# Principles of Open Science Monitoring

#### Draft submitted to UNESCO - March 28th 2024.

This document has been produced by the Open Science Monitoring Initiative. It has been prepared by a group of open science experts from August 2023. It has been discussed online and at UNESCO's workshop in Paris on December 19th 2023. It has been open to comments and suggestions from December to March 2024. It has been submitted to UNESCO in April 2024.

More and more open science policies are being developed in countries, universities, research-performing organisations and research-funding organisations around the world<sup>1</sup>. To estimate the advance of open science in different regions, numerous open science monitoring initiatives are emerging, which have the collective characteristic of being heterogeneous, non-comparable and not easily transposable to other geographical areas. These Principles of Open Science Monitoring aim to establish a set of best practices and guidelines that encourage pooling, comparison and reuse of the results of open science monitoring.

The goal of the initiative is to provide help for stakeholders like national governments, research-performing organisations and international organisations engaged in scientific research to set up their monitoring tools, embracing the diversity of monitoring approaches that are needed to account for the diversity of research and policies. The Principles of Open Science Monitoring initiative also seeks to establish standards that various stakeholders, including international organisations, national governments, research institutions, funding agencies and providers of data or related services can widely adopt. This initiative originates from organisations that monitor open science practices within their own domains. However, the principles outlined here are not tied to any specific monitoring service or particular software solution.

This document defines very broad, high-level principles. While the global aim is to monitor a comprehensive transformation to open science and its impacts on the research ecosystem and society, this document proposes a pragmatic approach. The scope of this document extends to all kinds of research outcomes and their impact on the scholarly ecosystem. In the future, it could expand to monitoring open science policies in their full breadth. This document is intended to be supplemented by technical specifications for each selected family of indicators such as open access to publications and academic publishing, sharing of clinical trial results, opening up or sharing of research data, publication of software used in research under open source licences, estimation of the costs of open science, estimation of the academic impact of open science, estimation of the societal impacts of open science, study of values and beliefs in the research community, study of the evolution of citizen science, study of open citations, study of the impact of open science approaches on daily research processes, etc.

We recommend that open science monitoring relies on open, transparent and reproducible methodologies that follow shared and agreed-upon principles.

Table of content	
Part 1: Relevance	2
Part 2: Transparency and reproducibility	2
Part 3: Self-assessment and responsible use	3
Annexes	4
Contributors	4
About the Open Science Monitoring Initiative	4
How to share a comment or contact the Open Science Monitoring Initiative	4
How to cite the principles	4
Bibliography	5

<sup>&</sup>lt;sup>1</sup> We adopt the <u>official definition of open science</u> from Unesco: "open science is defined as an inclusive construct that combines various movements and practices aiming to make multilingual scientific knowledge openly available, accessible and reusable for everyone, to increase scientific collaborations and sharing of information for the benefits of science and society, and to open the processes of scientific knowledge creation, evaluation and communication to societal actors beyond the traditional scientific community. It comprises all scientific disciplines and aspects of scholarly practices, including basic and applied sciences, natural and social sciences and the humanities, and it builds on the following key pillars: open scientific knowledge, open science infrastructures, science communication, open engagement of societal actors and open dialogue with other knowledge systems".

# **Part 1: Relevance**

- Applicable and clear in scope: indicators must be meaningful, applicable and relevant for the specific monitoring task at hand, and their scope should be clearly defined. Any limitations or constraints on the applicability of indicators should be explicitly stated and clearly communicated.
- **Meaningful for public policy**: indicators should be available for various stakeholders to design their own open science policies, fostering evidence-informed decisions and actions, adapted to their specific context. Indicators are intended to inform public policy and should not be used for assessing individual researchers, such as in grant applications, recruitment, or staff promotion.
- **Co-created**: as much as possible, the adoption of indicators should be based on active, voluntary participation from relevant stakeholders. Indicators should be co-created with research policymakers, research-performing organisations, research-funding organisations, funding agencies and the research community through public consultation and dialogue. Stakeholders should have the agency to opt out of adopting indicators if they choose to do so.
- Inclusive: indicators should take into account the diversity of academic domains, social contexts and languages.
- **Mature**: the level of maturity of each indicator needs to be made explicit. Experimental approaches should be transparently labelled as such. Meta-research should be involved to estimate the maturity of new indicators both before and after making them public.
- Indicators toolboxes for different contexts: to reduce complexity and minimise implementation efforts for organisations setting up open science monitors, the goal should be to identify toolboxes of basic indicators suitable for various contexts. This approach allows stakeholders to select indicators from these sets based on their specific needs and requirements.
- **Comparable**: indicators should aim to facilitate comparability between institutions, countries, regions, and research fields. Striving towards comparability while acknowledging the limitations to the applicability of any indicators in practice is a large part of the work of monitoring.
- Up-to-date indicators: Provide indicators that are regularly updated.

