



SCIENCE
EUROPE
Shaping the future of research

OPEN ACCESS MONITORING

Guidelines and Recommendations
for Research Organisations
and Funders

BRIEFING
PAPER

May
2021

Colophon

May 2021

"Open Access Monitoring: Guidelines and Recommendations for Research Organisations and Funders"

Available at: <https://scieur.org/monitoring-oa>

Authors: Tobias Philipp (Swiss National Science Foundation), Georg Botz (Max Planck Society), Jean-Claude Kita (Fund for Scientific Research F.R.S.-FNRS), Paul Richards (UK Research & Innovation), Astrid Sanger (German Research Foundation), Olaf Siegert (Leibniz Association), Mathilde Reumaux (Science Europe).

Acknowledgments: The authors would like to thank the members of the Science Europe Working Group on Open Access to Research Publications (chaired by Georg Botz), participants at the Science Europe Workshop on Open Access Monitoring (October 2020), and Lidia Borrell-Damian (Science Europe) for their contributions.

Editors: Lorna Stokes and Iwan Groeneveld (Science Europe).

For further information please contact:
office@scienceeurope.org

© Copyright Science Europe 2021. This work is licensed under a Creative Commons Attribution 4.0 International Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original authors and source are credited, with the exception of logos and any other content marked with a separate copyright notice. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/> or send a letter to Creative Commons, 444 Castro Street, Suite 900, Mountain View, California, 94041, USA.





OPEN ACCESS MONITORING:

Guidelines and Recommendations
for Research Organisations
and Funders





Table of Contents

1. Introduction	7
2. WHY Measure Publication Status?	8
2.1 Setting basic parameters – Recommendations	9
3. WHAT to Measure?	10
3.1 Demarcating target publications	11
3.2 Sampling sets of publications – Recommendations	13
4. HOW to Collect and Interpret Data on Publications?	14
4.1 Persistent and unique identifiers	15
4.2 Collecting data on publications	15
4.3 Interpretation and reporting	18
4.4 Gathering and interpreting data on publications – Recommendations	21
5. Sustainable and Comparable Monitoring	22
6. Conclusion and General Recommendation	22



Science Europe and its Member Organisations are committed to supporting and contributing to a smooth and rapid transition to Open Access at national and European level. They work together to develop the best possible solutions to practical aspects of the transition, and support the implementation of full Open Access policies that are adapted to different contexts. Science Europe's Member Organisations share insights, methods, sources, tools, and best practices to adapt the system and develop a culture of sharing. For more information on Science Europe's activities on Open Access, please visit its website at <https://scieur.org/oa>

1. Introduction

The global transition to Open Access has accelerated in the past few years. This transition is complex and involves a variety of approaches and multifaceted strategies. Many research stakeholders, including but not limited to research funders (RFOs) and research performing organisations (RPOs), have adopted Open Access policies and established supporting measures such as provision of funding and development of infrastructures and services to foster the transition.

The availability of data and information on the current state of scholarly publishing, be it institutional, national, and/or in an international context, is invaluable to help advance Open Access, including to inform policy development, implementation, and evaluation.

To gather such information, monitoring exercises are conducted, taking stock of publication output and measuring its Open Access status. Given the complexity of the scholarly publishing system, this is no trivial task and involves a multitude of decisions.

This briefing paper aims to support decision-makers at RFOs and RPOs develop new monitoring exercises or assess and improve existing processes to measure the Open Access status of publications. It is not a technical manual but seeks to bridge the gap between the technical aspects of Open Access monitoring and the strategic decisions necessary to implement such an exercise.

It proposes three main steps an organisation should take to develop a monitoring exercise. It begins by defining the purpose of the exercise, continues with identifying the scope of publications for which information is needed to inform the purpose, and finishes with practical considerations for data collection, interpretation, and reporting. For each step, the briefing identifies specific challenges and provides general recommendations on how to navigate them. Examples of different monitoring exercises have been selected to represent different use cases, organisational setups, data sources, and strategies of interpretation.

This briefing paper focuses on monitoring scholarly research articles, but many of the considerations also apply to academic books and other publication types, although differences are acknowledged, such as metadata availability.

Monitoring the costs of Open Access publishing is recognised as an essential issue for RFOs and RPOs. However, this is not included in the scope of this document due to the diverse funding and financial practices at national and institutional levels.

81%

of Science Europe Member Organisations who responded to a 2018 internal survey, declared that they planned to (further) develop Open Access monitoring mechanisms in the future.

Out of 36 Member Organisations, 26 responded to the survey on Monitoring Compliance with Open Access Policies.

