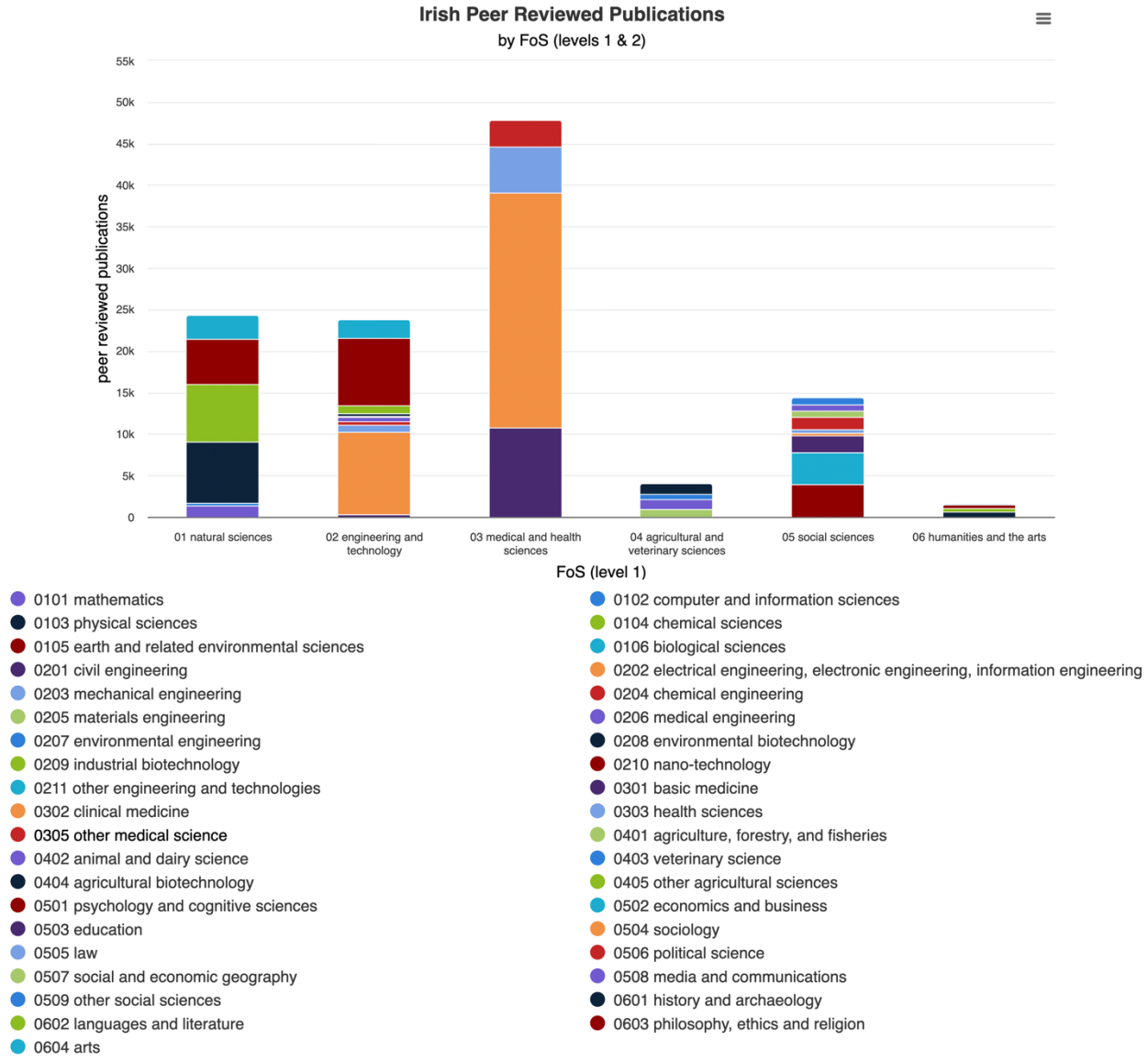

Figure 3: Irish Peer Reviewed Publications by FoS (levels 1 & 2)

These charts present the count of Irish peer-reviewed publications categorized by FoS domain (level 1) and topic (level 2). It becomes apparent that the Medical and Health Sciences domain, along with its Clinical Medicine and Basic Medicine topics, commands the highest numbers.

Table 2: Irish Peer Reviewed Publications by type

	# Irish peer-reviewed publications	Share of # Irish peer-reviewed publications
Article	*202,997*	*82.27%*
Book	*461*	*1.05%*

Conference object	*22,218*	*51.33%*
Contribution for newspaper or weekly magazine	*1*	*0.00%*
Dataset	*1*	*0.00%*
Part of book or chapter of book	*18,931*	*7.75%*
Unknown	*22*	*0.00%*

Within this table, it is evident that the predominant share of peer-reviewed publications comprises articles, with conference papers and books/book chapters following in sequence.

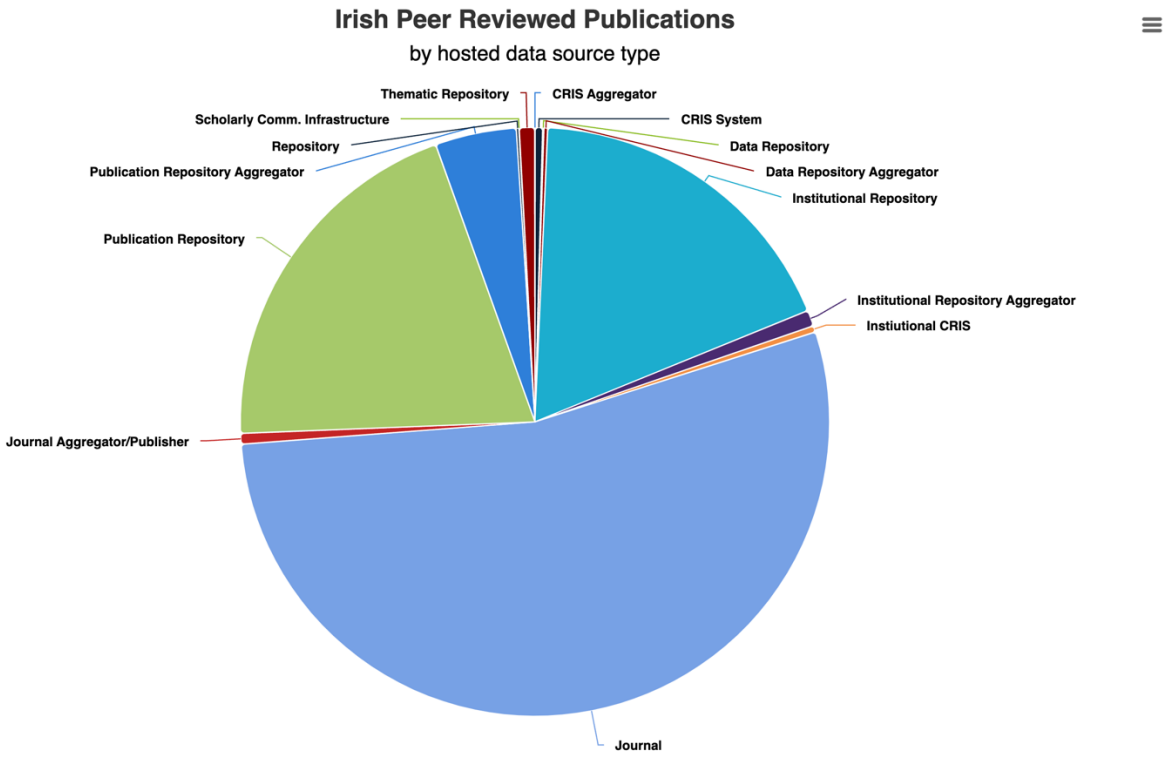

Figure 4: Irish Peer Reviewed Publications by hosted data source type

According to this chart, Journals serve as the primary data source for hosting the majority

of peer-reviewed publications, with Institutional and Publications Repositories coming in as the subsequent sources in line.

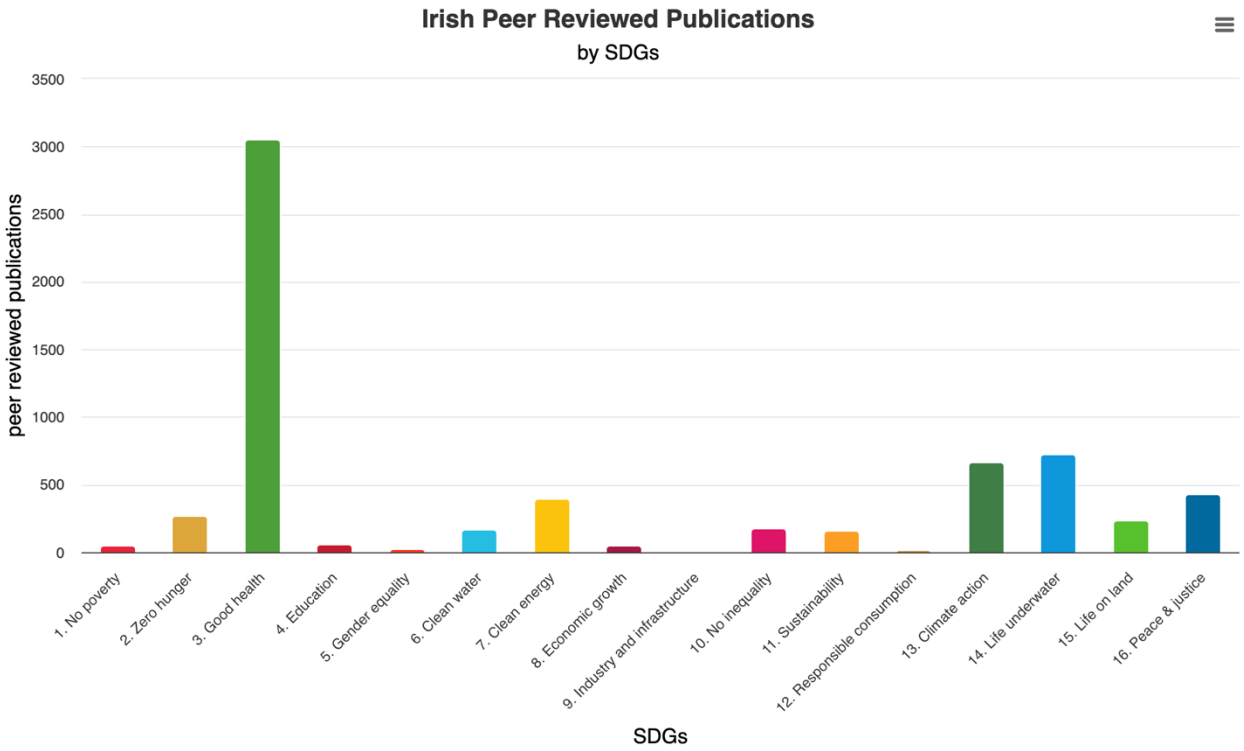

Figure 5: Irish Peer Reviewed Publications by SDGs

In the chart displaying the breakdown by Sustainable Development Goals (SDGs), one can discern a proxy for the societal relevance and potential impact of Irish peer-reviewed publications, with "Good Health" occupying the leading position, followed by "Life Underwater" and "Climate Action".

2.2 Open & FAIR

In the final version of the report, we will redefine our approach to align with the Budapest Open Access Initiative, which requires an open license for OA classification. Unlike our current method, where licensing is assessed separately, the revised analysis will include only licensed publications in the OA category. Publications without licenses, such as Bronze or Green without a license, will be discussed separately as part of unrealized OA.

In this section, we delve into a set of essential indicators that illuminate the openness and "FAIRness" of publications within Ireland's Open Access landscape. These indicators encompass various dimensions, including access rights, OA types/colours, embargoed access, licensing, persistent identifiers (PIDs), and the accessibility and interoperability of OA peer-reviewed publications.

By analysing these indicators, we gain valuable insights into the extent to which scholarly outputs are openly accessible and align with principles of FAIRness and transparency. 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. These indicators offer a comprehensive view of the multifaceted dimensions of openness and fairness in scholarly communication across domains, topics, publication types, and data source types.

OpenAIRE's PDF aggregation system has meticulously reviewed approximately 44K URLs extracted from the metadata of the 127K Irish OA peer-reviewed publications. Among these URLs, we have successfully retrieved 16K PDFs (~12,1%). Our ongoing efforts involve inspecting the remaining 28K URLs to further enhance our coverage.

2.2.1 Access Rights

The table that follows illustrates the distribution of Irish Peer-Reviewed Publications based on their access rights.

Table 3: Number and share of Irish Peer Reviewed Publications by Access Rights

Access Rights	# Irish peer-reviewed publications	Share of # Irish peer-reviewed publications
Open Access	*127,221*	*52.11%*
Embargo	*26*	*0.01%*
Restricted	*890*	*0.36%*
Closed Access	*70,534*	*28.89%*
not available	*45,485*	*18.63%*

To delve deeper into this trend, the following chart provides a chronological breakdown. We can observe an upward trend in the uptake of Open Access, with an increase of over 20% from 2014 to 2021.

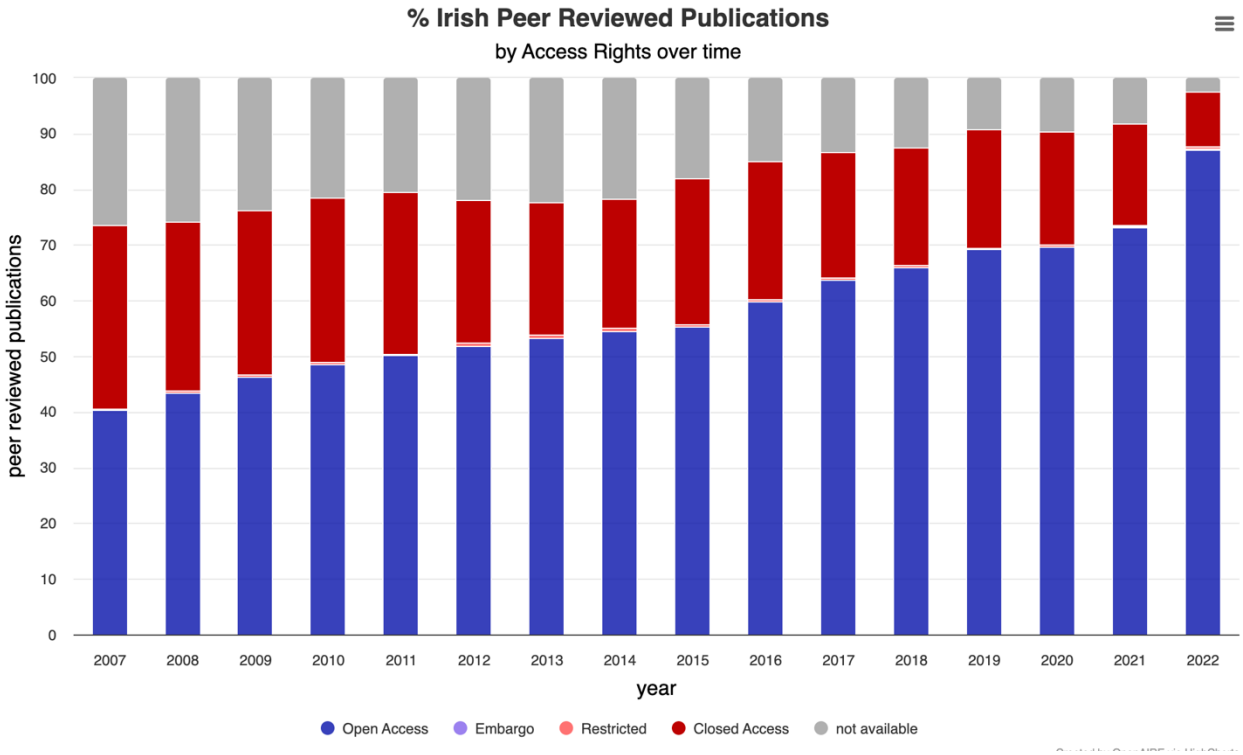

Figure 6: % Irish Peer Reviewed Publications by Access Rights over time

To further explore this trajectory, two charts that follow provide a detailed breakdown according to Field of Science domains (FoS level 1) and topics (FoS level 2). This finer-grained perspective can shed light on particular disciplines that might be influencing the overall trend. It offers insights for shaping future OA strategies, as exemplified by the increase in Open Access peer-reviewed publications within the Medical and Health Sciences domain, along with the Clinical Medicine topic.

