OPEN SCIENCE MONITOR

UPDATED METHODOLOGICAL NOTE

Brussels, 4th April 2019

Consortium partners:

The Lisbon Council

ESADE Business School

Centre for Science and Technology Studies (CWTS) at Leiden University

Subcontractor:

Elsevier

1 Introduction

This is the revised version of the methodology of the Open Science Monitor, based on the comments received online, the discussion in the experts' workshop.

Open science has recently emerged as a powerful trend in research policy. To be clear, openness has always been a core value of science, but it meant publishing the results or research in a journal article. Today, there is consensus that, by ensuring the widest possible access and reuse to publications, data, code and other intermediate outputs, scientific productivity grows, scientific misconduct becomes rarer, discoveries are accelerated. Yet it is also clear that progress towards open science is slow, because it has to fit in a system that provides appropriate incentives to all parties. Of course, dr. Rossi can advance his research faster by having access to Dr. Svensson's data, but what is the rationale for Dr Svensson to share her data if no one includes data citation metrics in the career assessment criteria?

The European Commission has recognised this challenge and moved forward with strong initiatives from the initial 2012 recommendation on scientific information (C (2012) 4890), such as the Open Science Policy Platform and the European Open Science Cloud. Open access and open data are now the default option for grantees of H2020.

The Open Science Monitor (OSM) aims to provide data and insight needed to support the implementation of these policies. It gathers the best available evidence on the evolution of Open Science, its drivers and impacts, drawing on multiple indicators as well as on a rich set of case studies. [1]

This monitoring exercise is challenging. Open science is a fast evolving, multidimensional phenomenon. According to the OECD (2015), "open science encompasses unhindered access to scientific articles, access to data from public research, and collaborative research enabled by ICT tools and incentives". This very definition confirms the relative fuzziness of the concept and the need for a clear definition of the "trends" that compose open science.

Precisely because of the fast evolution and novelty of these trends, in many cases it is not possible to find consolidated, widely recognized indicators. For more established trends, such as open access to publications, robust indicators are available through bibliometric analysis. For most others, such as open code and open hardware, there are no standardised metrics or data gathering techniques and there is the need to identify the best available indicator that allows one to capture the evolution and show the importance of the trend.

The present document illustrates the methodology behind the selected indicators for each trend. The purpose of the document is to ensure transparency and to gather feedback in order to improve the selected indicators, the data sources and overall analysis.

The initial launch of the OSM contains a limited number of indicators, mainly updating the existing indicators from the previous Monitor (2017). New trends and new indicators will be added in the course of the OSM project, also based on the feedback to the present document.

[1] The OSM has been published in 2017 as a pilot and re-launched by the European Commission in 2018 through a contract with a consortium composed by the Lisbon Council, ESADE Business School and CWTS of Leiden University (plus Elsevier as subcontractor). See https://ec.europa.eu/research/openscience/index.cfm?pg=home&section=monitor