2. WHY

Open Access monitoring exercises serve a range of purposes, from taking a snapshot of publication numbers, status, and venues, to trying to evaluate the impact of existing policy measures. Achieving actionable results by collecting and analysing a set of data on scholarly publications must take organisation-specific demands and requirements into account. To set up an effective monitoring exercise, its aims must be defined as clearly as possible.

2. WHY Measure Publication Status?

Assessing the Open Access status of scholarly publications emanating from particular funders or institutions is crucial to the timely transition to Open Access in the scholarly publishing system.

There are numerous guidelines and examples that can inform the development of an Open Access strategy or policy at institutional, organisational, or even national level.¹ However, general guidelines or orientations for Open Access strategy development may not be enough.

Implementing an Open Access policy and identifying measures to support it require current knowledge on the status quo of scholarly publishing: How many publications are there each year? By whom? Where

are they published? How accessible and reusable are they?

Before devising and implementing a monitoring exercise, an organisation should clearly define the objectives and information it wants to extract. Requirements can differ considerably, for example, between monitoring at national or institutional level and between RFOs and RPOs.

If an Open Access strategy or policy is already in place, information requirements could be derived from this. Additional external requirements may also have an influence, for example, discussions on scientific impact and communicating science to broader audiences.

2.1 Setting basic parameters – Recommendations

Organisations setting up a monitoring exercise should, at the very least, address the following four key aspects:

1. **Purpose:** Should results guide policy development, evaluation, compliance assessments, implementation of supporting measures, negotiations with publishers or communication towards individual disciplines, institutions or scholars?
2. **Comparability:** Should outputs of the monitoring exercise be compared to pre-existing monitoring outputs, which might prescribe a specific approach and methodology?

3. **Scope and level of granularity:** Should results be reported on the level of individual scholars, institutes, organisations, nations or a combination of all these?
4. **Timeframe:** Will results for one point in time suffice or does the analysis need to be repeated, investigating changes over time?

Having a solid idea of what needs to be achieved by a given monitoring exercise is essential. With this in place, organisations should then decide on what to monitor and how to collect and interpret data on publications.

Monitoring national progress towards Open Access

Universities UK

In 2017 the multi-stakeholder Universities UK Open Access Coordination Group published its second report looking at UK progress towards Open Access. Using Scopus data and various other data sources, the report considered trends in various aspects of Open Access publication such as publication status, licensing, and costs.^A



A. <https://www.universitiesuk.ac.uk/policy-and-analysis/reports/Pages/monitoring-transition-open-access-2017.aspx>

1. An example is the Science Europe Principles on Open Access to Research Publications, 2015: <https://scieur.org/opennew>

3. WHAT

With purpose and information needs clearly defined, the set of publications to analyse needs to be determined. Typically, target publications are identified by authorship, institutional affiliation, or funding sources. The description of this set also indicates what metadata must be obtained.

3. WHAT to Measure?

Over three million new scholarly articles are published each year in English-language journals, in addition to an uncounted myriad of articles in other languages, books, book chapters, proceedings, and

other publication types.² Every monitoring exercise needs to select a specific sample of publications from this huge volume, based on the specific purposes of the exercise.

3.1 Demarcating target publications

Scholarly publishing takes place via a variety of publication types. To date, the predominant focus of Open Access has been scholarly articles. Yet, there are several other important types such as books, proceedings, and book chapters for which Open Access is becoming more common. When setting up a monitoring exercise that examines publication status, two primary decisions have to be taken: which types of publications are to be covered; and the set of publications to be considered.

Although this document focuses on scholarly articles, many of the considerations and recommendations outlined below can also apply to other types of publications.

Target publication types

Peer-reviewed scholarly articles published in journals are, by far, the most analysed type of academic publication. Standardised workflows of article publishing and discovery have resulted in well-developed infrastructures, providing general and Open Access-specific metadata. Hence, scholarly articles are the least challenging type of publication to monitor and have been the focus of every large-scale monitoring exercise to date.

Going beyond the scholarly article might require additional approaches to gather, analyse, and interpret publication data. Open Access to academic books is at an earlier stage of development and the publishing landscape differs considerably from that of scholarly articles.³ To monitor the publication status of books, book chapters, and proceedings, additional metadata are needed. For example, to differentiate between editors of books or anthologies and authors of individual contributions. The infrastructures and data sources for other types are not yet as developed as they are for articles. The advent of Open Access publishing platforms currently adds complexity to this landscape.