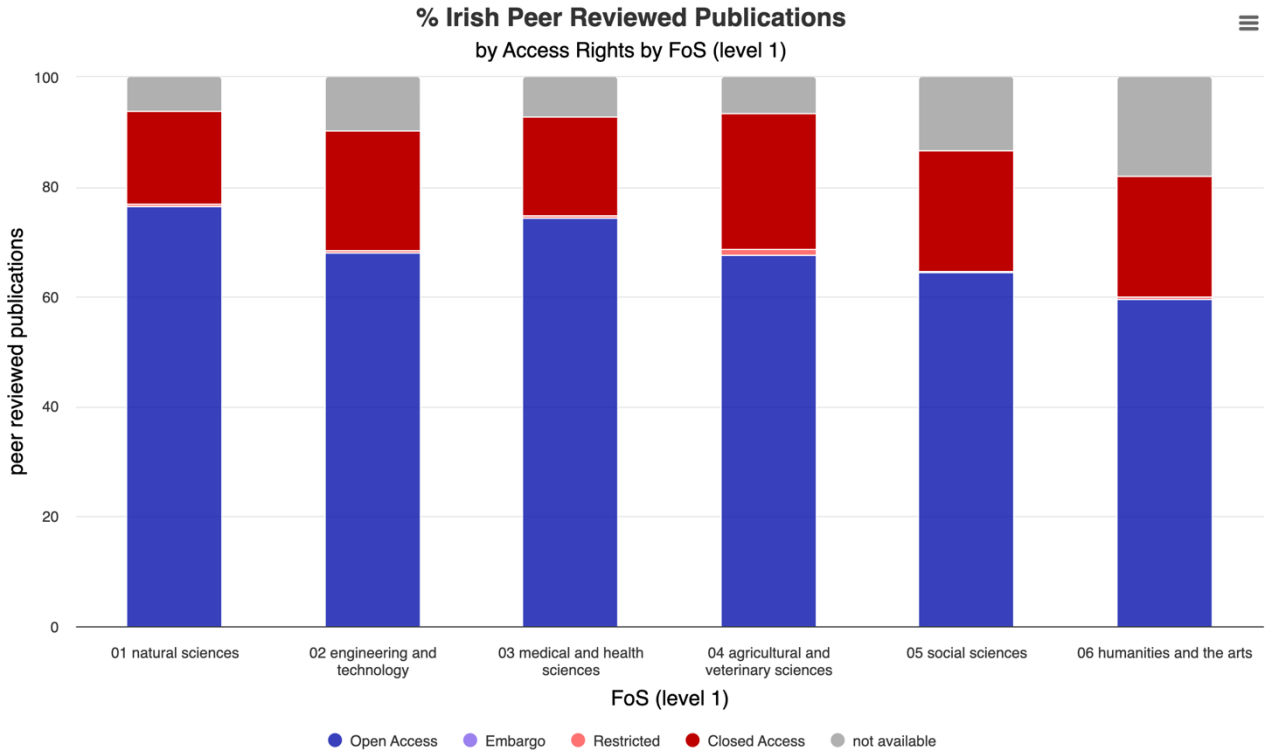

Figure 7: % Irish Peer Reviewed Publications by Access Rights by FoS (level 1)

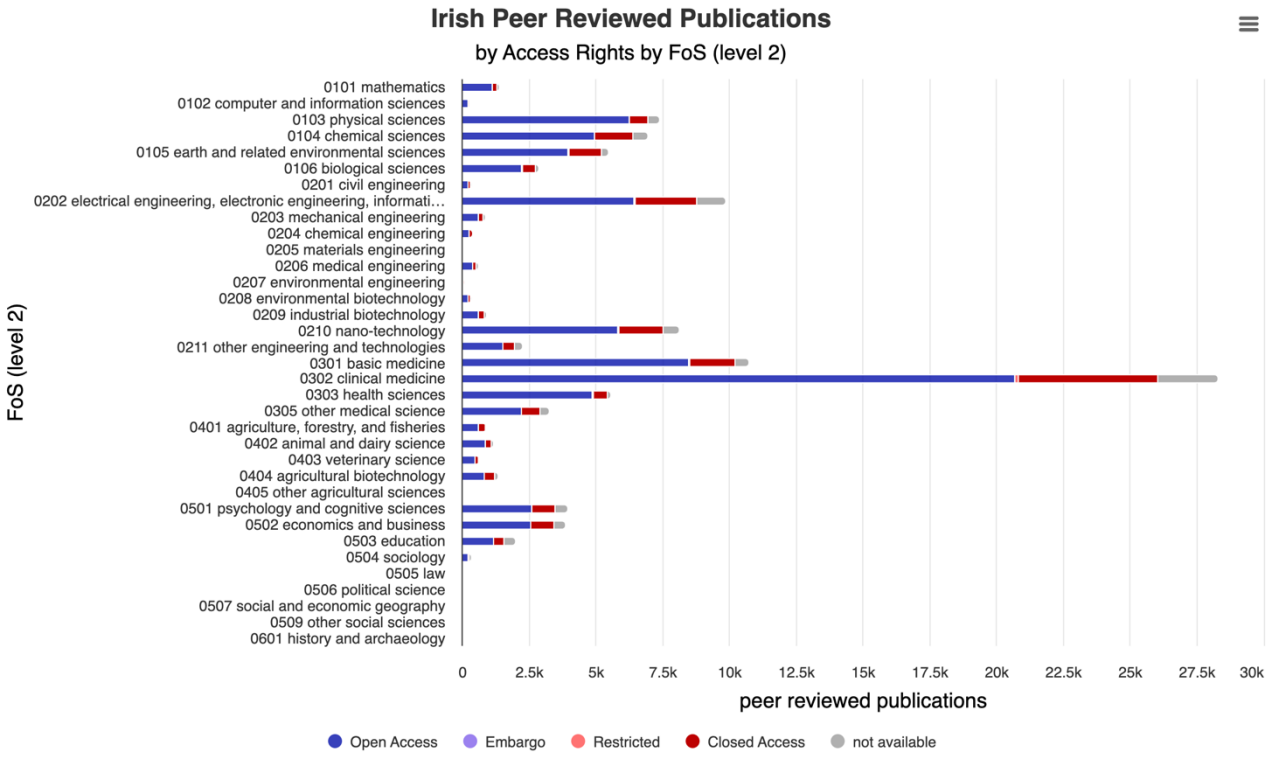

Figure 8: Irish Peer Reviewed Publications by Access Rights by FoS (level 2)

In this table, we find that the majority of Open Access peer-reviewed publications are articles, with conference papers and books/book chapters being the subsequent categories in order as a share of their overall production.

Table 4: Number of Irish Peer Reviewed Publications by Access Rights by Publication Type

	Open Access	Embargo	Restricted	Closed Access	not available
Article	*114,943*	*21*	*804*	*59,774*	*27,455*
Book	*145*		*5*	*75*	*236*
Conference object	*9,052*		*31*	*1562*	*11,573*
Contribution for newspaper or weekly magazine	*1*				
Dataset	*1*				
Other ORP type	*3*				
Other literature type	*1,742*	*1*	*15*	*327*	*21*
Part of book or chapter of book	*3,848*	*5*	*50*	*8,827*	*6,201*
Preprint	*4*				
Thesis	*2*				*3*
Unknown	*13*				

As indicated by the two charts presented below, Journals emerge as the foremost data source for hosting the bulk of Open Access peer-reviewed publications, closely followed by Institutional and Publications Repositories. Meanwhile, in the chart depicting the distribution by Sustainable Development Goals (SDGs), one can discern the societal

relevance of Irish Open Access peer-reviewed publications. Notably, "Good Health" claims the top spot, with "Life Underwater" and "Climate Action" and following suit.

In the final version of the report, these visualizations will also be displayed in terms of their percentage shares, providing additional insights.

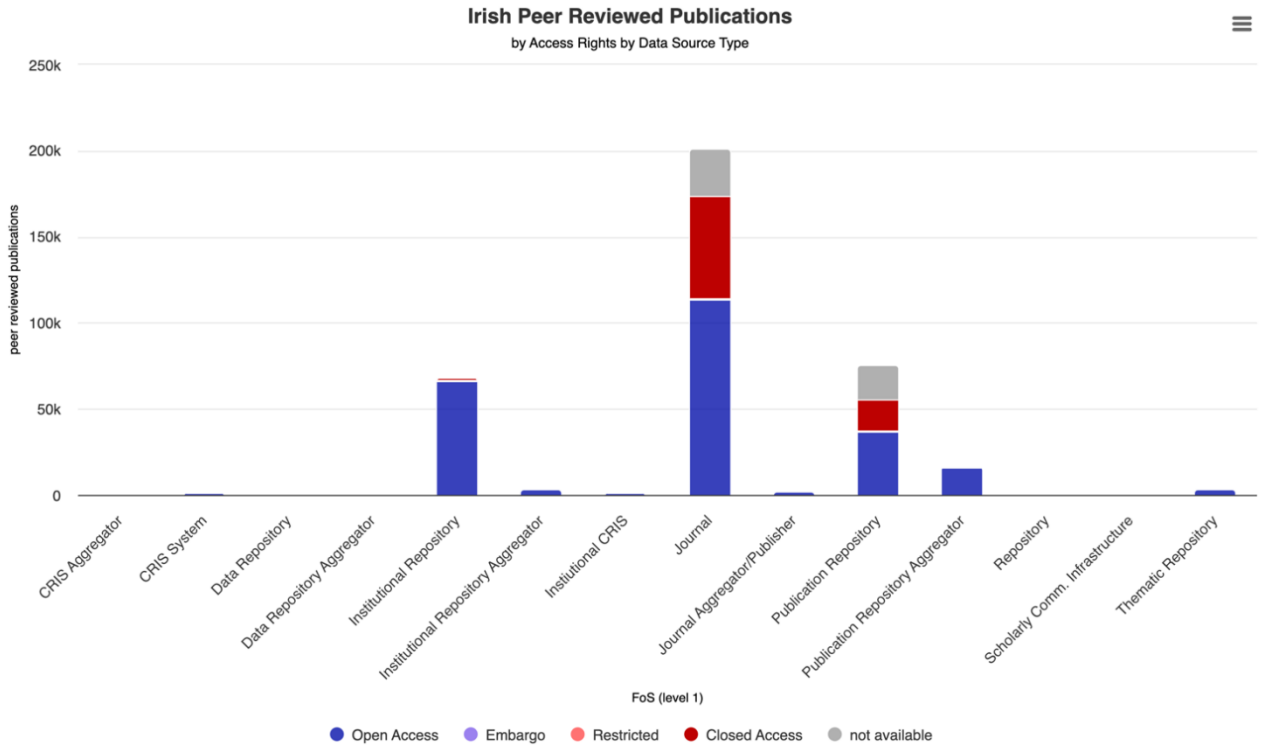

Figure 9: Irish Peer Reviewed Publications by Access Rights by Data Source Type

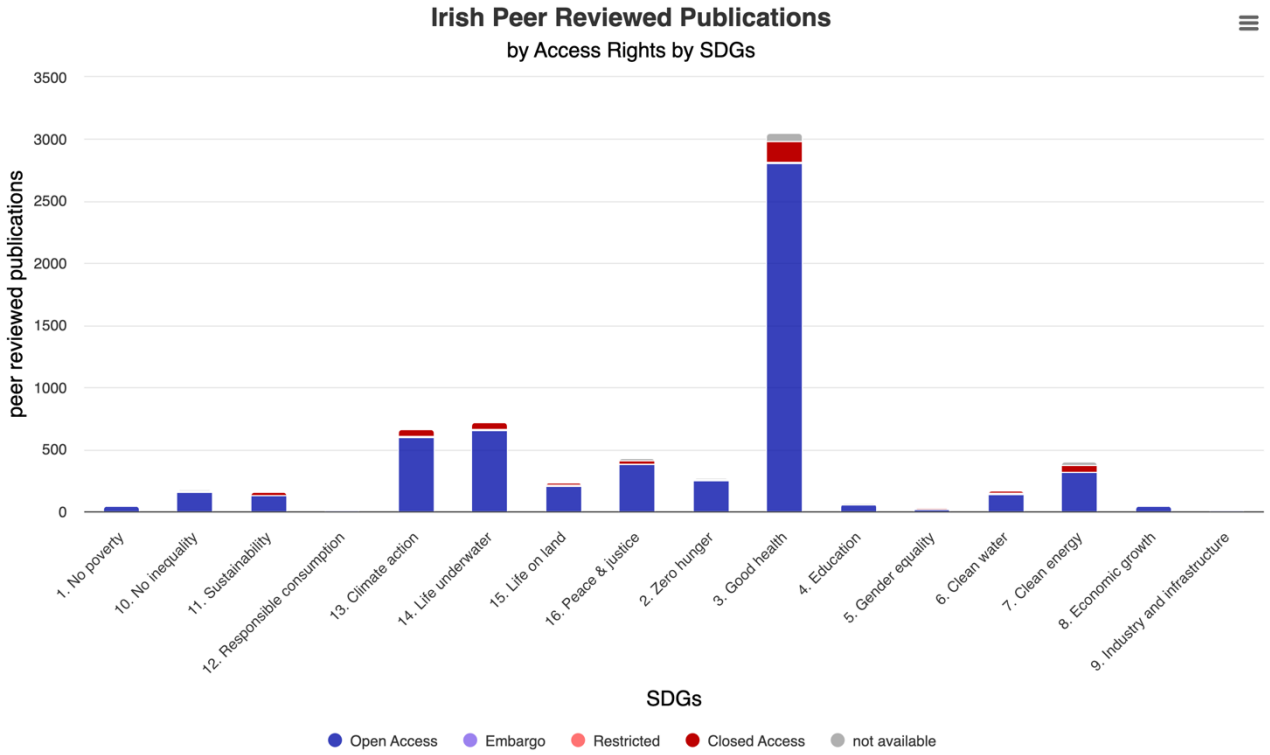

Figure 10: Irish Peer Reviewed Publications by Access Rights by SDGs

2.2.2 Open Access types

The analysis in this section will be aligned with the Budapest Open Access definition, encompassing the following actions:

Distinguishing Green Open Access publications based on the presence or absence of a CC license.

***Excluding Bronze publications from the Open Access category** due to the absence of a CC license. However, comprehensive data and metrics for Bronze publications will still be presented as they will be repositioned to the section designated for Unrealized Open Access.*

Soliciting feedback from the Advisory Group through various visualisations, allowing them to decide the preferred representations for the dashboards.

Table 5 provides a detailed breakdown of Open Access, distinguishing Repository-mediated OA (Green) from Publisher-mediated OA (Gold, Hybrid, Bronze). We note that approximately one-quarter of Irish Open Access publications fall under Repository-

mediated Open Access, while nearly three-quarters are categorized as Publisher-mediated Open Access.

Table 5: Number and share of Repository mediated vs publisher mediated OA Irish Peer Reviewed Publications

Open access type	# Irish OA peer-reviewed publications	Share of # Irish OA peer-reviewed publications
Repository mediated OA (Green)	*28,277*	*22.31%*
Publisher mediated OA (Gold)	*34,592*	*27.30%*
Publisher mediated OA (Hybrid)	*19,966*	*15.76%*
Publisher mediated OA (Bronze)	*10,880*	*8.59%*

In the year-wise distribution chart and table 6 below, we discern a rising trend in both Repository-mediated and Publisher-mediated Open Access. Notably, Publisher-Mediated Open Access has increased by over 50% from 2014 to 2021. This trend aligns with the direction of the Irish research landscape towards greater openness and underscores the effectiveness of mechanisms established by the global Open Science movement and related initiatives.