## **Part 2: Transparency and reproducibility**

- **Communication for the general public**: a clear communication of the conclusions drawn from the indicators, generated data and visualisations, is needed for the interested parties.
- **Public documentation of processes and methodology**: a publicly available documentation should describe data collection, data provenance, processing and implementation choices made by the monitoring project.
- **Explicit data provenance**: the whole data pipeline should be transparent and open. It should thus rigorously document the origin and the licence of each data point to provide clear provenance information. Partially undocumented information or unclear licences would prevent reuse of the whole data set.
- **Open output data**: output data should be open. New knowledge and data thus produced should be distributed under an open licence. In the particular case of indicators coming from text and data mining applied on texts protected by copyright, the output should be open, in compliance with copyright rules and privacy protection.
- Inclusion of FAIR and CARE principles: output data should be as much as possible compliant with the FAIR principles (Findable, Accessible, Interoperable, Reusable) as well as the CARE principles (Collective Benefit, Authority to Control, Responsibility and Ethics).
- **Sampling**: when exhaustive data cannot be obtained, where this can be reasonably justified, indicators can employ randomised sampling and other well-established statistical techniques, interpolation and transparent heuristics, as long as these techniques are properly documented.
- **Open-source software**: software used for data acquisition and processing should be open source, versioned and published with adequate documentation on a platform that facilitates collaboration and contribution, to foster open collaboration and reuse over the long term.
- **Standardised indicators**: core indicators are standardised using a public and collective definition procedure. Each indicator is publicly documented into a specification document.
- **Transparency in the quality of indicators**: accuracy (e.g. precision and recall<sup>2</sup>), coverage and timeliness should be evaluated and ideally made public for each indicator.
- **Reusable by design**: monitoring results and indicator outcomes should be accessible through APIs and bulk download features, updated on a regular basis.

<sup>&</sup>lt;sup>2</sup> "**Precision** and **recall** are performance metrics that apply to data retrieved from a collection, corpus or sample space. **Precision** (also called positive predictive value) is the fraction of relevant instances among the retrieved instances. **Recall** (also known as sensitivity) is the fraction of relevant instances that were retrieved" (<u>source</u>).

## Part 3: Self-assessment and responsible use

- **Self-assessment**: initiatives that implement these principles are invited to assess their monitoring work on the basis of this document, and make the assessment public.
- Monitor for improvement: monitoring should be used to incentivise rather than punish lack of practices.
- **Continuous assessment of the monitoring initiative**: as any indicator could potentially facilitate gaming attempts, there is a need for continuous assessment of indicators and methodologies, and the potential shift of practices in unintended directions.
- Adaptable: specifications may evolve over time. In such cases, changes and developments should be transparent and publicly documented.
- Avoid rankings: these indicators must not be used to create rankings, as mentioned in the CoARA Agreement.