Target authors and funding sources

RFOs are primarily interested in monitoring publications emanating from their own funding. RPOs typically monitor publications that have been authored by affiliated researchers. To satisfy these interests when determining the set of publications to monitor, one should consider the following:

- Performing and publishing research is often a highly collaborative and international endeavour. A multi-author publication might only be attributed to the corresponding author's institution, or could be attributed regardless of position in the author line. Disciplinary differences in the role and meaning of first or last authors need to be considered.
- Acknowledgements of funding and establishing traceable links between publications and funding sources facilitate monitoring. However, since it is often the case that funding from more than one source is acknowledged, a practice needs to be established on how to handle such cases.

2. The STM Report: An overview of scientific and scholarly publishing (1968–2018), International Association of Scientific, Technical and Medical Publishers, 2018, p. 5: https://www.stm-assoc.org/2018_10_04_STM_Report_2018.pdf

3. Science Europe Briefing Paper on Open Access to Academic Books, 2019: <https://scieur.org/oa-books>
A Landscape Study on Open Access and Monographs: Policies, Funding and Publishing in Eight European Countries, Knowledge Exchange, 2017: https://repository.jisc.ac.uk/6693/1/Landscape_study_on_OA_and_Monographs_Oct_2017_KE.pdf

Open Access policies of RPOs and RFOs are of great significance when deciding which outputs should be included in an analysis. They can, for example, decide whether a particular publication is covered by the Open Access policy of an organisation and should thus be monitored.

Since authorship alone does not reflect the diversity of contributions culminating in a publication, other approaches are being developed. One example is the Contributor Roles Taxonomy CRediT,⁴ which provides a fine-grained framework to attribute contributions to scholarly outputs. Future monitoring exercises might rely on such additional information to define in greater detail which contributions should be considered.

Monitoring the compliance with funder's Open Access policies

Austrian Science Fund

FWF

Der Wissenschaftsfonds.

In order to evaluate researchers' compliance with their Open Access requirements, the Austrian Science Fund (FWF) monitors the Open Access rate of peer-reviewed publications in FWF-funded projects.^A The basis for the calculations is the Open Access status of publications at the time of the final project reports.

A. <https://www.fwf.ac.at/de/ueber-den-fwf/publikationen/publication-types/43/>



4. <https://casrai.org/credit/>

3.2 Sampling sets of publications – Recommendations

A solid understanding of what should be measured is necessary to ensure efficient data collection and to guide the interpretation process. Organisations setting up a monitoring exercise should address the following:

Publication types

- The selection of publication types to monitor must reflect the scope of the monitoring exercise and publishing practices of covered researchers. Disciplinary differences regarding the role of scholarly articles, books, book chapters, or proceedings must be taken into account.

Authors and funding

- The minimum requirement to attribute a publication to a RPO should be a specific affiliation of any of its authors, irrespective of the author's position in the author line.
- To attribute a publication to a RFO there should be an acknowledgement of the received funding in the publication itself.
- If the scope of a monitoring exercise needs to be limited, established metadata should be used, such as only considering certain classes of authorship (corresponding author, first or last author) or funding (only acknowledgements stating specific grants, limiting the number of RFOs involved).

Metadata

- Only consider publications as Open Access that are available through legitimate channels, and disregard sources such as Sci-Hub. Access to publications in academic social networks is often limited to members of those networks, thus not qualifying as Open Access. Furthermore, institutional websites are not a reliable location to host publications and should be disregarded, as long as they are not linked to an institutional or disciplinary repository.
- Metadata should be collected describing the following key characteristics of Open Access publications:
 - Verification of the open and free accessibility of the full text of the publication online.
 - Version of the available full text: a preprint, the text submitted for publication, the author's accepted manuscript (AAM), or the final version of the record (VoR). This may also indicate whether a publication has passed quality assurance processes.
 - Licence applied to the available text (if any), and accordingly possibilities and limitations to use and re-use the text.
 - Location of the available text: sustainable and legal source, ensuring discoverability and long-term preservation.

4. HOW

There is no single, comprehensive source that provides complete coverage of the publication data and specific metadata required to monitor Open Access status. For most monitoring exercises, data from multiple sources will need to be gathered, aggregated, and reconciled. Being able to unambiguously identify authors, publications, institutions, and contributing funding sources through persistent identifiers is crucial to generate a reliable dataset. The advantages, shortcomings, and potential biases of different data sources need careful consideration. Uncertainties have to be accounted for when interpreting such data. When reporting results of any monitoring exercise, transparency is required on all steps taken prior to, and during the analysis, as well as limitations.