Table 6: Repository mediated vs publisher mediated OA Irish Peer Reviewed Publications over time

Category	Repository mediated OA (Green)	Publisher mediated OA (Gold)	Publisher mediated OA (Hybrib)	Publisher mediated OA (Bronze)
2007	*591*	*414*	*210*	*266*
2008	*704*	*607*	*253*	*284*
2009	*837*	*761*	*425*	*370*
2010	*911*	*961*	*546*	*342*
2011	*1015*	*1159*	*498*	*507*
2012	*1060*	*1385*	*558*	*533*
2013	*1187*	*1621*	*711*	*526*

2014	*1413*	*1740*	*822*	*581*
2015	*1476*	*2298*	*973*	*530*
2016	*2106*	*2381*	*1235*	*616*
2017	*2456*	*2798*	*1377*	*542*
2018	*2810*	*3104*	*1621*	*902*
2019	*3354*	*3573*	*1752*	*1030*
2020	*2818*	*4119*	*2632*	*922*
2021	*2319*	*4539*	*4021*	*643*
2022	*767*	*1072*	*1030*	*41*

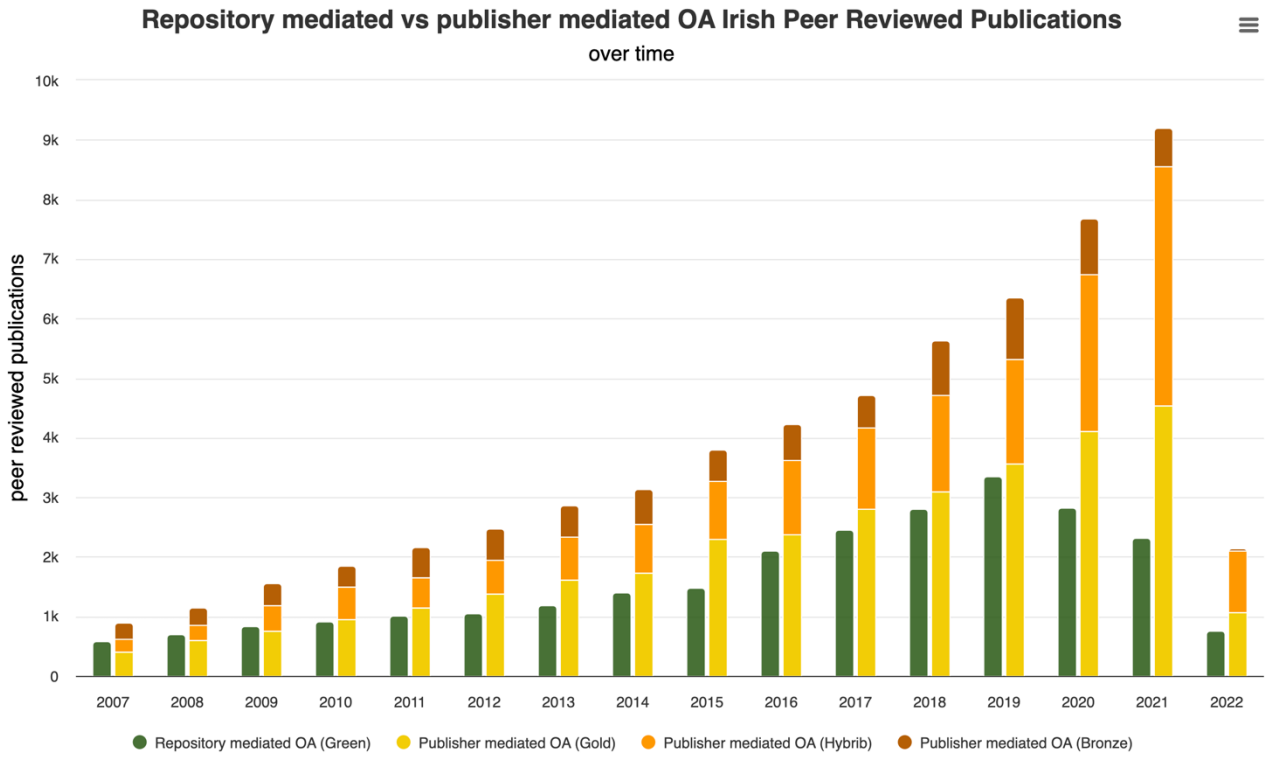

Figure 11: Repository mediated vs publisher mediated OA Irish Peer Reviewed Publications over time

To delve deeper into this trend, the following charts provide a detailed breakdown by Field of Science (FoS). This finer-grained perspective can shed light on particular disciplines that might be contributing to the overall rise in Open Access peer-reviewed publications, offering valuable insights for shaping future OA strategies.

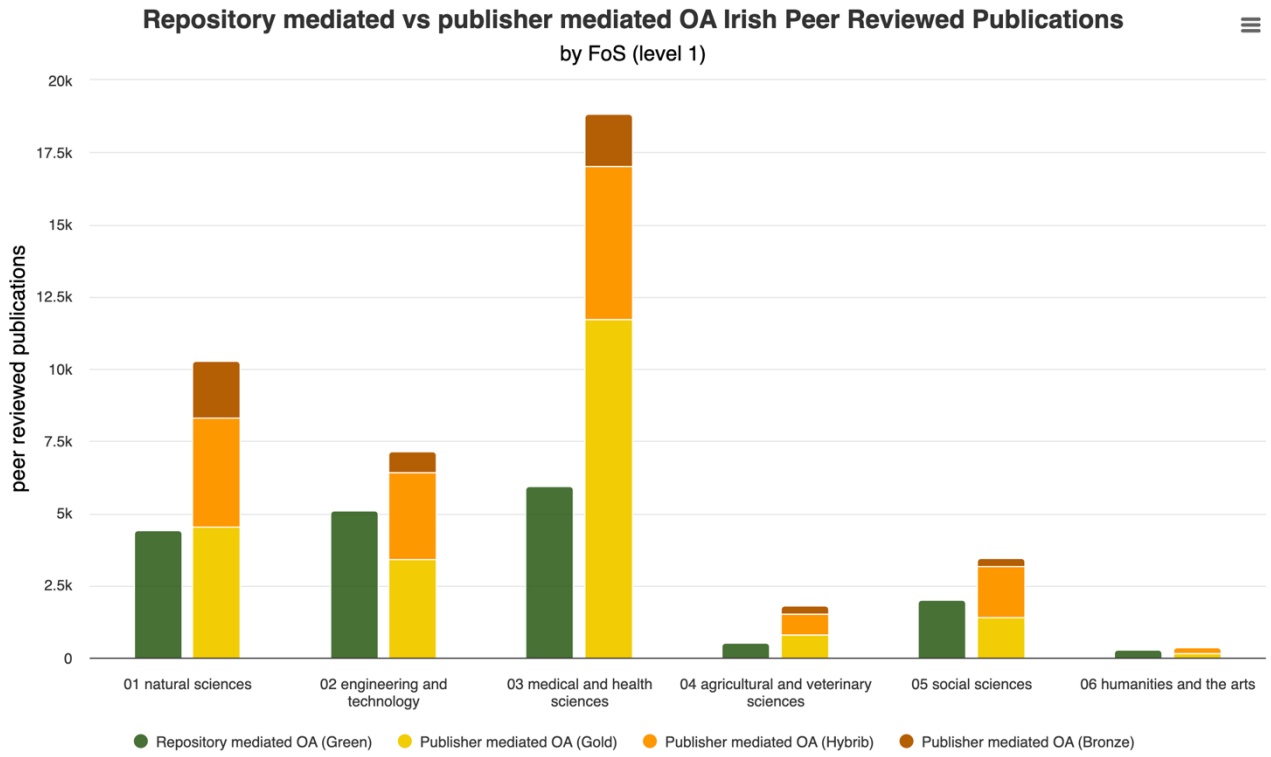

Figure 12: Repository mediated vs publisher mediated OA Irish Peer Reviewed Publications by FoS (level 1)

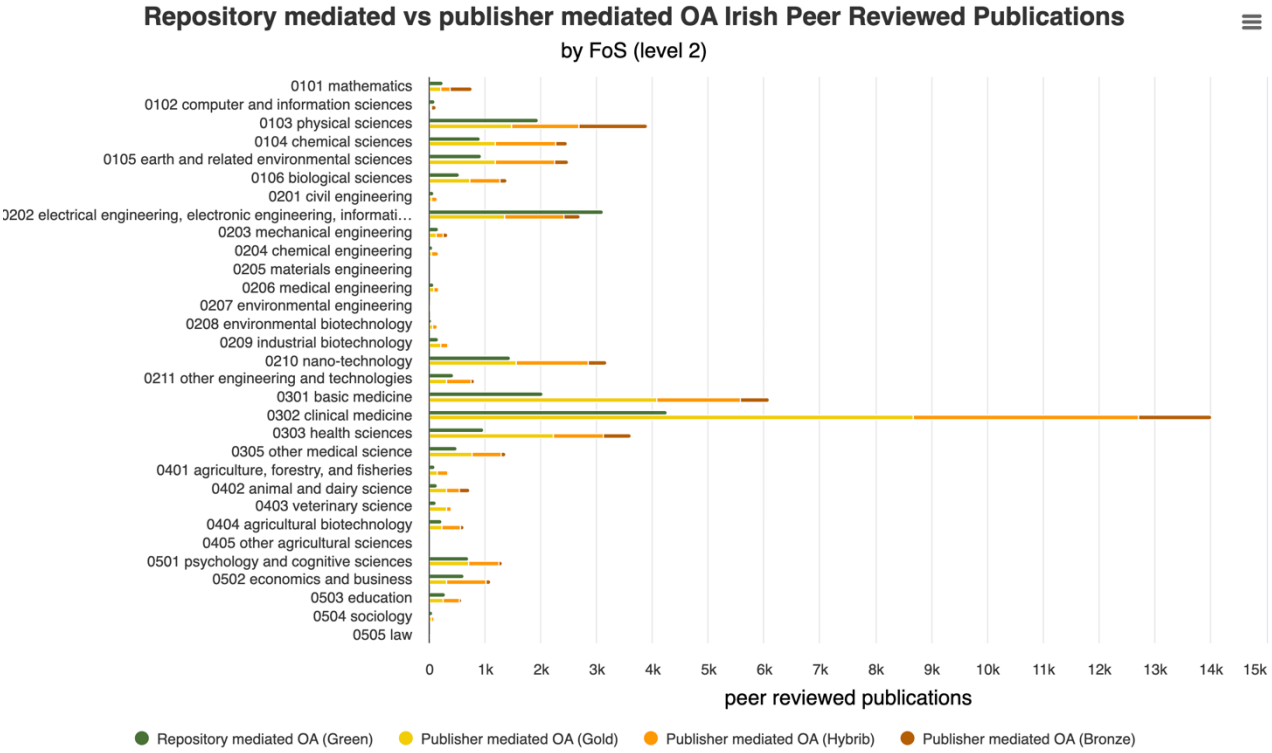

Figure 13: Repository mediated vs publisher mediated OA Irish Peer Reviewed Publications by FoS (level 2)

2.2.3 Embargoed vs. Immediate Open Access

As shown in Table 7, which outlines the count and proportion of Immediate Open Access and Embargoed Irish Peer-Reviewed Publications, half of the Irish Open Access peer-reviewed publications are readily available in an Open Access type (Immediate OA). As immediate OA we define and identify the Gold and Hybrid OA publications.

To categorize Green OA publications as Immediate OA, the deposition date of the publications must be obtained from the relevant data sources. This information is typically maintained by the repositories. While the OpenAIRE Guidelines do incorporate the Publication Date stating that "Typically, Date will be associated with the creation or availability of the resource", its identification and population with the correct date is contingent upon data sources exposing it through their OAI-PMH protocol.

Table 7: Number and share of All vs Immediate Open Access Irish Peer-reviewed Publications

All vs Unrealised OA	# Irish OA peer-reviewed publications	Share of # Irish OA peer-reviewed publications
All OA	*127,221*	*100.00%*
Immediate OA	*54,558*	*42.9%*
Embargoed	*26*	*0.02%*

2.2.4 Unrealised Open Access

As shown in Table 8, which outlines the count and proportion of Unrealised Open Access Irish Peer-Reviewed Publications, it is apparent that there are 81,414 publications falling under the "unrealised Open Access" category. The latter comprises of publications initially designated as non-Open Access and Bronze publications.

As aforementioned, in the final edition of the report, Bronze publications will not be deemed as Open Access and will be repositioned to this section.

Table 8: Number and share of Open Access vs Unrealised Open Access Irish Peer-reviewed Publications

OA vs Unrealised OA	# Irish peer-reviewed publications	Share of # Irish peer-reviewed publications
OA	*127,221*	*52.1%*

Unrealised OA	*116,935*	*47.9%*

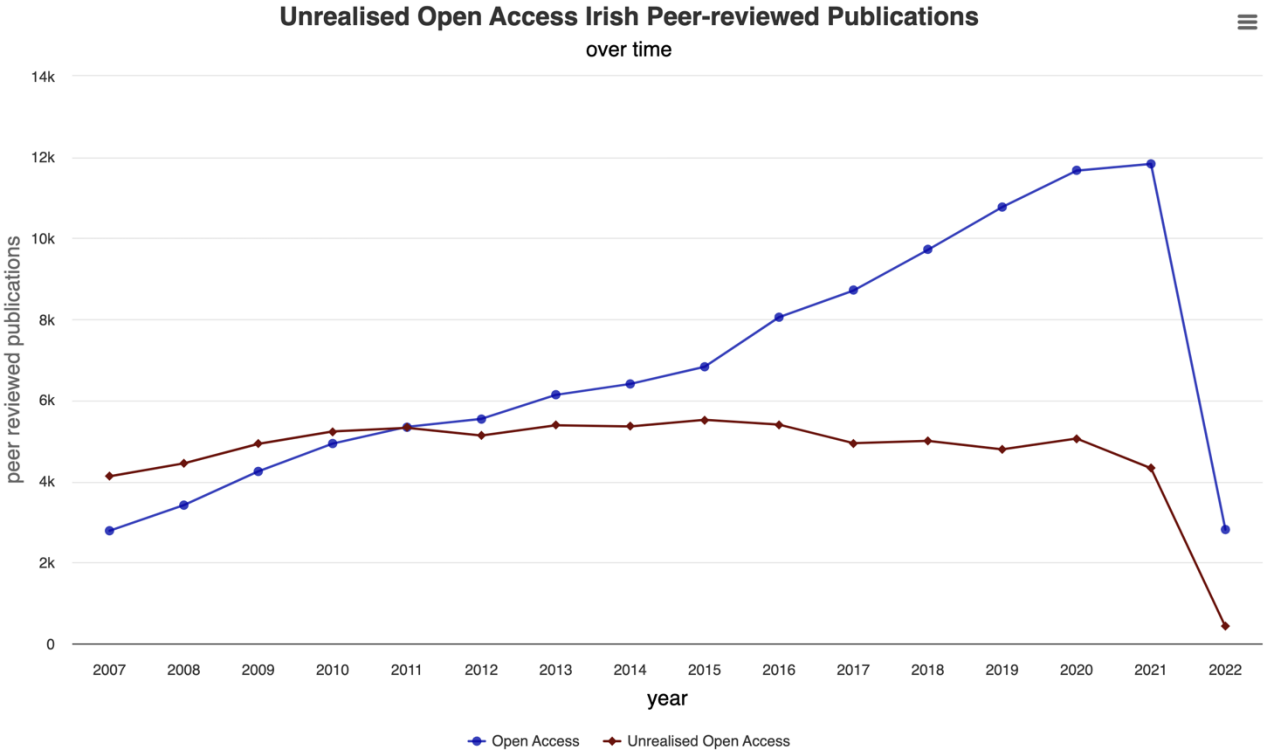

Figure 14: Unrealised Open Access Irish Peer-reviewed Publications over time

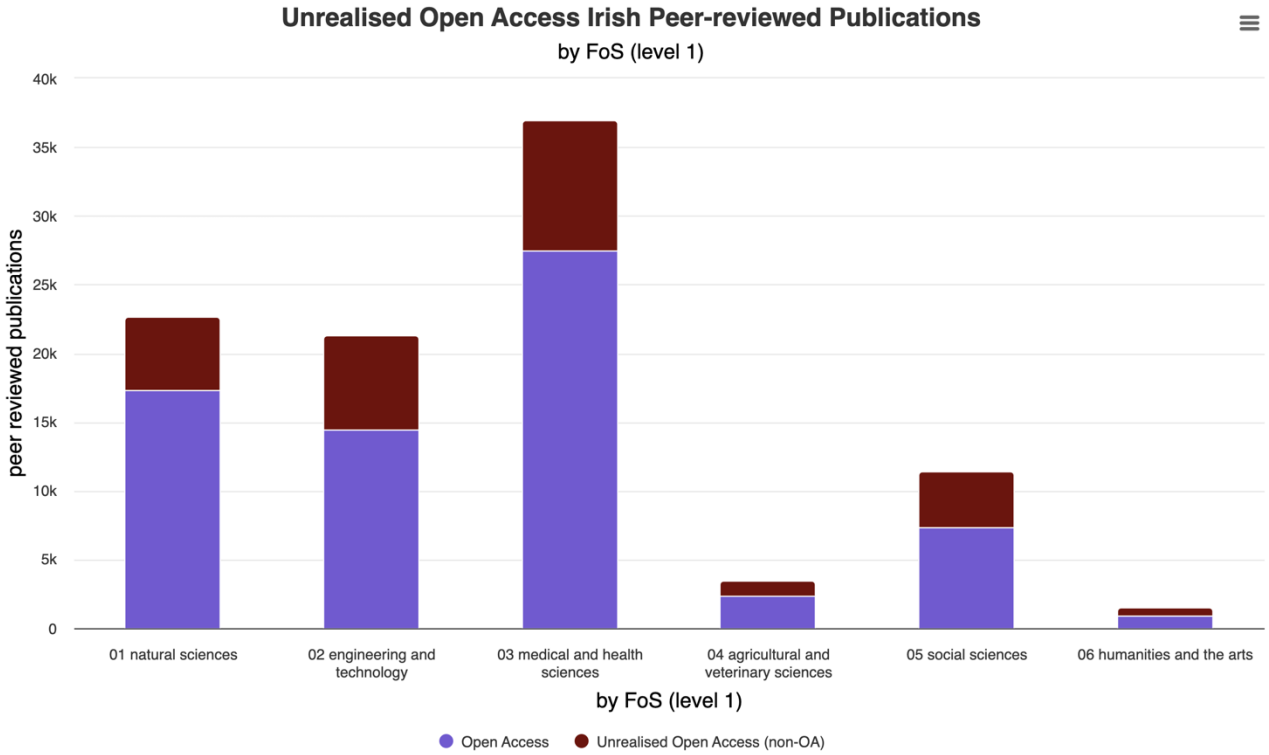

Figure 15: Unrealised Open Access Irish Peer-reviewed Publications by FoS (level 1)