1.1 Objectives

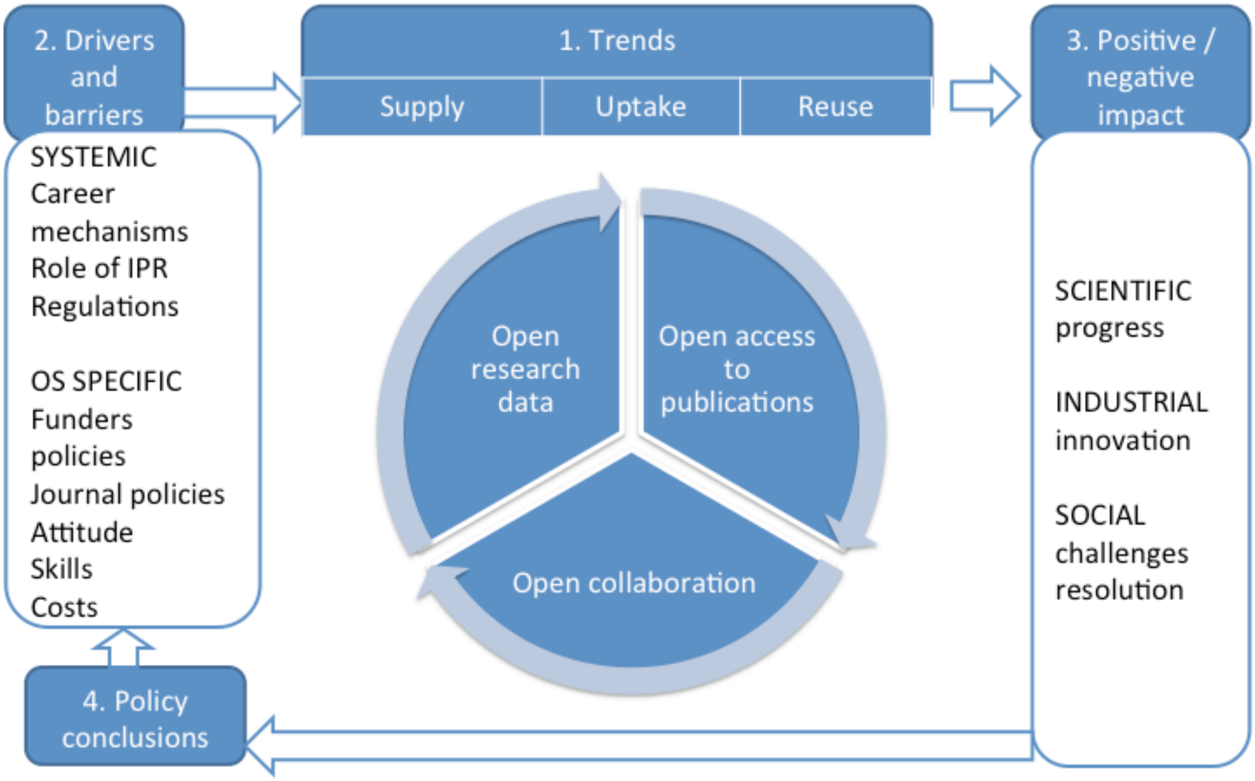
The OSM covers four tasks:

1. To provide metrics on the open science trends and their development.
2. To assess the drivers (and barriers) to open science adoption.
3. To identify the impacts (both positive and negative) of open science
4. To support evidence-based policy actions.

The indicators presented here focus mainly on the first two tasks: mapping the trends, and understanding the drivers (and barriers) for open science implementation.

The chart below provides an overview of the underlying conceptual model.

Figure 1: A conceptual model: an intervention logic approach



The central aspect of the model refers to the analysis of the open science trends and is articulated alongside three dimensions: *supply*, *uptake* and *reuse* of scientific outputs.

In the OSM framework, *supply* refers to the emergence of services such as data repositories. The number of data repositories (one of the existing indicators) is a *supply* indicator of the development of Open Science. On the demand side, indicators include, for example, the amount of data stored in the repositories, the percentage of scientists sharing data. Finally, because of the nature of Open Science, the analysis will go beyond usage, since the reuse dimension is particularly important. In this case, relevant indicators include the number of scientists reusing data published by other scientists, or the number of papers using these data.

On the left side of the chart, the model identifies the key factors influencing the trends, both positively and negatively (i.e. *drivers* and *barriers*). Both drivers and barriers are particularly relevant for policy-makers as this is the area where an action can make greatest difference, and are therefore strongly related to policy recommendations. These include "policy drivers", such as funders' mandates. It is important to assess not only policy drivers dedicated to open science, but also more general policy drivers that could have an impact on the uptake of open science. For instance, the increasing reliance on performance-based funding or the emphasis on market exploitation of research are general policy drivers that could actually slow down the uptake of open science.

The right side of the chart in the model, illustrates the *impacts* of open science to research or the scientific process itself; to industry or the capacity to translate research into marketable products and services; to society or the capacity to address societal challenges.