4. HOW to Collect and Interpret Data on Publications?

The collection and interpretation of data on publications must build on well-defined selection criteria. Before undertaking any larger-scale analysis, it is advisable to conduct a small-scale exploratory analysis, to assess if the chosen approach fulfils the purpose of the exercise. During such an exploration, the chosen approach might prove to be too narrow

or too wide in scope, for example, by missing too many relevant or including too many irrelevant publications. This can be used to improve the definitions of what to monitor and lead to a more efficient data collection and interpretation process, potentially improving the alignment of data collection and overarching goals of the monitoring exercise.

4.1 Persistent and unique identifiers

Monitoring exercises require the ability to unambiguously identify publications, authors, and associated research institutions. Identifying involved funders and specific grants is also a frequent requirement. Monitoring workflows relies on unique and persistent identifiers (PIDs) in a machine-readable format, which, for example, act as proxies for full names of authors, institutions, or proprietary grant numbers. Such standardised identifiers solve the challenge of disambiguation. They help to differentiate between scholars with the same name, different ways to write or abbreviate institutions and they can help to overcome typing errors. Without PIDs, even small-scale analyses of a few hundred publications involve a large amount of manual data processing, matching, and checking rendering them difficult to conduct, let alone in a transparent and interoperable fashion.

Current monitoring workflows already rely on widely available Digital Object Identifiers (DOIs) for publications.⁵ Identifiers for researchers are not yet standard but are becoming more widespread (for example, ORCID⁶). Depending on the use case, a clear differentiation between research institutions can also benefit from using identifiers (for example, Research Organisation Registry ROR⁷). In addition to publications, authors, and institutions unique identifiers for grants and other modes of research funding complete the picture and further improve monitoring workflows. Such identifiers are not yet widespread, the first major service having been launched by Crossref in 2019.⁸

4.2 Collecting data on publications

There is no single source that provides a comprehensive and ready-to-use dataset for Open Access monitoring purposes. Given the diversity of use cases and the heterogeneity of scholarly publishing, monitoring almost always requires aggregating data from different sources.

Data collection for monitoring purposes is a three-step process:

1. A list of target publication outputs is generated.
2. Open Access specific metadata on these publications is collected.
3. The data obtained from different sources is aggregated into one dataset. This step comprises cleaning, converting, and reorganising data into an interoperable format that is usable for the specific analysis.

5 <https://www.doi.org/>

6 <https://orcid.org/>

7 <https://ror.org/>

8 <https://www.crossref.org/community/grants/>

Combination of data sources to monitor the implementation of a national Open Access strategy

Danish Open Access Indicator

Since 2018, the Danish Service for Science and Higher Education produces the annual Danish Open Access Indicator.^A Results are interpreted and presented within the framework of the national strategy. For the indicator, a standardised workflow builds on data from Danish universities' own CRIS, enriched with metadata from the Directory of Open Access Journals and SHERPA/RoMEO^B which is a service providing information on publishers' Open Access policies.



The French Open Science Monitor



The French Open Science Monitor^C was developed in the context of the French National Plan for Open Science. As in Denmark, it aims to steer Open Science in the country and to monitor Open Access trends on a regular basis. The tool is based on openly available data and makes use of large-scale systems such as Unpaywall, HAL^D (the main open repository in France, part of the French National Centre for Scientific Research (CNRS)),

ORCID, and IDRef (referential for French Higher Education and Research). It uses rule-based and machine learning techniques to enrich the metadata of the publications.

A. <https://www.oaindikator.dk/en/>

B. <https://v2.sherpa.ac.uk/romeo/>

C. <https://www.ouvrirelascience.fr/the-french-open-science-monitor/>

D. <https://hal.archives-ouvertes.fr/>

Sources for publication data

Since scholarly publications are analysed for various purposes, not just monitoring Open Access, there is a market for this kind of data. Currently, a small number of commercial providers dominate this market, most notably Clarivate Analytics (Web of Science)⁹, Elsevier (Scopus)¹⁰ and Digital Science (Dimensions).¹¹ However, other open data sources, such as Crossref,¹² CORE,¹³ OpenAIRE,¹⁴ or Microsoft Academic,¹⁵ may become interesting alternatives.

Keeping track of publication output is also of concern for RFOs and RPOs. In many such organisations, their Current Research Information Systems (CRIS) are used to systematically collect and provide metadata on their research activities, including scholarly publications. Other potential data sources are national, institutional, and/or disciplinary repositories, which provide the full text and metadata

of archived publications. Three prominent examples of disciplinary repositories are PubMed Central and Europe PMC for the biomedical and life sciences, and arXiv for physical sciences. All of these sources can also provide publication data that are useful for monitoring, in addition to their primary functions.