These principles will be completed by upcoming technical specifications dedicated to their implementation

## Annexes

#### **Contributors**

- Evgeny Bobrov (Berlin Institute of Health at Charité, QUEST Center for Responsible Research, <u>https://orcid.org/0000-0002-2368-464X</u>)
- Laetitia Bracco (Université de Lorraine, <u>https://orcid.org/0000-0002-2939-9110</u>)
- Marin Dacos (French Ministry of Higher Education and Research, <u>https://orcid.org/0000-0002-9361-5295</u>)
- Roberto Di Cosmo (INRIA, https://orcid.org/0000-0002-7493-5349)
- Christopher Erdmann (SciLifeLab, <u>https://orcid.org/0000-0003-2554-180X</u>)
- Nicolas Fressengeas (French Ministry of Higher Education and Research, <u>https://orcid.org/0000-0002-5534-712X</u>)
- Alice Howarth (University of Liverpool, https://orcid.org/0000-0003-2288-1745)
- Iain Hrynaszkiewicz (PLOS, <u>https://orcid.org/0000-0002-9673-5559</u>)
- Anastasiia Iarkaeva (Berlin Institute of Health at Charité, QUEST Center for Responsible Research, https://orcid.org/0000-0002-7370-1663)
- Malika Ihle (LMU Munich, LMU Open Science Center, <u>https://orcid.org/0000-0002-3242-5981</u>)
- Neil Jacobs (UKRN Open Research Programme, https://orcid.org/0000-0002-8050-8175)
- Nees Jan van Eck (Leiden University, <u>https://orcid.org/0000-0001-8448-4521</u>)
- Eric Jeangirard (French Ministry of Higher Education and Research, <u>https://orcid.org/0000-0002-3767-7125</u>)
- Petr Knoth (Open University, <u>https://orcid.org/0000-0003-1161-7359</u>)
- Bianca Kramer (Sesame Open Science, <u>https://orcid.org/0000-0002-5965-6560</u>)
- Ignasi Labastida I Juan (Universitat de Barcelona, https://orcid.org/0000-0001-7030-7030)
- Larissa Leiminger (Max Planck Digital Library, <u>https://orcid.org/0000-0002-6491-3197</u>)
- Patrice Lopez (science-miner, <u>https://orcid.org/0000-0002-9959-9441</u>)
- Paolo Manghi (OpenAIRE, <u>https://orcid.org/0000-0001-7291-3210</u>)
- Lucy Montgomery (Curtin University, https://orcid.org/0000-0001-6551-8140)
- Florian Naudet (University of Rennes, <u>https://orcid.org/0000-0003-3760-3801</u>)
- Cameron Neylon (Curtin University, <u>https://orcid.org/0000-0002-0068-716X</u>)
- Max Paulus (Vrije Universiteit Amsterdam, https://orcid.org/0000-0001-9651-3086)
- Esther Plomp (Delft University of Technology, Faculty of Applied Sciences, https://orcid.org/0000-0003-3625-1357)
- Vanessa Proudman (SPARC Europe, https://orcid.org/0000-0001-5623-9051)
- Karthik Ram (Berkeley Institute for Data Science. University of California, Berkeley, https://orcid.org/0000-0002-0233-1757)
- Laura Rovelli (Latin American Forum for Research Assessment (FOLEC) and the Latin American Council of Social Science (CLACSO), <u>https://orcid.org/0000-0001-7059-149X</u>)
- Inge Stegeman (University Medical Center Utrecht, <u>https://orcid.org/0000-0001-5154-7178</u>)
- Didier Torny (CNRS, <u>https://orcid.org/0000-0001-6661-9680</u>)
- Giannis Tsakonas (University of Patras, <u>https://orcid.org/0000-0002-8786-9440</u>)
- Athanasia Salamoura (HEAL-Link, https://orcid.org/0000-0002-5173-8588)
- Ludo Waltman (Leiden University, https://orcid.org/0000-0001-8249-1752)

## About the Open Science Monitoring Initiative

The **Open Science Monitoring Initiative** brings together institutions and individuals involved in monitoring open science. OSMI aims to encourage the adoption of open science monitoring principles and to promote their practical implementation.

### How to share a comment or contact the Open Science Monitoring Initiative

You can contact the Open Science Monitoring Initiative to share comments or for any other purpose by writing to <u>osmf@groupes.renater.fr</u>

### How to cite the principles

**How to cite :** Open Science Monitoring Initiative, *Principles of Open Science Monitoring*, Paris, 2024. DOI : <u>10.52949/4</u> **Licence :** This work is published under a CC BY Licence.