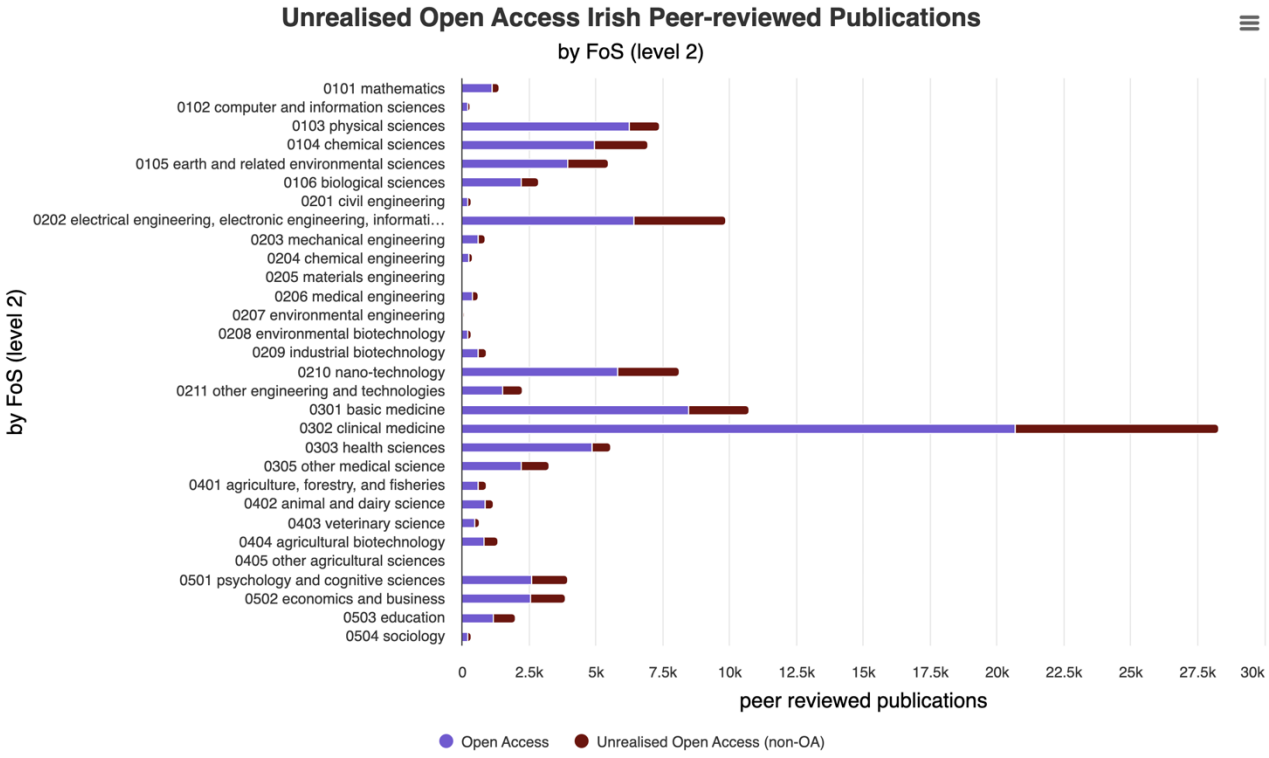

Figure 16: Unrealised Open Access Irish Peer-reviewed Publications by FoS (level 2)

2.2.5 FAIRness

As previously mentioned, we will revise the charts and tables accordingly in the final version of the report to align with the Budapest Open Access Initative definition.

Table 9: Number and share of Open Access Irish Peer-reviewed Publications with a CC license

	# Irish OA peer-reviewed publications	Share of # Irish OA peer-reviewed publications
with CC license	191,929	78.61%
without CC license	52,227	21.39%

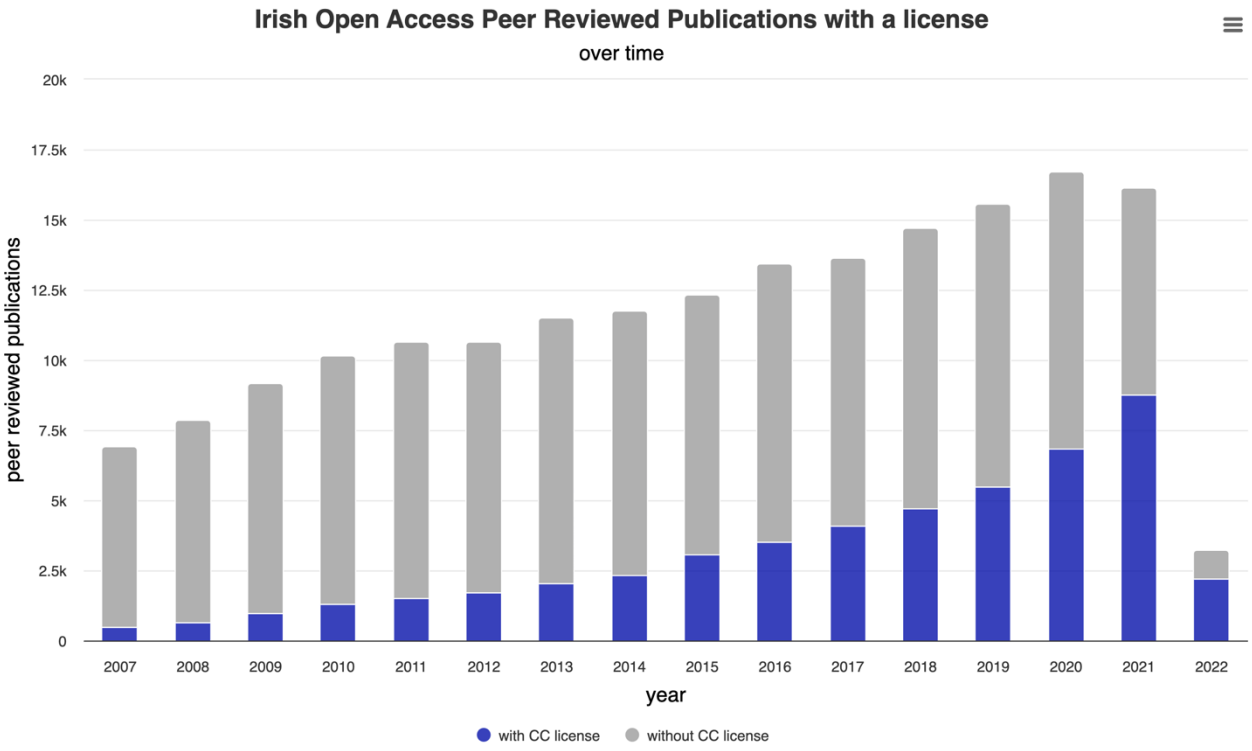

Figure 17: Irish Open Access Peer Reviewed Publications with a license over time

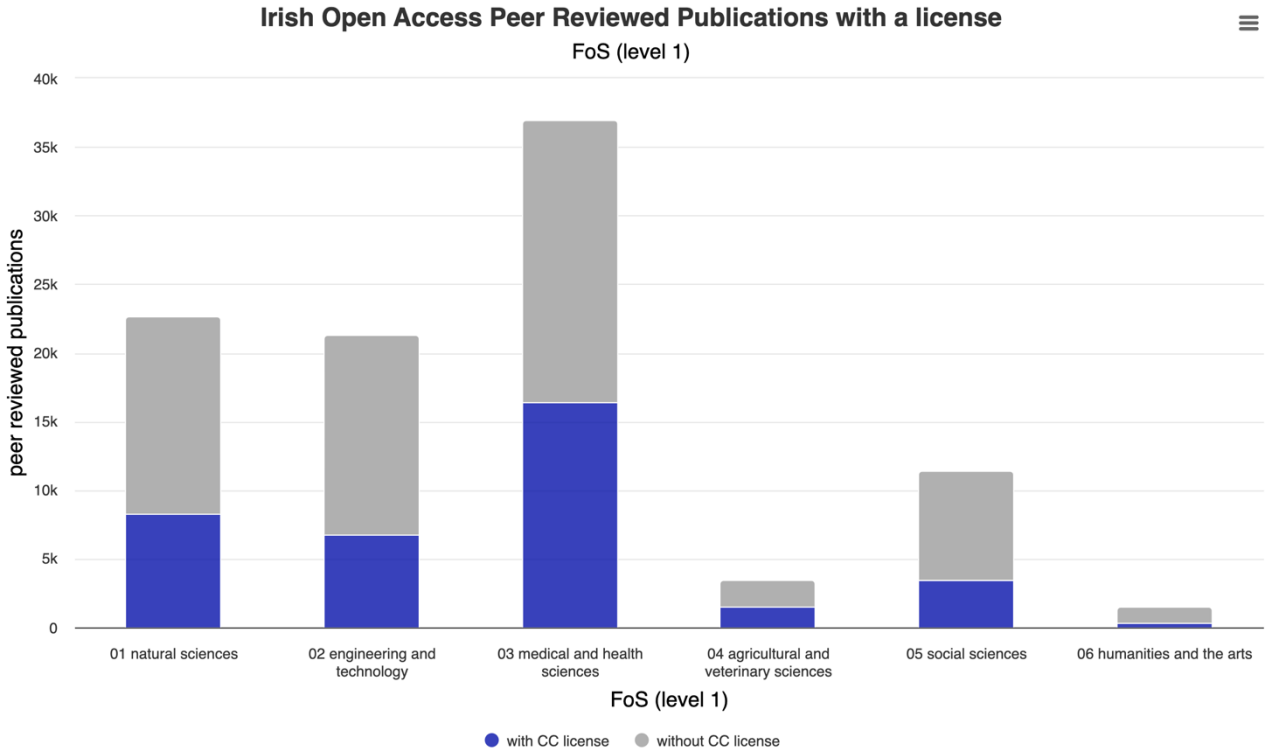

Figure 18: Irish Open Access Peer Reviewed Publications with a license by FoS (level 1)

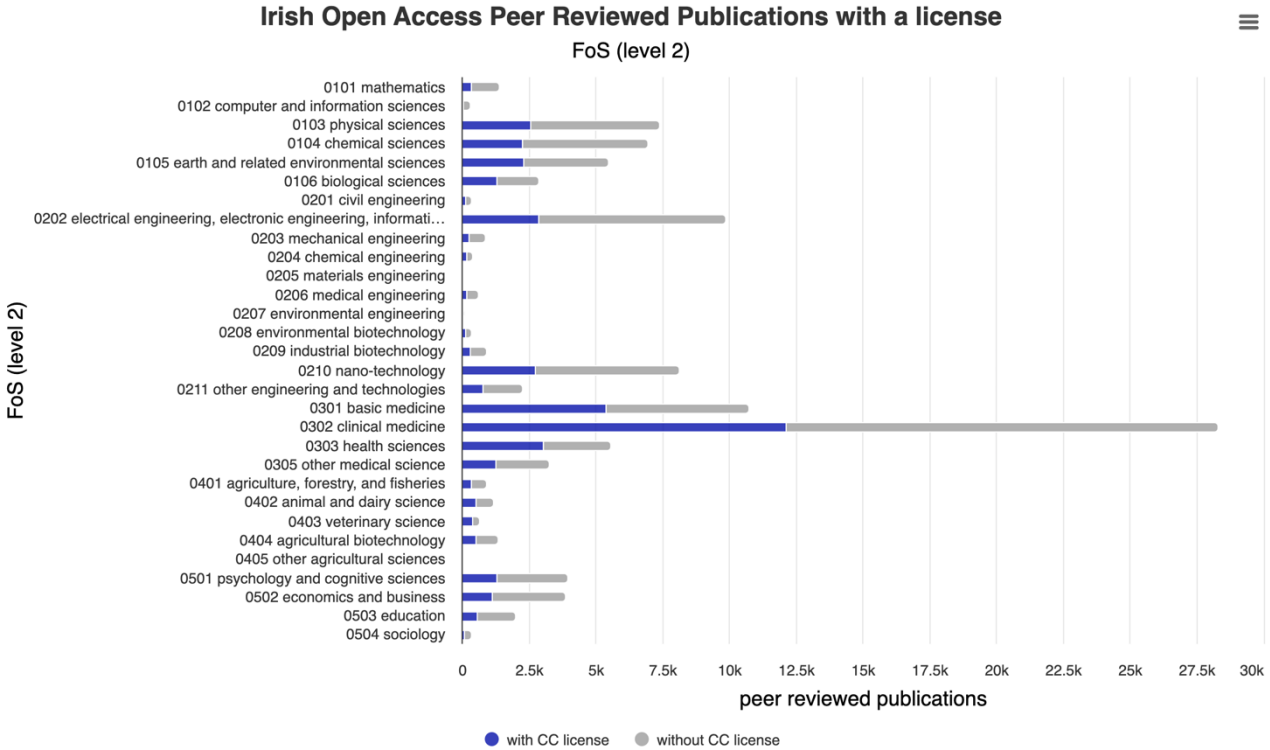

Figure 19: Irish Open Access Peer Reviewed Publications with a license by FoS (level 2)

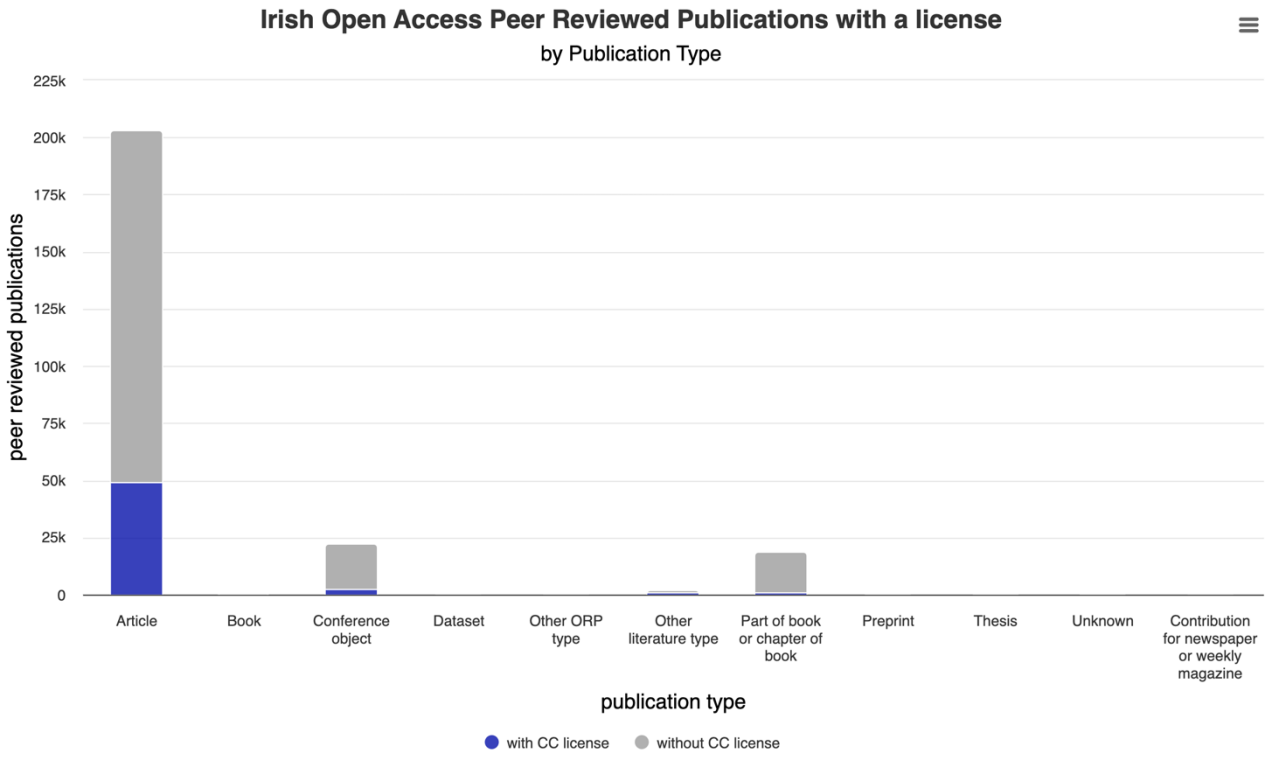

Figure 20: Irish Open Access Peer Reviewed Publications with a license by Publication Type