Every data source has benefits and downsides. Multi-purpose publication databases such as Web of Science, Scopus, or Dimensions collect and curate publication data on their own, according to their specific selection criteria, and as part of their business model. They do not require authors or institutions to actively register publications themselves. Their coverage is broader than that of any other mentioned source, although still incomplete. English language publications are better covered than publications in other languages, and publications from the Social Sciences and Humanities disciplines are under-represented.¹⁶ Multi-purpose databases often provide

9. <https://clarivate.com/webofsciencelgroup/solutions/webofscience-platform/>

10. <https://www.scopus.com/>

11. <https://www.dimensions.ai/>

12. <https://www.crossref.org/>

13. <https://core.ac.uk/>

14. <https://www.openaire.eu/>

15. <https://academic.microsoft.com/home>

16. Kulczycki, E., Engels, T.C.E., Pölonen, J. et al., Publication patterns in the social sciences and humanities: evidence from eight European countries, *Scientometrics* 116, 463–486 (2018): <https://doi.org/10.1007/s11192-018-2711-0>

easily available data on corresponding authors and grant acknowledgement statements. This facilitates the attribution of publications to research institutions.

On the other hand, CRIS and institutional repositories often incentivise or require researchers to enter their publications into those systems. Institutions running such infrastructures usually actively curate the data themselves. Attribution of publications to a specific institution can thus be derived or further improved from publications being present in the institutions' databases, but sometimes CRIS do not capture all of the metadata elements that could be relevant, such as corresponding authors, and the quality of metadata often varies.

Depending on the vendor, workflows to retrieve data from proprietary databases and public repositories vary. Commercial providers require licences to access their services, which vary in price and access type, for example, if they provide raw data or an API.

Sources for Open Access specific metadata

Unpaywall,¹⁷ provided by Our Research, is currently the most frequently used service to gather Open Access specific metadata. By using DOIs to uniquely identify publications, the service provides detailed publication characteristics. For example, it indicates whether an article is published in a fully Open Access journal and what version is openly available under which licence. It does not offer information on the embargo periods applied to individual articles, however. Unpaywall aggregates data on scholarly publications from a number of different sources,

including publishers and institutional repositories and it uses the Directory of Open Access Journals (DOAJ)¹⁸ to distinguish between open and toll access journals. Other international non-profit or public organisations have the primary function of providing common metadata on scholarly publications. For example, CrossRef is dedicated to creating and linking metadata on academic publishing and OpenAIRE aggregates metadata from other public providers throughout Europe. Finally, the ESAC initiative¹⁹ provides information on transformative agreements in place, making it possible to differentiate this form of Open Access from hybrid publications.

Aggregation and cleaning

Data obtained from various sources have to be merged into a single dataset. This is not a trivial task. The efficiency and accuracy of this aggregation depends heavily on the availability and use of PIDs across the different datasets that are to be combined. Usually, some manual standardisation is required, for example, affiliation of corresponding author or grant acknowledgement statements. Furthermore, the data will have to be cleaned and validated: double entries need to be removed; authors, research institutions, and funders might have to be disambiguated; and the dataset needs to be checked for potential errors. Finally, the data must often be converted and reorganised into an interoperable format that is suitable for the intended analysis. The complexity of this multi-stage process is often underestimated. It needs additional time and expertise thus sufficient resources need to be allocated. Typically, a balance must be struck between the goal of generating the best possible dataset and inevitable resourcing.

Negotiations of Transformative Agreements with publishers

Open Access 2020

The Open Access 2020 initiative^A aims to support the transition to Open Access by transforming today's scholarly journals from subscription to Open Access. One means of action for research organisations is the negotiation of Transformative Agreements with publishers. In this process, having robust data on the publications in question and their Open Access status is highly valuable. The ESAC Data Analytics Working Group focuses on data-related issues in the context of Open Access publishing agreements.^B



A. <https://oa2020.org/take-action/>

B. <https://esac-initiative.org/about/data-analytics/esac-data-working-group/>

17. <https://unpaywall.org/>

18. <https://doaj.org/>

19. Efficiency and Standards for Article Charges (ESAC): <https://esac-initiative.org/>

4.3 Interpretation and reporting

Only in rare use cases is the goal of monitoring achieved only by providing data. Reporting on scholarly publishing necessitates synthesising often complex information, arising from diverse practices, into actionable advice for different internal and external stakeholders. The categorisation of publications by 'Open Access colours' such as 'gold' and 'green' is one approach used frequently to reach a level of detail that is efficient to communicate and easier to digest.