#### DOI: 10.52949/49

## **Bibliography**

- Bassinet A., Bracco L., L'Hôte A., Jeangirard E., Lopez P., Romary L. (2023). "Large-scale Machine-Learning analysis of scientific PDF for monitoring the production and the openness of research data and software in France". hal-04121339v3
- Bilder G, Lin J, Neylon C. (2020). "The Principles of Open Scholarly Infrastructure". https://doi.org/10.24343/C34W2H
- Bracco A., L'Hôte A., Jeangirard E., Torny D. (2022). "Extending the open monitoring of open science: A new framework for the French Open Science Monitor (BSO)". hal-03651518
- Carroll S., Garba I., Figueroa-Rodríguez O., Holbrook J., *et al.* (2020). "The CARE Principles for Indigenous Data Governance". *Data Science Journal*, 19: XX, pp. 1–12. https://doi.org/10.5334/dsj-2020-042
- Coalition for Advancing Research Assessment (2023). "Agreement on Reforming Research Assessment". <u>https://coara.eu/agreement/the-agreement-full-text/</u>
- Cobey KD., Haustein S., Brehaut J., Dirnagl U., Franzen DL., Hemkens LG., *et al.* (2023). "Community consensus on core open science practices to monitor in biomedicine". *PLoS Biol* 21(1): e3001949. https://doi.org/10.1371/journal.pbio.3001949
- Diprose, J. P., Hosking, R., Rigoni, R., Roelofs, A., Chien, T., Napier, K., Wilson, K., Huang, C., Handcock, R. N., Montgomery, L. & Neylon, C. (2023). "A User-Friendly Dashboard for Tracking Global Open Access Performance". *The Journal of Electronic Publishing* 26(1). https://doi.org/10.3998/jep.3398
- European Commission, Directorate-General for Research and Innovation (2021). "Towards a reform of the research assessment system Scoping report". *Publications Office*. <u>https://data.europa.eu/doi/10.2777/707440</u>
- Hicks, D., Wouters, P., Waltman, L. *et al.* (2015). "Bibliometrics: The Leiden Manifesto for research metrics". *Nature* 520, 429–431. https://doi.org/10.1038/520429a
- Hrynaszkiewicz I., Kiermer V. (2022). "PLOS Open Science Indicators principles and definitions". figshare. https://doi.org/10.6084/m9.figshare.21640889.v1
- Hrynaszkiewicz I. "How Do We Measure Success for Open Science?". (2022). https://scholarlykitchen.sspnet.org/2022/12/13/guest-post-how-do-we-measure-success-for-open-science/
- Jeangirard E. (2019). "Monitoring Open Access at a national level: French case study". *ELPUB 2019 23rd* edition of the International Conference on Electronic Publishing, Marseille, France. 10.4000/proceedings.elpub.2019.20
- Moher D., Bouter L., Kleinert S., Glasziou P., Sham MH., Barbour V., *et al.* (2020). "The Hong Kong Principles for assessing researchers: Fostering research integrity". *PLoS Biol* 18(7): e3000737. https://doi.org/10.1371/journal.pbio.3000737
- PathOS. "Open science indicator handbook". https://handbook.pathos-project.eu/
- Rafols I., Molas-Gallart J., Meijer I. (2024). "Monitoring Open Science as Transformative Change: Towards a Systemic Framework". https://doi.org/10.31235/osf.io/knhzt
- DORA (2013). "San Francisco Declaration on Research Assessment". <u>https://sfdora.org/</u>
- https://op.europa.eu/en/publication-detail/-/publication/36ebb96c-50c5-11ec-91ac-01aa75ed71a1/language-en
- UK Committee on Research Integrity. (2023). "Research Integrity in the UK: Annual statement 2023". https://doi.org/10.5281/zenodo.8117154
- Wilkinson M., Dumontier M., Aalbersberg I. *et al.* (2016). "The FAIR Guiding Principles for scientific data management and stewardship". *Sci Data* 3, 160018. https://doi.org/10.1038/sdata.2016.18