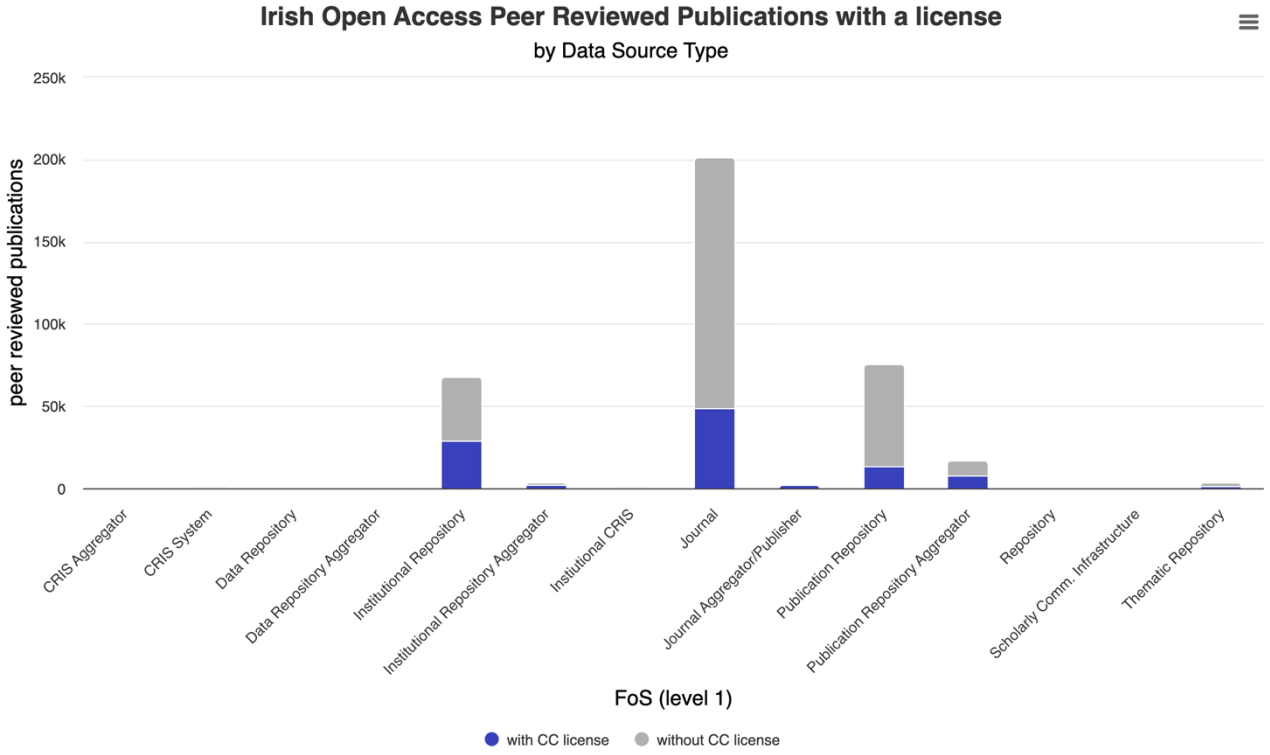

Figure 21: Irish Open Access Peer Reviewed Publications with a license by Data Source Type

License information is integral to the Reusability aspect of FAIRness in metadata, as it encompasses details regarding the permissions and conditions under which the publications can be reused.

We are currently enhancing the Open License coverage of the Irish Publications and improving the license normalization algorithm, aiming for a more comprehensive representation in the National Open Access Monitor of Ireland.

The table below provides an overview of the progress made in grouping and normalizing licenses. It is important to note that a merged record of a publication may include multiple licenses.

Table 10: Licence normalisation

Licence	# Irish OA peer-reviewed publications
CC-BY	*30,561*
CC-BY-NC-ND	*13,679*
CC-BY-NC	*4,759*
CC-BY-NC-SA	*3,755*
CC-0	*664*
CC-BY-SA	*375*
CC-BY-ND	*97*
CC-BY-ND-SA	*22*
Elsevier TDM	*10,584*
Elsevier OA	*8,135*
Wiley VOR	*4,123*
Springer TDM	*5,618*
Wiley TDM	*4,611*
arXiv non-exclusive distribution	*3,055*
Oxford Academic reuse	*2,802*

Sage TDM	*2,595*
CambridgeCORE	*2,454*
IEEE	*2,377*
APS	*1,545*
ASM TDM	*1,042*
% Irish OA peer-reviewed publications with non-normalised licenses	*11.63%*

2.3 Plan S

Plan S[10] represents a pioneering initiative with the primary objective of ensuring the immediate and universal accessibility of publicly funded scientific research. It stands as a significant milestone in the broader open-access movement within scientific publishing.

At its core, Plan S embodies a set of fundamental principles that dictate researchers funded by cOAlition S members must publish their research findings through open-access journals or platforms. Alternatively, they can choose to make their articles available in recognized open repositories without any embargo period. This commitment to openness encompasses diverse types to compliance, including publishing in open-access journals, depositing articles in recognized repositories, and participation in transformative journals transitioning towards full Open Access.

Moreover, Plan S places emphasis on transparency and accountability regarding publication fees. It seeks to establish limits on charges levied by publishers for OA publishing. Additionally, it underscores the importance of licensing, mandating that articles be published under an open license, preferably the Creative Commons Attribution licence(CC BY), which enables others to reuse the works with appropriate credit.

In this section, we explore the indicators associated with Plan S compliance, offering insights into the progress and adherence to these transformative principles within the Monitor.

Publishing in hybrid journals under transformative agreements meets Plan S compliance and will be represented in the Plan S section in the final version of the report.

[10] https://www.coalition-s.org/about/

Furthermore, we will investigate the possibility of refining the indicators on the date of the Plan S Transformative Journal concept launch in April 2020.

Table 11: Irish Open Access Peer Reviewed publications by Journal Business Model

	# Irish OA peer-reviewed publications
Full OA with APCs	*31,224*
Diamond	*146*
Transformative Journal	*5,467*

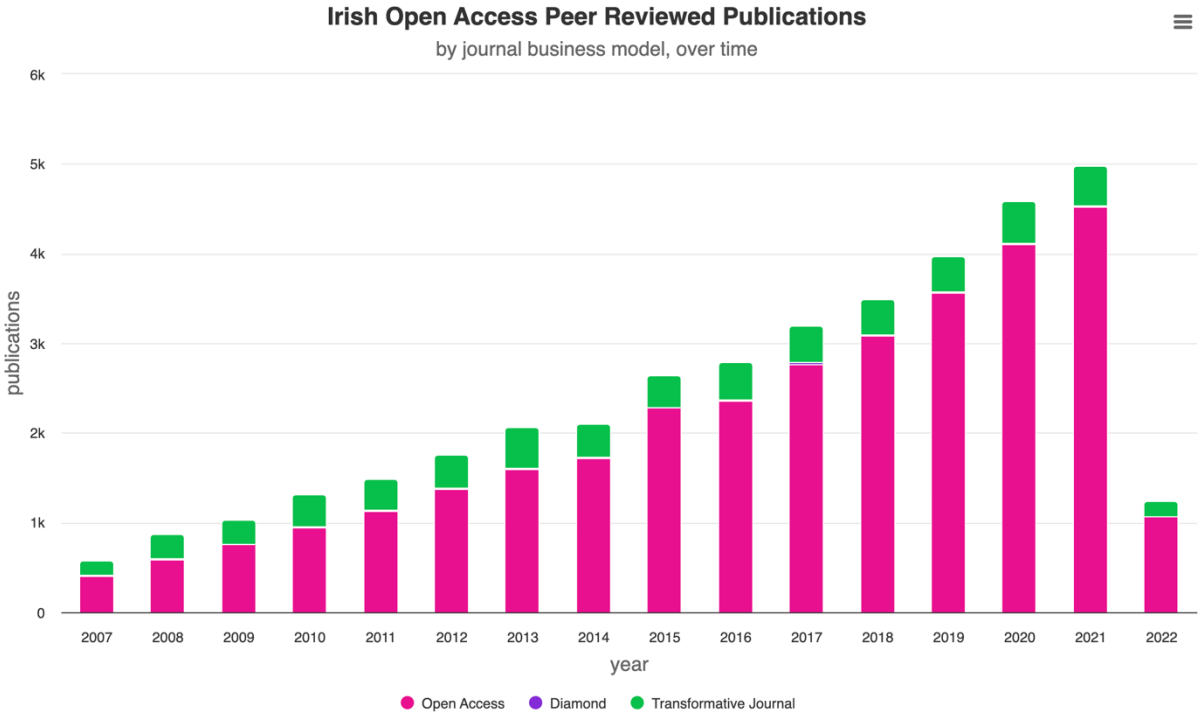

Figure 22: Irish Open Access Peer Reviewed Publications by journal business model, over time

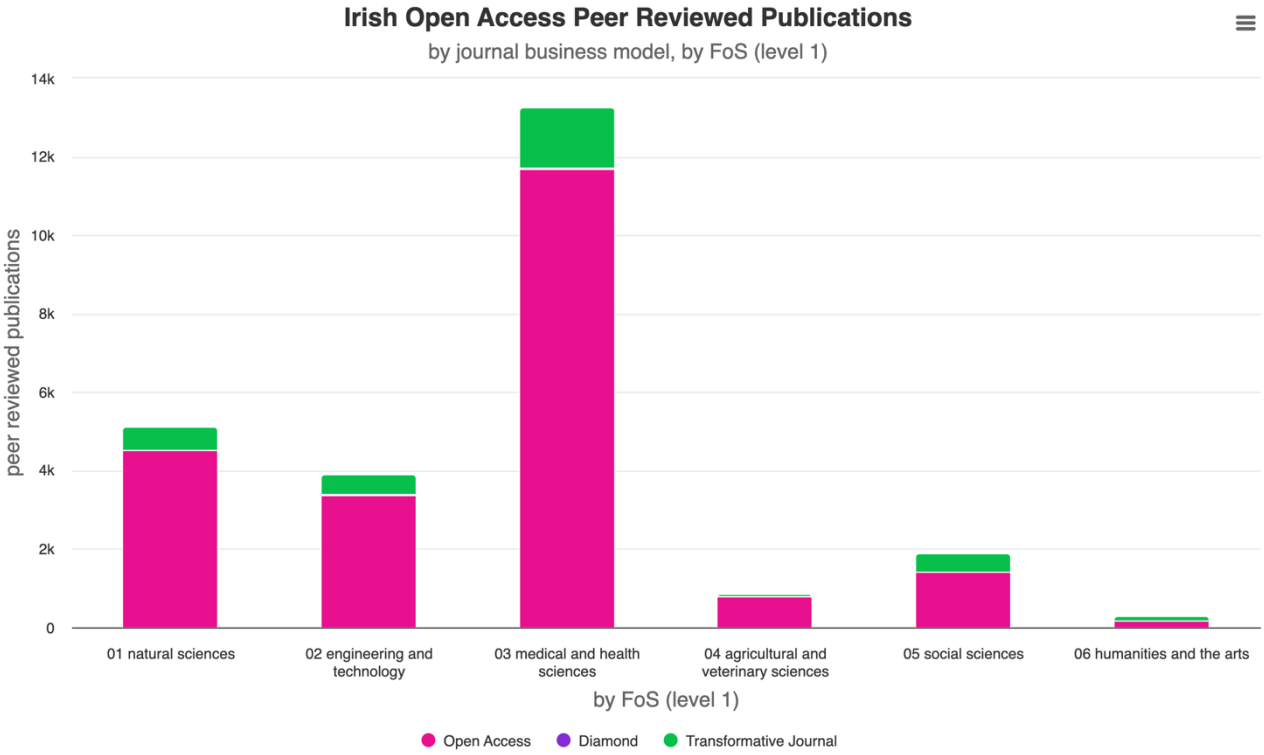

Figure 23: Irish Open Access Peer Reviewed Publications by journal business model, by FoS (level 1)

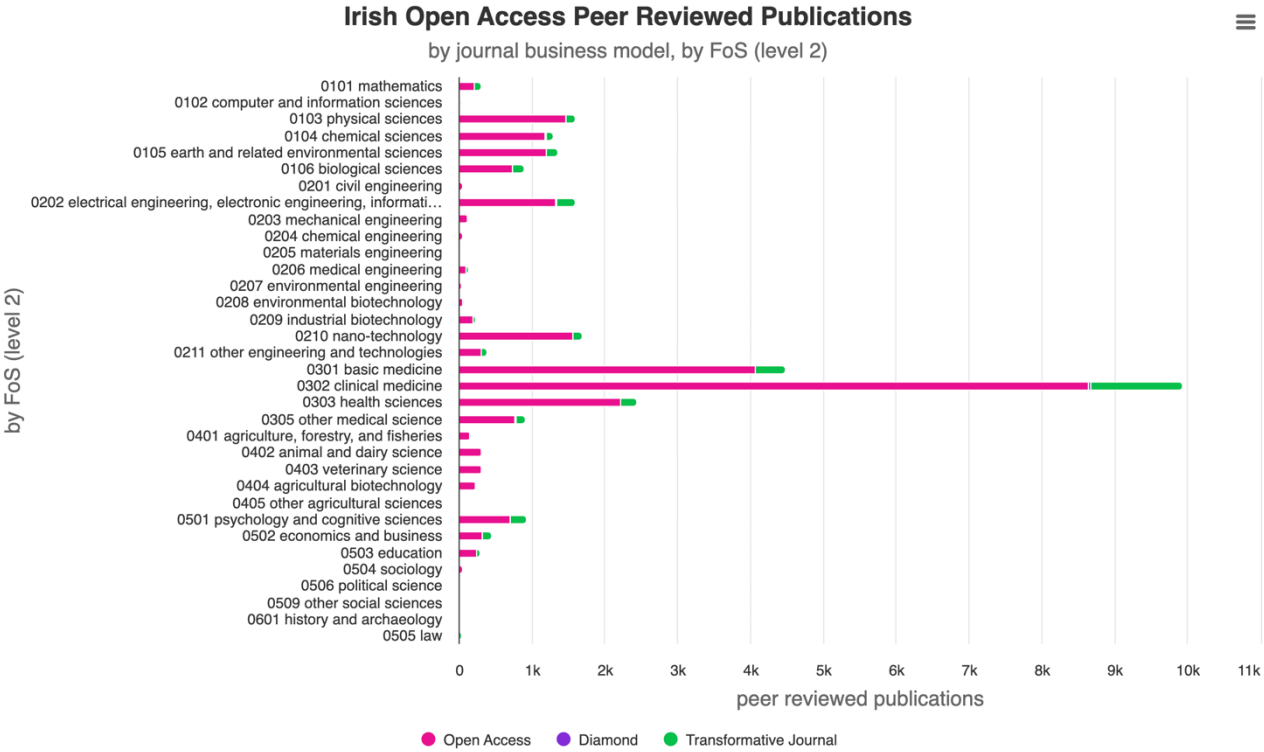

Figure 24: Irish Open Access Peer Reviewed Publications by journal business model, by FoS (level 2)

The Irish OA Peer Reviewed Publications by journal business model table and charts reveal that the Full OA with APCs business model holds a substantial proportion comparing to the Diamond and Transformative model.