Some form of categorisation – the most basic being a differentiation between open and closed publications – will almost always be applied. To ensure transparency and comparability of monitoring results, it is of vital importance for categorisation to be carried out in a comprehensible way. Open Access categories are never self-explanatory and differences in application can lead to serious misunderstandings. For example, there are different views on whether publications are considered 'open' as soon as their full text is available online, or whether they must be available under a specific open licence. Thus, monitoring exercises must inform about the specific categorisations applied and should explain why they are applied in that way.

Multi-use initiatives

German Open Access Monitor



The publicly funded German Open Access Monitor^A measures the publication output of German research institutions while also reporting on their Open Access status. The way they interpret and present the data is very user-friendly and facilitates re-use. Publication data is obtained from Dimensions and Web of Science and additional Scopus data should be added in 2021. The metadata on Open Access status is gathered from Unpaywall.

Australian Curtin Open Knowledge Initiative (COKI)

In an effort to give insights into the Open Access status of research publications worldwide, the Australian Curtin Open Knowledge Initiative (COKI)^B draws on data from public sources around the world, giving an overview over Open Access statistics for a large number of countries. Publication data are collected from Crossref and Microsoft Academic. Information from Crossref's Funder Registry^C and from the Global Research Identifier Database (GRID)^D is used to link publications to funders and to research institutions.

A. <https://open-access-monitor.de/>

B. <http://openknowledge.community/dashboards/coki-open-access-dashboard/>

C. <https://www.crossref.org/services/funder-registry/>

D. <https://www.grid.ac/>

Uncertainty in monitoring scholarly publishing

Some uncertainties and sources of error are impossible to avoid when carrying out monitoring exercises: for example, the constant flow of new publications and the inevitable time lag between publication and discoverability of the publication itself and all the needed metadata. Moreover, the Open Access status of a single article can change over time. This happens, for example when an embargo runs out and it becomes available on a repository, or the publisher uses a 'rolling paywall', making the version of record openly accessible after a certain delay.

The variety of publication types, disciplinary publishing practices, and sheer volume make complete coverage by individual databases and repositories unlikely. When interpreting gathered data, and especially when reporting results, this uncertainty has to be accounted for and should be openly communicated. For example, if it is known that a specific discipline will be under-represented in the data sources used in the monitoring exercise, or if a type of publication important to that discipline was excluded, this needs to be communicated along with the detailed results.

There will also be a number of articles in every monitoring exercise that do not fit predefined Open Access status categories. Typically, these publications cannot be classified as 'open' or 'closed' and will require a residual category such as 'other'. Such a category will inevitably pose a challenge when reporting results; well-chosen examples go a long way in explaining the reasons for such a category. In cases that are hard to interpret it might be considered to reach out to responsible authors or institutions for clarification, mitigating potential misunderstandings and increasing acceptance for results when publicly reporting.

Transparency in monitoring processes

To fully understand the final results of any monitoring exercise and ensure their robustness and integrity, the data used, the analysis steps taken, and underlying assumptions and limitations need to be made transparent. As there is no single approach to monitoring, and no generally accepted standard with regard to data collection and data processing, monitoring results are never self-explanatory. When reporting results, the following information should be provided as a minimum:

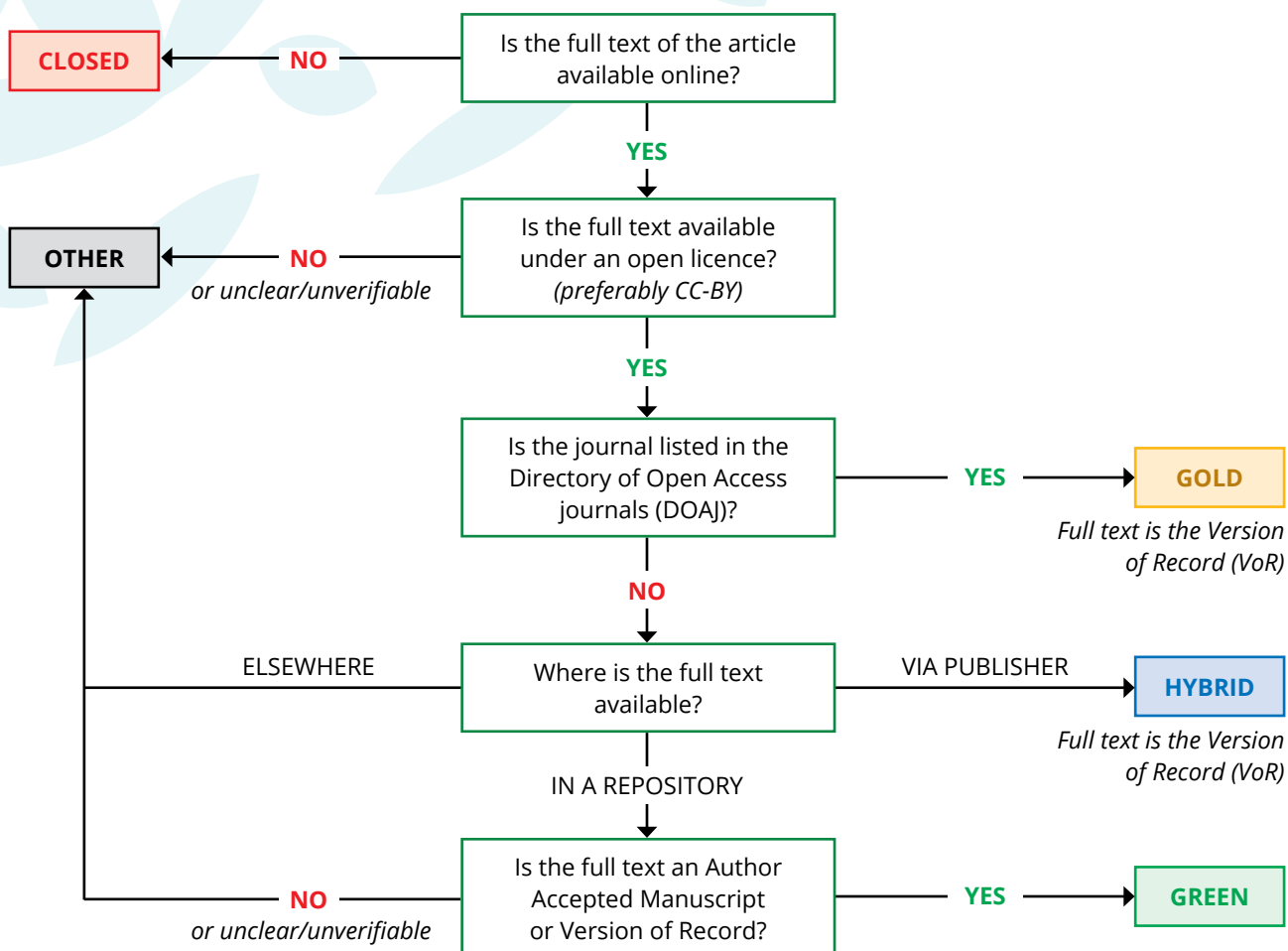
- Time of data collection.
- Which classes of publication types and groups of authors were analysed.
- How the data was sourced.
- How different Open Access categories were defined.

Example of categorising publication status – From publication characteristics to Open Access colours

The flowchart on page 20 is a minimal example on how to distinguish articles into three Open Access categories and one residual category. It assumes that articles available from sources such as academic social networks, institutional websites, or Sci-Hub are not to be considered Open Access. It is based on the Science Europe Principles on Open Access to Research Publications,²⁰ which build on the widely accepted core principles of the Berlin Declaration.²¹

This example illustrates the logic behind such a categorisation, applications to different use cases might vary.

From Characteristics to the Colours of Open Access Publishing



Open Access in international institutional comparison

CWTS Leiden Ranking



Since 2019, the CWTS Leiden Ranking^A includes the number and proportion of Open Access publications as one metric to assess the scientific performance of 1000 universities worldwide. CWTS uses publication data from Web of Science and Open Access metadata from Unpaywall. Their indicators differentiate different types of Open Access using Open Access colours.^B

A. <https://www.leidenranking.com/>

B. <https://www.leidenranking.com/information/indicators>

20. <https://scieur.org/opennew>

21. Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities (2003): <https://openaccess.mpg.de/Berlin-Declaration>

4.4 Gathering and interpreting data on publications – Recommendations

Organisations conducting a monitoring exercise should consider the following when gathering and interpreting data on publications:

Identifiers

- Persistent unique identifiers drastically simplify workflows and increase reliability. They should be used whenever possible as they improve reliability and comparability.
- Open or community-governed services should be chosen over proprietary offers to avoid lock-ins with single providers and foster the development of open infrastructure.
- For publications, DOIs are required.
- For authors, ORCID is recommended. However, since not every author has registered an ORCID, monitoring exercises cannot only rely on this identifier.
- For institutions, services such as the community-led Research Organization Registry (ROR) should be used.
- For funding sources, services such as Crossref's Funder Registry are recommended.