Table 12: Total and average APCs of # Irish OA peer-reviewed publications

	# Irish OA peer-reviewed publications
APCs for all affiliated publications	*€1,775,412*
Average APCs per publication (For affiliated publications with reported APCs)	*€2,233.22*

The table presented above provides information on the overall and average APCs associated with Irish Open Access peer-reviewed publications. At present, the APCs are computed in accordance with the described methodology, combining the APCs reported by the institution with those for publications featuring co-authors affiliated with other institutions (APCs not funded by the institution). Currently we showcase indicators based on APCs calculated at the individual publication level rather than at the institutional level.

We are currently in the process of refining this calculation to specifically identify APCs reported solely by the institution. This enhancement will be integrated before the launch of the Monitor pilot.

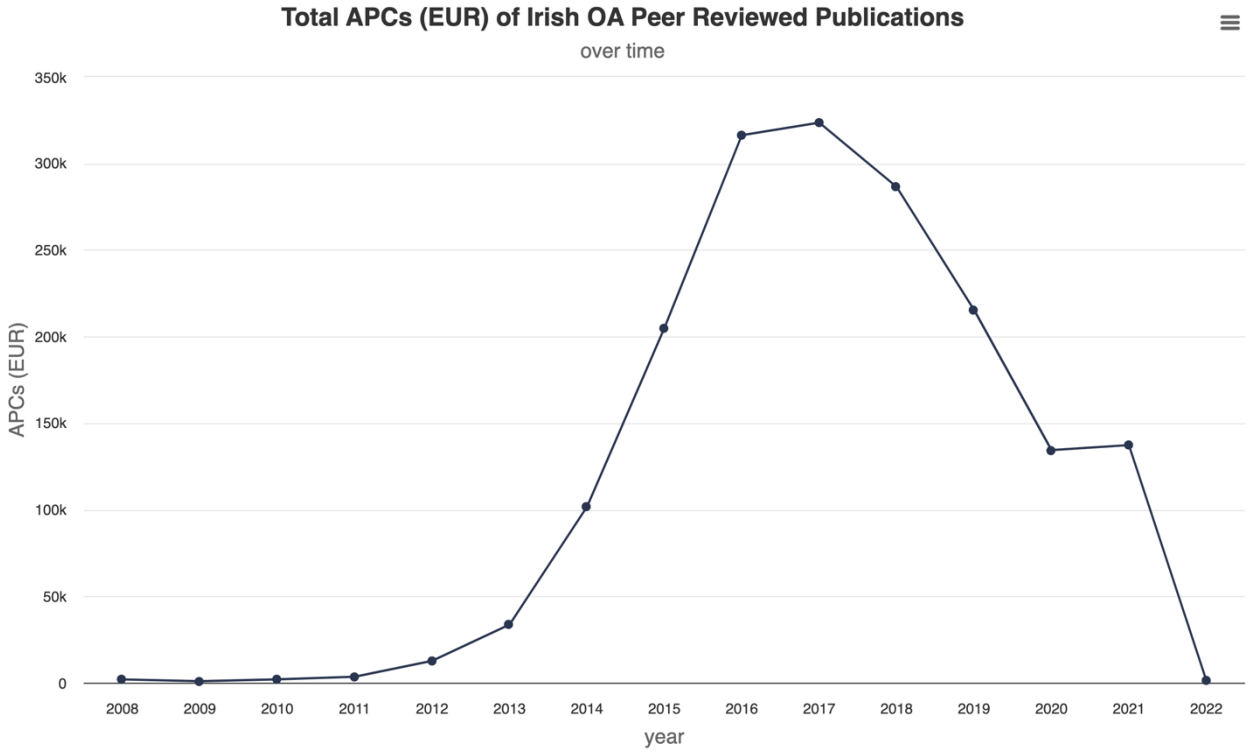

Figure 25: Total APCs (EUR) of Irish OA Peer Reviewed Publications over time

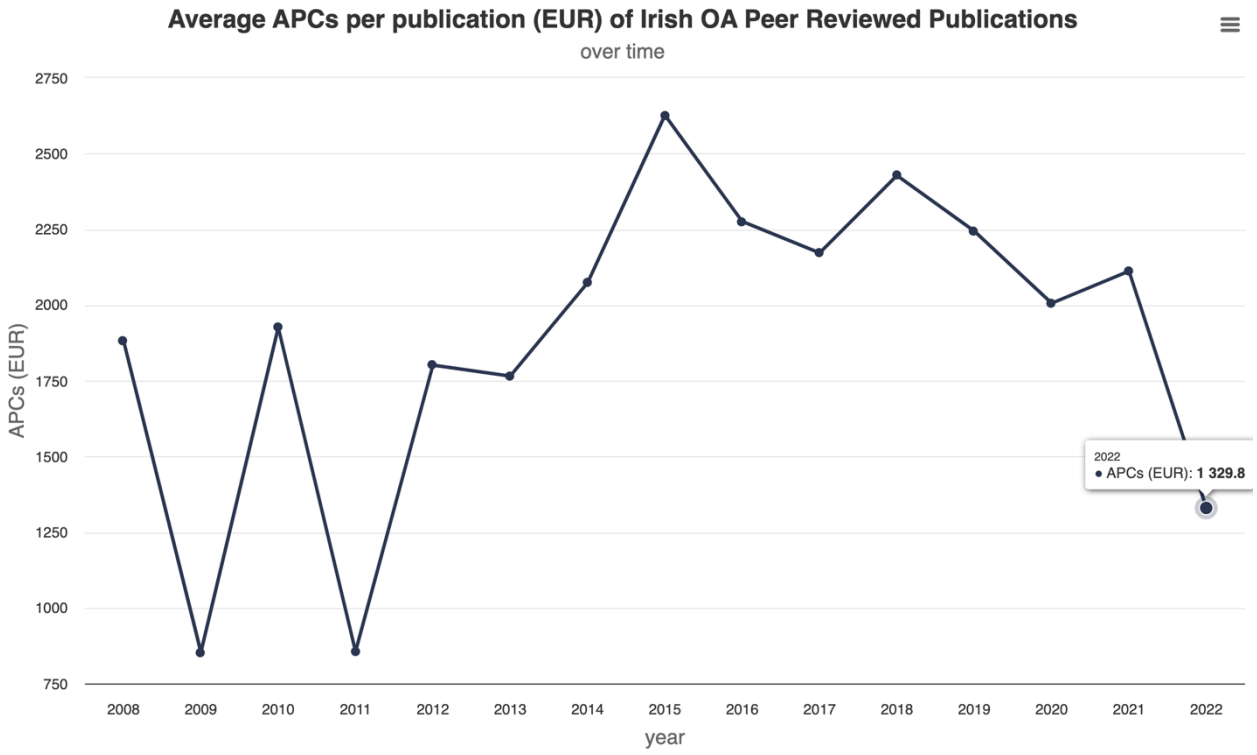

Figure 26: Average APCs per publication (EUR) of Irish OA Peer Reviewed Publications over time

Examining the breakdown of Total APCs by year, a noteworthy pattern emerges, indicating a substantial upsurge commencing in 2012, reaching its zenith during 2016-2017. The cumulative APCs and the average APCs per publication present another potential metric for evaluating the effectiveness of Open Access policies within RPOs or RFOs. A prevailing assumption suggests that the APC level serves as an indicator of the prestige or impact level of an Open Access journal. However, previous analyses have demonstrated that this assumption does not hold true, as high APCs do not consistently align with the high impact of specific articles. Empirical studies have revealed that exorbitant fees charged by certain journals for publishing articles do not necessarily correlate with increased citation rates.

We are in the process of aggregating and conducting additional assessments of APC data sourced from the DOAJ public data dump and OpenAPC. It is worth noting that the OpenAPC platform depends entirely on institutions supplying their APC data. Decentralised APCs paid by faculties or individual authors are not covered in the database.

Upon reviewing the two charts below, which feature the Top 20 Journals ranked by Total APCs and Average APCs per publication, we can discern the distribution of APC expenditures among the journals most frequently utilized by Irish researchers for Open Access publishing. Moreover, in the Average APCs per publication chart, we will include the number of publications to derive more meaningful results.

In the final iteration of the report, we will incorporate a significant and meaningful indicator: "APCs breakdown by publisher".

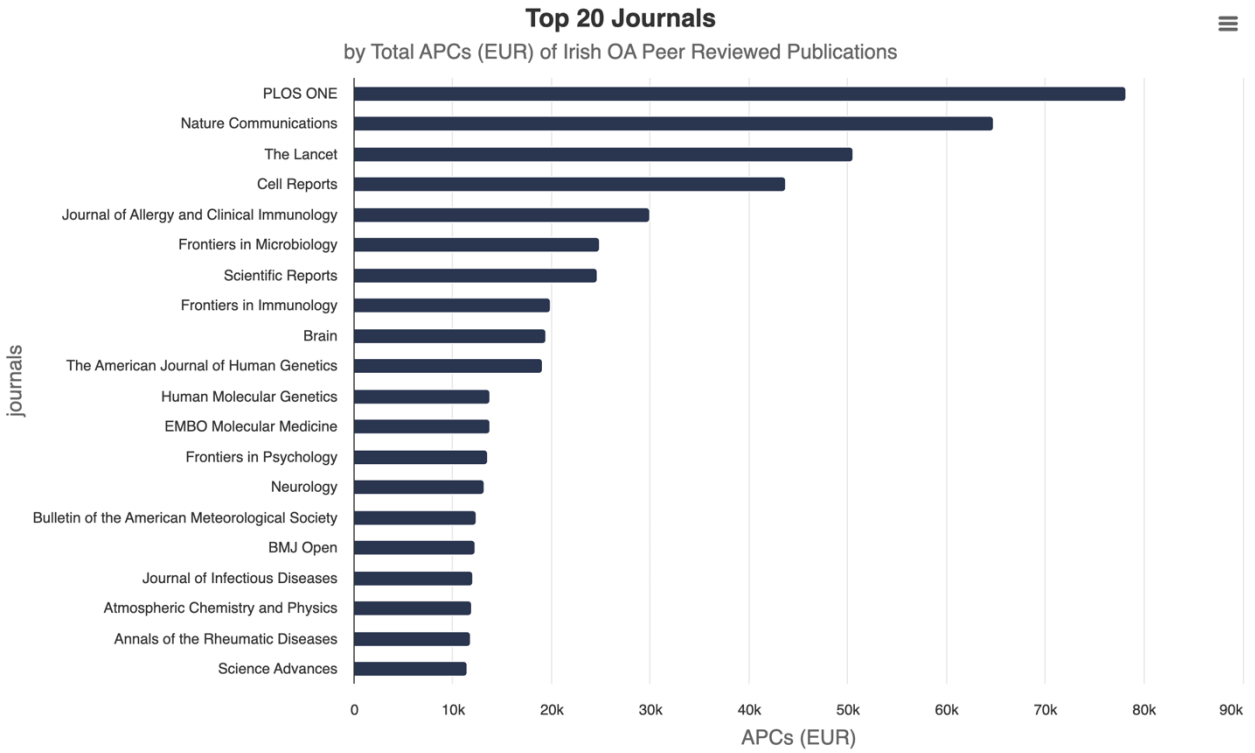

Figure 27: Top 20 Journals by Total APCs (EUR) of Irish OA Peer Reviewed Publications

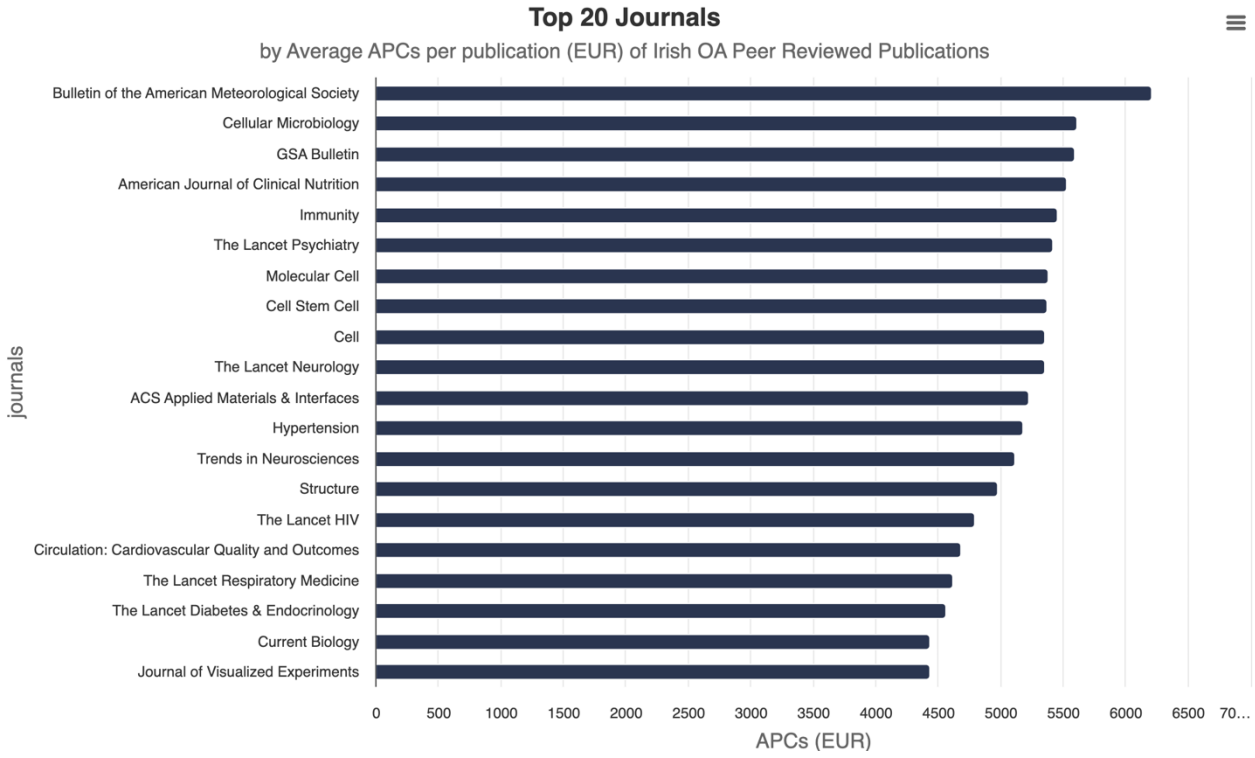

Figure 28: Top 20 Journals by Average APCs per publication (EUR) of Irish OA Peer Reviewed Publications

To delve deeper into this trend, the following charts provide a detailed breakdown by Field of Science (FoS). This fine-grained perspective highlights Medical & Health Sciences domain driving the overall load of APCs. Further investigation will showcase more qualitative results especially with the increase of the FoS coverage until the delivery of the Monitor.