Data sources

- No single data source can guarantee coverage of the complete publication output even of a single research institution. Data sources used should be carefully checked for 'blind spots' and potential biases, for example, disciplinary representation.
- It is recommended to include publication data available from local or national sources such as CRIS or institutional or national repositories whenever possible, as they regularly cover publications not indexed in multi-purpose publication databases.
- Coverage of general multi-purpose publication databases is typically much lower for publications in languages other than English. This needs to be considered when choosing data sources to report on research areas or disciplines that do not predominantly publish in English.

Data aggregation

- The compilation of a reliable dataset is a complex task that should be budgeted for accordingly.
- Aggregating, cleaning, and analysing data on publications is a frequent task in the field of bibliometrics. It is recommended to consult experts from that field.

Interpretation and reporting

- Categorisations to simplify reporting must be made transparent. This also applies if categorisations made by third parties are used, for example from metadata providers.
- It is highly recommended to document every decision taken in a monitoring exercise and communicate them as openly and transparently as possible. This will also facilitate comparisons with other monitoring exercises.
- Reporting must account for lack of coverage of the data sources used and not claim validity for publication types or disciplines that are not covered or are considerably less covered.
- Time frames for analysis and reporting have to be chosen carefully. Possible time lags between the time of publication and the availability of all required metadata should be taken into account.
- The closer a monitoring exercise gets to the level of individual researchers, the more important it becomes to assure the quality of the monitoring results. On the aggregate level of nations or institutes, missing data or wrong classification of some publications will likely have little impact on the overall results. When the number of publications becomes small, manual checking of at least a sample of publications is recommended.

5. Sustainable and Comparable Monitoring

The landscape and infrastructure of scholarly publishing are in constant flux, with new providers and infrastructures entering and leaving the market quite frequently. Research organisations and their practices impact these research infrastructures and service providers, and processes established today will have lasting effects. When setting up monitoring exercises, sustainability and comparability should therefore be kept in mind. What happens if a provider offering relevant monitoring services terminates that service or fundamentally changes its offers? How well can a chosen approach include new publishing channels, such as Open Access publishing platforms? To ensure long-term sustainability and comparability of obtained monitoring results, it is recommended to reflect on the dependencies created, for example, the decision for a specific provider. Changing providers

or substantially altering conditions might be costly. To mitigate such risks while in parallel practising openness, it is strongly recommended to require transparent processes and detailed documentation.

Furthermore, the infrastructure and services used should follow open standards and ideally be community-governed and not for profit. It is recommended that RFOs and RPOs contribute to the development of such open services by stipulating in contracts with publishers that bibliographic metadata must be made openly available. When monitoring Open Access, openness of the processes and parties involved should obviously be a priority.

6. Conclusion and General Recommendation

Evidence on the open availability of scholarly publications has become an important driver to put the core ideas of Open Access into practice. Open Access monitoring enables deeper insight into publishing trends, can inform future strategies at institutional and national levels, provides guidance for policy development and review, helps to assess the effects of funding mechanisms and is crucial to negotiate transformative agreements with traditional subscription publishers. Furthermore, it contributes to a factual basis for decisions in support of new and improved fully Open Access publishing venues and platforms. Producing such evidence, tailored to the specific needs at hand, always involves decisions on Why, What, and How to monitor.

Although purpose, timeframe and other parameters might differ between organisations interested in monitoring Open Access, an aligned implementation and collaboration can improve the impact these efforts will have. This would not only save resources, but also contribute to sustainable practices overall, and facilitate the inevitable comparisons between results of Open Access monitoring exercises.

The practices presented in the here above use-cases have proven to be effective. However scholarly publishing and communication is an area of rapid development. Accordingly, a final and important recommendation must be to remain attentive to changes in the publishing landscape. New publishing channels and ever evolving practices of research communication will have implications for any future monitoring exercise.

© Copyright 2021

Science Europe AISBL, Rue de la Science 14, 1040 Brussels, Belgium

Science Europe is the association representing major public organisations that fund or perform excellent, ground-breaking research in Europe.

We bring together the expertise of some of the largest and best-known research organisations in the world to jointly push the frontiers of how scientific research is produced and delivers benefits to society.

We advocate science and the scientific community to help build the European Research Area and shape the global scientific agenda.

More information on our mission and activities is provided at www.scienceeurope.org