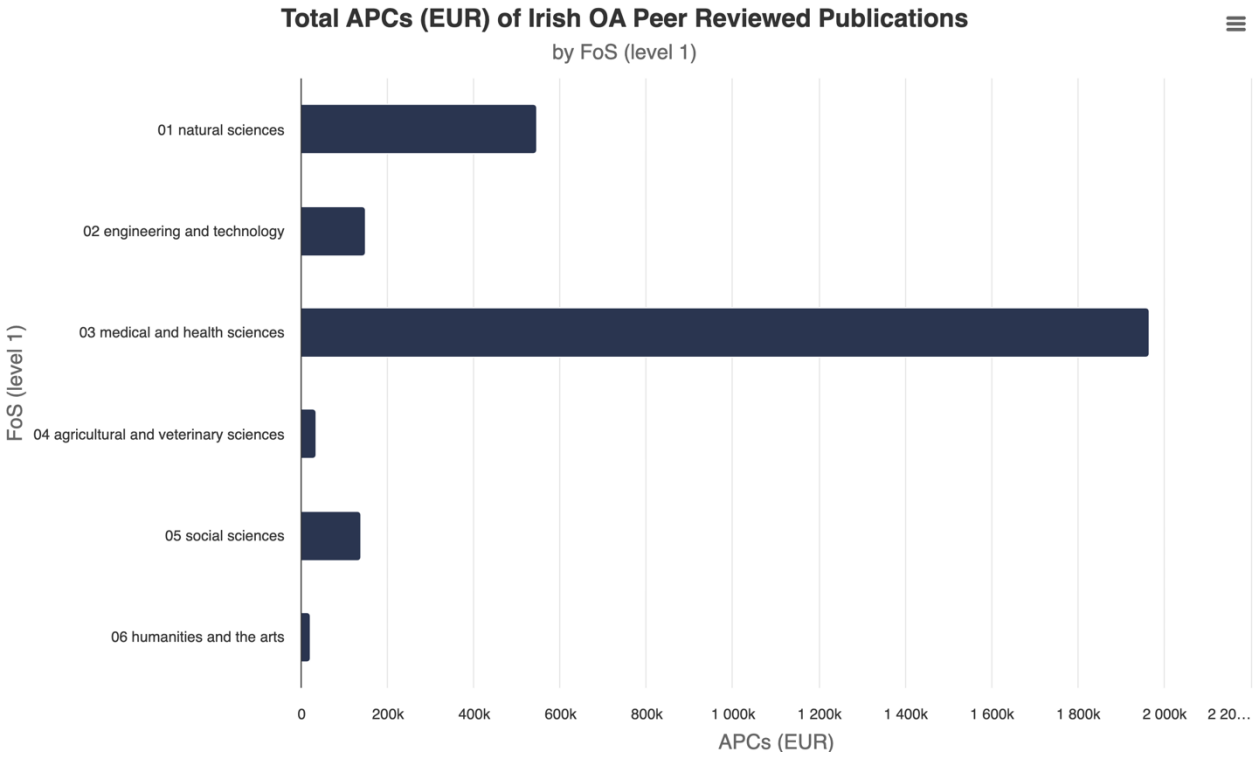

Figure 29: Total APCs (EUR) of Irish OA Peer Reviewed Publications by FoS (level 1)

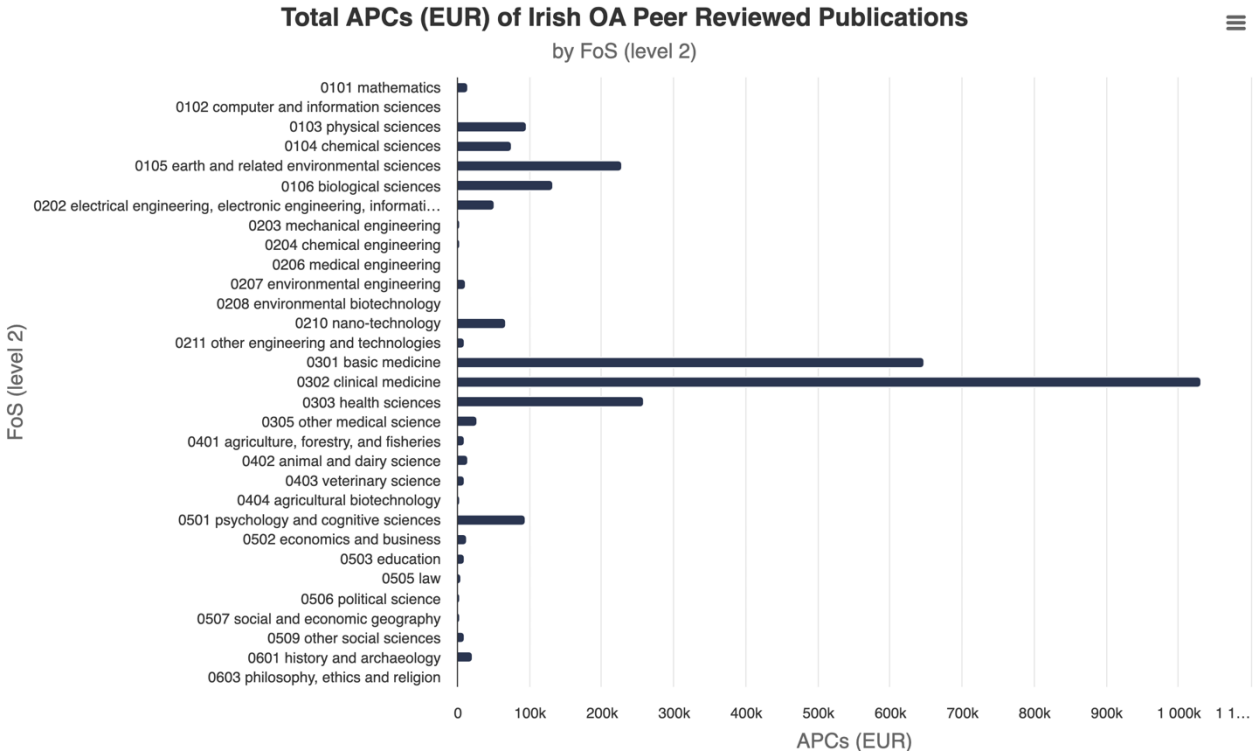

Figure 30: Average APCs per publication (EUR) of Irish OA Peer Reviewed Publications by FoS (level 1)

The following tables provide insights into the count and proportion of Irish peer-reviewed publications that have received grant support from Science Foundation Ireland, a funder member of the cOAlition S consortium. Of the total Irish Open Access peer-reviewed publications, 11,797, comprising approximately 4.83%, have been funded by Science Foundation Ireland. Consequently, these publications adhere to Plan S requirements, ensuring their immediate availability as Open Access resources. An analysis of the distribution of grant-supported Irish peer-reviewed publications across various Open Access types underscores Gold OA as the preferred publishing type, with Green, Hybrid, and Bronze options following. Furthermore, it's worth noting that approximately 40% of these publications are released under a Creative Commons (CC) license.

We will refine the tables to distinguish between pre- and post-Plan S periods, utilizing either the date of the Plan S Transformative Journal concept launch in April 2020 or the year 2021, when the Plan S principles began to predominantly take effect.

Table 13: Number and share of Grant supported Irish Peer Reviewed publications by a Plan S Funder

Plan S Funders	# Grant supported Irish peer-reviewed publications	Share of # of Grant supported Irish peer-reviewed publications

Science Foundation Ireland (SFI)	*11,797*	*4.83%*

Table 14: Number of Grant supported Irish Peer Reviewed publications by Open Access Type

Plan S Funders	Green OA	Gold OA	Hybrid OA	Bronze	With CC license
Science Foundation Ireland	*2,090*	*2,892*	*1,514*	*5,60*	*4,417*

As aforementioned, in the final edition of the report, Bronze Open Access publications will not be deemed as Open Access.

Table 15: Open Access vs Closed Access Grant supported Irish Peer Reviewed publications

Plan S Funders	Open Access	Closed Access
Science Foundation Ireland	*10,071*	*1,223*

Table 16: Share of Grant supported Irish Peer Reviewed publications by Open Access Type

Share of Grant supported Irish Peer Reviewed publications by Open Access Type					
Plan S Funders	**Green OA**	**Gold OA**	**Hybrid OA**	**Bronze**	**With CC license**
Science Foundation Ireland	*17.72%*	*24.51%*	*12.83%*	*0.05%*	*37.44%*

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.[11] This transparent approach ensures the integrity and reliability of our findings, guiding readers through the structured framework that underpins our assessment of the OA landscape.

3.1 OpenAIRE Graph: Foundation of the Monitor

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

The Graph aggregates around 450 million metadata records from more than 120,000 trusted scholarly communication sources worldwide, including Crossref, Unpaywall, ORCID, institutional and thematic repositories (registered in OpenDOAR, re3data.org and FAIRSharing.org), Open Access journals, data archives, and the EOSC Service Catalogue. 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 license.

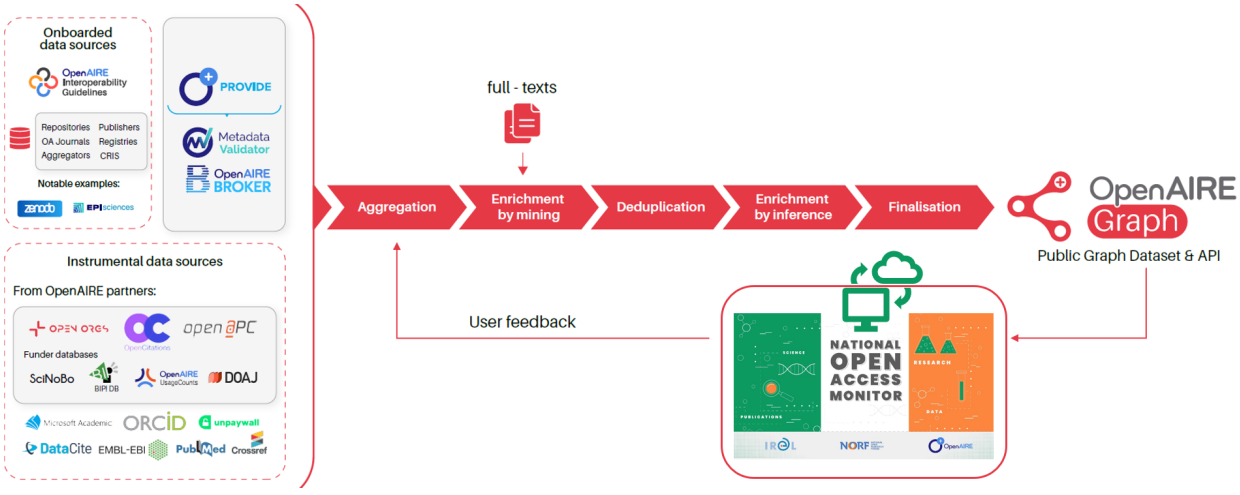

Figure 31: The Monitor and the OpenAIRE Graph Pipeline

[11] Due to its relevance some of the text has been copied from the Inception Report.

3.2 The Publication Set of the Monitor

This section outlines the processes through which the Monitor compiles and utilizes data from Irish Research Performing Organizations (RPOs) and Research Funding Organizations (RFOs), ensuring comprehensive representation of 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 **244.156** peer-reviewed publications.

Identification of Irish Research Performing Organisations' (RPOs) Publications

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

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[12].

The table below provides some summary statistics for the RPOs in the Monitor[13].

Table 17: RPOs & their publications in the Monitor

# Irish RPOs	# publications of Irish RPOs
36	*337,121*

Irish institutional data sources registered in OpenAIRE

[12] https://explore.openaire.eu/

[13] The RPO list that will be provided by IReL, through the "National Open Access Monitor Survey: Organisational Identity" will be used to refine this set.

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.

Table 18: Irish Institutional Data Sources & Alignment with OpenAIRE

Institutional Data Source Type	**Registered in OpenAIRE**		**Harvested by OpenAIRE (directly or via a compatible aggregator)**	
	# sources	# publications	# sources	# publications
Repositories	*17*	*98,151*	*0*	*0*
CRIS Systems	*0*	*0*	*0*	*0*
OA Journals	*0*	*0*	*18*	*2,628*

Among the registered Irish repositories within OpenAIRE, a cumulative count of 98,151 publications is recorded. However, it's important to note that, among these repositories, 13 adhere to the BASIC and to version 2.0 of the OpenAIRE Guidelines. As a consequence, 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. These repositories also fall short of alignment with Open Science mandates and established standards, as they do not endorse well-established metadata schemas and namespace abbreviations for project identifications, including EC funding programs, among others. 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, preferably upgrading to version 3.0 or, preferable, version 4.0.

OA Journals that are harvested through a compatible aggregator are not inherently linked to the institution or publisher within OpenAIRE. The subsequent actions involve institutions either registering their OA Journals in OpenAIRE or facilitating the registration and harvesting process. This is contingent upon the platforms supporting the OAI-PMH harvesting protocol.

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[14].
- Collecting links from users via the "link" functionality.
- Exchanging data with the EC's IT systems for EC/FP funding.
- Text mining of full text publications to identify the grants for 30+ funders that have joined OpenAIRE (see next paragraph). Science Foundation Ireland (SFI) is one of them.

Irish Funders in OpenAIRE

Irish funders are represented in the OpenAIRE graph through two primary avenues. The first is by directly joining OpenAIRE[15], 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 19: Irish Funders that have joined OpenAIRE

Irish RFOs that have joined OpenAIRE	Projects	Publications
Science Foundation Ireland	*6,384*	*11,797*

The second avenue for representation is through the Open Funder Registry (OFR)[16] 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 20:General Representation of Irish Funders in OpenAIRE (excluding SFI)

# Irish RFOs integrated via OFR	# Irish RFOs integrated via OFR with publications	# Irish RFOs' publications
142	*79*	*11,351*

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

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

[16] Crossref, https://www.crossref.org/services/funder-registry/

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 articles, based on the following criteria:

1. **Curated Peer-Review Assessment:** The OpenAIRE team has engaged in a curation process to determine the peer-review status of journals. This hand-curated assessment has been integrated into the Graph and is continuously under development.
2. **Exclusion of Grey Literature:** We also filter out 'grey literature', which includes document types that typically bypass the peer review process, such as reports, theses, 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.
3. **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 three criteria gives the following number of Irish peer-reviewed publications of the Monitor.

Table 21: Irish Peer Reviewed Publications

Irish peer-reviewed publications
244,156

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 as we have been given examples of peer-reviewed data sources that get DOI from DataCite.

Table 22: Irish publications with DOI not assigned by Crossref and NOT included in current set

Irish publications with DOI not assigned by Crossref and NOT included in current set
5,202

- **Reference Count by Field of Study (FOS)**: We are also exploring the possibility of adding a reference count, below which a publication can be considered non-peer-reviewed. Key in this process is examining the mean and standard deviation in the number of references per scientific field to ascertain whether this is a meaningful criterion.

Table 23: Mean and standard deviation in # of references by FoS (level 1)

# References in publications	*Mean*	*Standard deviation*
01 Natural Sciences	*2.45*	*3.88*
02 Engineering & Technology	*1.92*	*2.02*
03 Medical & Health Sciences	*2.69*	*2.72*
04 Agricultural & Veterinary Sciences	*3.13*	*5.71*
05 Social Sciences	*1.93*	*1.83*
06 Humanities & The Arts	*2.20*	*1.69*

The final draft of our report will solidify our selection criteria. Nevertheless, the OpenAIRE infrastructure is designed for adaptability. Given that all processes and indicator workflows are fully integrated, we can easily make modifications in the future to meet evolving requirements or to refine our approach.

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 24: Irish peer-reviewed publications by PID type

PID type	# Irish peer-reviewed publications
Publication PIDs	
Digital Object Identifier (DOI)	244,156
Handle	19,354
PubMed Central ID	29,772
PubMed ID	87,020
arXiv	1
Organisation PIDs	
Participant Identification Code	224,434
ISNI	373,695
OrgRef	364,789
Open Funder Registry	353,241
Wikidata	377,701
GRID	379,940
RingGold	25,781
ROR	380,093
OrgReg	363,754
ORCID iDs	
ORCID iD	116,584

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 Licenses: 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 is bolstered by custom text similarity algorithms.

Additionally, OpenAIRE is systematically working on normalizing licenses. The table below provides an overview of the progress made in grouping and normalizing licenses. It is important to note that a merged record of a publication may include multiple licenses.

Table 25: Licence normalisation

Licence	*# Irish OA peer-reviewed publications*
CC-BY	30,561
CC-BY-NC-ND	13,679
CC-BY-NC	4,759
CC-BY-NC-SA	3,755
CC-0	664
CC-BY-SA	375
CC-BY-ND	97
CC-BY-ND-SA	22
Elsevier TDM	10,584

Licence	# Irish OA peer-reviewed publications
Elsevier OA	8,135
Wiley VOR	4,123
Springer TDM	5,618
Wiley TDM	4,611
arXiv non-exclusive distribution	3,055
Oxford Academic reuse	2,802
Sage TDM	2,595
CambridgeCORE	2,454
IEEE	2,377
APS	1,545
ASM TDM	1,042
% Irish OA peer-reviewed publications with non-normalised licenses	11.63%

As of now, approximately 89% of licenses have been successfully grouped and normalized. However, determining how these licenses should be accurately compared and categorized, particularly in the case of non-CC licenses, is a concern. For the construction of Hybrid OA publications (in a hybrid journal with a licence) we are currently only including CC licenses and are investigating which other ones should be included.

Authors: Researcher dashboards will be seamlessly integrated with their ORCID profiles. Originally, we had considered using text mining for author disambiguation. However, by anchoring the profiles to ORCID IDs, we have streamlined the process, allowing for easy identification of researchers *using all possible name variations* associated with their ORCID ID. This approach eliminates the need for additional text mining efforts.

3.4 Enrichment via Text Mining

In order to enrich metadata and enhance the comprehensiveness of scholarly records, OpenAIRE employs several effective 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.
- **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.

For the Monitor, we have successfully completed the PDF aggregation exercise, which involved collecting accessible PDF documents of Irish RPOs.

Thus, we have meticulously reviewed 84K URLs extracted from the metadata of the 226K Irish OA publications. Among these URLs, we have successfully retrieved 61K PDFs. Our ongoing efforts involve inspecting the remaining 23K URLs to further enhance our coverage.

These documents will pass through the graph pipeline, enabling the implementation of the text-mining methods mentioned above. We are also mining them to classify them into FoS (FoS) and Sustainable Development Goals (SDGs) and identify the corresponding author affiliation. Methodology for the latter is addressed in Section 3.6

FoS (FoS) Classification System: To categorize into distinct levels FoS[17], we have integrated an advanced classification system (Kotitsas, et al. 2023). 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: In order 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

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

adaptation, biodiversity preservation, mitigation of environmental contaminants, to socioeconomic upliftment.

3.5 Indicators

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

Table 26: Construction methodology of indicators

Attribute	Construction Methodology
Journal Business Models	
Open Access	Utilizing OpenAIRE's curated Gold ISSN list, Unpaywall metadata, and DOAJ journals.
Subscription	Journals without any Open Access articles.
Hybrid	Journals with Open Access articles that are not OA journals.
Transformative	We identify Transformative Journals by ISSN matching with the publicly available Transformative Journals data[18] from Plan S initiative.
Journal APC Business Models	
Diamond OA	We obtain APC data from DOAJ using DOAJ's Public Data Dump[19] (an exportable version of the journal metadata). We cross-reference this with the OpenAPC[20] data integrated in the OpenAIRE Graph.
OA Types/Colours	

[18] https://journalcheckertool.org/transformative-journals/

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

[20] https://openapc.net/

Attribute	Construction Methodology
Green OA	An Open Access scientific publication deposited in a repository
Gold OA	A scientific publication published in an OA journal as defined above.
Hybrid OA	An Open Access scientific publication published in a hybrid journal with an open license. At this point we consider only CC licenses "open"[21]. In principle, this means that we may be underestimating the number of hybrid OA articles and overestimating the number of bronze. In the final version of the report, we will adhere to the Unpaywall definition of "open licenses"[22] encompassing designations of the "ACS Editors' Choice", "APS License for Accepted Manuscripts" and "Open Access for APA Journals Authors". The inclusion of "Many other less-common licenses, as long as they grant users sufficient rights to freely use and redistribute content" will not be incorporated due to its lack of clear definition for inclusion.
Bronze	An accessible scientific publication published in a hybrid journal without an open license.
Accessibility – Interoperability	
Accessible	Accessible publications are sourced from OpenAIRE's full text collection, which holds PDFs of over twenty million OA publications. OpenAIRE's PDF aggregation system, recalibrated to prioritize Irish publications, examines the URL links in each publication's metadata to retrieve the corresponding PDF document. Given that multiple links can be associated with one publication, each is navigated. With ongoing automation, the system's coverage consistently expands, verifying if OA publications are accessible through thorough URL link inspection and PDF retrieval.
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.

[21] See discussion in Section 3.3

[22] https://support.unpaywall.org/support/solutions/articles/44002063718-what-is-an-oa-license-

3.6 Additional Aspects

In this section, we outline the construction methodologies for specific metadata elements vital to the Monitor. 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 in the Graph, we face limitations due to the lack of explicit tagging of this role in metadata from integrated data sources. While some sources are in development stages to provide such details, we devised two methodologies to bridge this gap:

1. ***Contributor Rank Analysis:*** We identify the corresponding author based on their position in the author list. Typically, *if the list is not alphabetical*, we consider the first author as the corresponding author. However, the sequencing of authors is not commonly represented in metadata, and we have managed to identify the first author in just 0.04% of Irish peer-reviewed publications using this method.
2. ***Text Mining:*** Exclusively for Irish publications, we are employing text mining on PDFs to discern the corresponding author's affiliation. It is worth noting that this approach is only possible for OA publications, as those PDFs are available.

Our final report will offer a comprehensive evaluation of the efficiency and limitations of these approaches.

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.," several steps are underway:

1. A comprehensive desk research is in progress to discern public RPOs and RFOs.
2. Survey currently conducted by IReL.
3. OFR's metadata offers pertinent information on funder type which is being integrated in the Graph. The OFR is donated by Elsevier, and is updated around every 4-6 weeks with new and updated funder records. Existing entries are also reviewed to make sure that they are accurate and up-to-date. The available funder types are "Government" and "Private". Private funding subtypes include: academic, corporate, foundation, international, other non-profit (private), professional associations and societies. Government funding subtypes include: federal (national government), government non-federal (state/provincial government).
4. Metadata from the European Commission's project participant RPOs includes their activity type and identifies the public ones.

Methodological Concerns: We have identified the following issues with the above methodology. Research might be funded indirectly by public spending, e.g., a private RPO receiving public grants or subsidies and then funding a research project, in which case the publication metadata may not include an acknowledgement of public funding.

As of now, within OFR, Science Foundation Ireland is the only entity classified as "government." Consequently, we have a total of 11,797 Irish publications that are publicly funded. The assessment of data quality will be conducted in the final report and will be enhanced by applying the aforementioned methodology to identify additional publicly funded Irish scholarly publications.

Detailed results and further insights will be provided in the final report.

4 Data Evaluation

In the context of Open Access (OA) monitoring, the quality and completeness of data are of utmost importance. This section conducts a thorough evaluation, focusing on essential metadata elements crucial for displaying data, indicators and filters on the platform. The following table provides a detailed overview of these elements, including their coverage and data quality within the context of Open Access in Ireland. Additionally, we offer practical comments and steps that both OpenAIRE and stakeholders can take to improve data accuracy and depth, ultimately enhancing the effectiveness of OA monitoring.

In the final version of the report, we will enhance the following table by offering more actionable information on stakeholder actions related to the assessment of metadata elements for data quality and completeness.

Table 27: Metadata Analysis

Metadata element & Issue	Quality	OpenAIRE Actions	Stakeholder Actions
Affiliations: Identification and association of research publications to Irish RPOs	The provenance of affiliation links in the Graph includes • Institutional data sources registered in OpenAIRE (repositories, CRIS, Open Access Journals) • Metadata from	Besides including about 129K data sources, the OpenAIRE pipeline includes affiliation text mining algorithms (see Section 3.4). OpenAIRE is focusing on refining data harvesting from Irish institutional sources and enhancing text mining specifically for Irish affiliations. The survey that is currently being conducted by IReL is expected to	Stakeholders can proactively register their institutional data sources with OpenAIRE[23], ensuring that the most up-to-date and comprehensive data is accessible and integrated into OA monitoring.

[23] Via OpenAIRE PROVIDE (https://provide.openaire.eu/home)

Metadata element & Issue	Quality	OpenAIRE Actions	Stakeholder Actions
	harvested data sources such as Crossref. - Inferred links via text mining. - Links created via the claim and link functionalities in OpenAIRE EXPLORE. Additionally, OpenORGs is utilized for name deduplication, ensuring consistency and precision in recognizing and distinguishing distinct RPOs and their outputs.	enhance the quality of these actions.	
Corresponding author affiliation: Determining the institution of the	0.04% publications with *first* author affiliation identified.	We currently have data on the first author of publications, and are text mining the collected PDFs to assess	Publishers, can assist by consistently highlighting corresponding author affiliations by making sure make sure the corresponding author is clearly indicated in published articles, facilitating easier identification through automated processes.

Metadata element & Issue	Quality	OpenAIRE Actions	Stakeholder Actions
corresponding author.		the possibility of identifying the corresponding author affiliation (see Section 3.6).	
Funded publications: Establishing comprehensive funder-publication links.	While SFI (Science Foundation Ireland) is curated due to having joined OpenAIRE, for other public funders, we rely on the harvested metadata. Its quality is difficult to assess externally.	OpenAIRE is still in the process of analysing data quality for integrated Irish Funders (those via OFR). The survey that is currently being conducted by IReL is expected to additionally enhance the quality.	Funders can join OpenAIRE by providing a project list, to allow us to build dedicated text mining algorithms that discover project-publication links and clean harvested metadata. Stakeholders can ensure clear labelling of funding sources in metadata and acknowledgements eg: For RPOs by having populated the respective fields in their metadata records according to their compatibility with the OpenAIRE Guidelines: • "Project Identifier" for v.3 • "Funding Reference" for v.4 For RFOs by having Recommended acknowledgement statements for the funded research outputs thus clearly stating the project reference. Additionally, we will investigate and utilise if applicable the IReL OpenAPC dataset as it includes the DOI and the RPO, which by definition of the IReL OA publishing agreements is the RPO the corresponding author is staff or student of. Funding sources can have PIDs.
Grant award ID: Ensuring availability &	The quality assessment of this metadata element	OpenAIRE is in the process of analysing data for Irish Funders	As above

Metadata element & Issue	Quality	OpenAIRE Actions	Stakeholder Actions
establishing project-publication links.	will be detailed in the final report. SFI has full coverage since they have joined OpenAIRE and provided a project list.	(those in OFR) to examine the metadata coverage of this variable. For funders that have joined OpenAIRE (SFI for now) we have Grant award IDs and mine for project-publications links.	
Publicly-funded publications: Identification of all publications resulting from publicly funded research	Data Quality to be assessed in final report. As of now, within OFR, Science Foundation Ireland is the only entity classified as "government." Consequently, we have a total of 11,797 Irish publications that are publicly funded.	OpenAIRE is working on identifying the list of public RFOs and RPOs. The survey that is currently being conducted by IReL is expected to enhance this list. See Section 3.6 for a discussion.	As above
Publication type: Differentiating between	It is challenging to assess the quality of this field. About 3.3% of publications	Comparing date of deposition in a repository and date of publication of an	Publishers and repositories can ensure clear metadata tags for Version of Record and pre-print instances. This is applicable in the DOI metadata for publishers and in the metadata records of the repositories, following the OpenAIRE

Metadata element & Issue	Quality	OpenAIRE Actions	Stakeholder Actions
Version of Record (VoR) and pre-print.	labelled as articles are also explicitly labelled as pre-prints, but checks have shown that there are more. Version of record is not a commonly shared metadata element and the sources that do expose it have low coverage.	article. If the article was deposited after it was published it is likely that it is the VoR that was deposited. We are currently examining this option.	Guidelines. (e.g. for the latest version: https://openaire-guidelines-for-literature-repository-managers.readthedocs.io/en/v4.0.0/field_publicationtype.html)
PIDs: Identifying and ensuring consistent coverage of Persistent Identifiers (PIDs) for publications, organisations, and authors.	As of now all Irish Peer-Reviewed publications have been identified as having a PiD. For authors right now we have almost 50% coverage of peer-reviewed publications with authors with an ORCID iD.	For the Organisations the OpenOrgs platform is essential in order to disambiguate the organisation different names and PiDs. For authors, the ORCID iDs publications coverage can be further enriched by several methods including the organisations disambiguation in OpenOrgs, the signing in of the Irish	Stakeholders, especially publishers and repositories, are encouraged to consistently apply and expose PiDs for publications, authors, and organisations. For repositories this is applicable in the metadata records of the repositories, following the OpenAIRE Guidelines. (eg. for the latest version: https://openaire-guidelines-for-literature-repository-managers.readthedocs.io/en/v4.0.0/field_resourceidentifier.html)

Metadata element & Issue	Quality	OpenAIRE Actions	Stakeholder Actions
		researchers into the Monitor providing their ORCID iD, the claiming and linking of research products functionality.	
Peer-reviewed publications: Identifying and verifying the peer-review status of Irish publications.		Refer to Section 3.2 for a discussion.	Publishers can label journals or proceedings as peer-reviewed in designated metadata fields[24].
Year of publication: Low(er) coverage for 2022	We have 3.2K peer-reviewed Irish publications in 2022, whereas in 2020 and 2021 they are 16.1K and 16.7K, respectively	The delay in the availability of 2022 Irish peer-reviewed publications, as of October 2023, can be attributed to the natural timeline associated with academic publishing. OpenAIRE is actively addressing this delay by focusing on two main strategies: refining data	Self-deposition of peer-reviewed articles in repositories.

[24] We will further investigate to improve the suggestions for these metadata fields.

Metadata element & Issue	Quality	OpenAIRE Actions	Stakeholder Actions
		harvesting from Irish institutional sources, enhancing text mining for affiliations specifically for Ireland and re-examining the peer-reviewed criteria.	
Publisher & Journal: Challenges in deduplication	Multiple instances of the same publisher or journal may be present due to variances in naming conventions, abbreviations, or metadata inconsistencies.	OpenAIRE has started the deduplication of publishers and journals (see Section 3.3).	Publishers can provide authoritative lists of names and identifiers.
Licence: Challenges in deduplications and interpretation of licence agreement	While about 89% of licenses have been grouped and normalized, there remains a challenge in categorizing non-CC licenses.	OpenAIRE has started the deduplication of licenses (see Section 3.3).	
FoS: Coverage and potential under-representation	Currently the FoS coverage for the peer-reviewed	OpenAIRE will classify the remaining Irish publications by running them through our FoS classification system to	We encourage researchers, institutions, and other stakeholders to provide feedback on any observed discrepancies in the classification.

Metadata element & Issue	Quality	OpenAIRE Actions	Stakeholder Actions
of specific disciplines	publications is 22.71%.	ensure comprehensive coverage (see Section 3.4).	
SDGs: Coverage	Currently the SDGs coverage for the peer-reviewed publications is 1.81%.	The remaining Irish publications are scheduled to undergo SDG classification to ensure comprehensive coverage (see Section 3.4).	Feedback on perceived mismatches or overlooked research contributions towards SDGs is encouraged.
Access rights 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.	After the Pilot Monitor's delivery on M6, OpenAIRE will begin uploading regular data dumps to Zenodo. Additionally, we will display the progression of indicators like % OA, % Green OA, etc., on the trendline graphs, utilizing monthly historical snapshots.	For the embargoed access repositories can expose the Embargo Period Date according to the OpenAIRE Guidelines. https://openaire-guidelines-for-literature-repository-managers.readthedocs.io/en/v4.0.0/field_embargoenddate.html

Metadata element & Issue	Quality	OpenAIRE Actions	Stakeholder Actions
URLs to PDF full texts: Coverage and validity of URLs linking to PDFs in publication metadata.	Of the 84K OA publications tested thus far, 61K had accessible PDF files linked directly from their metadata URLs.	OpenAIRE will continue testing Irish URLs as they show up in publication metadata.	Publishers and repositories could enhance metadata integrity by providing accurate, unambiguous links to PDFs.

5 Conclusion

In this draft report, we've endeavoured to provide a transparent and detailed overview of our findings from the National Open Access Monitor. It serves as both an evaluative tool and a roadmap, pinpointing the accomplishments in the OA landscape and highlighting areas that warrant further examination and action.

To sum up, for the current baseline analysis, the Irish publications database encompasses a total of 345,316 publications. Among these:

- 244,156 (70.7%) have been identified as Peer Reviewed.
- 226,025 (65.45%) have been classified as Open Access.

The baseline analysis has provided us with valuable initial insights into the Open Access landscape in Ireland. It serves as a solid foundation for our understanding of the current state of Open Access publication trends in the country. However, it's important to note that this is just the beginning. We are actively engaged in a process of continuous improvement, working diligently to refine our data and analysis.

The ongoing enhancements we are currently implementing will significantly enrich the Monitor Data, allowing for a deeper and more nuanced understanding of Open Access within Ireland. These improvements are not merely technical; they represent our commitment to delivering high-quality, reliable data that can inform decisions, policies, and strategies related to Open Access.

As we move forward, the Monitor's data will become an even more valuable resource for researchers, institutions, and policymakers. It will serve as a powerful tool for tracking the progress of Open Access initiatives, identifying areas for improvement, and fostering a culture of transparency and openness in Irish research landscape. Our dedication to enhancing data quality and expanding coverage ensures that the Monitor will continue to play a vital role in advancing Open Access in Ireland and beyond.

The final iteration of the report will build upon the data and insights shared here, offering a holistic perspective of Ireland's OA environment, including challenges identified and recommended solutions.

References

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.