1.2 Scope

By definition, open science concerns the **entire cycle** of the scientific process, not only open access to publications. Hence the macro-trends covered by the study include: open access to publications, open research data and open collaboration. While the first two are self-explanatory, open scientific collaboration is an umbrella concept to include forms of collaboration in the course of the scientific process that do not fit under open data and open publications.

Table 1: Articulation of the trends to be monitored

Categories	Trends
Open access to publications	• Open access policies (funders and journals), • Green and gold open access adoption (bibliometrics).[2]
Open research data	• Open data policies (funders and journals) • Open data repositories • Open data adoption and researchers' attitudes.
Open collaboration	• Open code, • Altmetrics, • Open hardware, • Citizen science.

New trends within the open science framework will be identified through interaction with the stakeholder's community by monitoring discussion groups, associations (such as Research Data Alliance- RDA), mailing lists, and conferences such as those organised by Force11 (www.force11.org).

[2] According to the EC, "'Gold open access' means that open access is provided immediately via the publisher when an article is published, i.e. where it is published in open access journals or in 'hybrid' journals combining subscription access and open access to individual articles. In gold open access, the payment of publication costs ('article processing charges') is shifted from readers' subscriptions to (generally one-off) payments by the author.[...] 'Green. open access' means that the published article or the final peer-reviewed manuscript is archived by the researcher (or a representative) in an online repository." (Source: H2020 Model Grant Agreement)

The study covers **all research disciplines**, and aims to identify the differences in open science adoption and dynamics between diverse disciplines. Current evidence shows diversity in open science practices in different research fields, particularly in data-intensive research domains (e.g. life sciences) compared to others (e.g. humanities).

The **geographic coverage** of the study is 28 Member States (MS) and G8 countries, including the main international partners, with different degrees of granularity for the different variables. As far as possible, data has to be presented at **country level**.

Finally, the analysis focuses on the factors at play for **different stakeholders** as mapped in the chart below (table 2). For each stakeholder's category, OSM will deliberately consider both traditional (e.g. Thomson Reuters) and new players in research (e.g. F1000).

Table 2: Stakeholders types

Researchers	Professional and citizens researchers
Research institutions	Universities, other publicly funded research institutions, and informal groups
Publishers	Traditional publishers New OA online players
Service providers	Bibliometrics and new players
Policy makers	At supranational, national and local level
Research funders	Private and public funding agencies.

2 Indicators and data sources

Because of the fast and multidimensional nature of open science, a wide variety of indicators have been used, depending on data availability:

- Bibliometrics: this is the case for open access to publications indicators, and partially for open data and altmetrics.
- Online repositories: there are many repositories dedicated to providing a wide coverage of the trends, such as policies by funders and journals, APIs and open hardware.
- Surveys: surveys of researchers shed light on usage and drivers. Preference is given to multi-year surveys.
- Ad hoc analysis in scientific articles or reports: for instance, reviews of journals policies with regard to open data and open code
- Data from specific services: open science services often offer data on their uptake, as for Sci-starter or Mendeley. In this case, data offer limited representativeness about the trend in general, but can still be useful to detect differences (e.g. by country or discipline). Where possible, in this case, we present data from multiple services.

2.1 Open access to publications

This trend has received lots of attention by people commenting, mainly because of the exclusive reliance on the Scopus database. The consortium has not received evidence to dispute that Scopus data allow for the necessary data quality, especially since the "open access" tagging is exclusively performed by the consortium partners. But in addition, to improve the robustness, we updated the methodology by adding Unpaywall data to provide the best possible coverage, and by adding dedicated analysis that will perform controls of the effects on data of using alternative databases such as Web of Science. More details are provided in the updated Annex 1. Additionally, data from Scopus can be made available to individual academic researchers to assess or replicate the OSM methodology, under the standing policy of Elsevier to permit academic research access to Scopus data.

Beside the long list of indicators below, the detailed methodology for calculating the percentage of OA publications is presented in the annex 1.

Indicator	**Source**
Number of Funders with open access policies (with caveat that it is skewed towards western countries)	Sherpa Juliet[3]
Number of Journals with open access policies (with caveat that it is skewed towards western countries)	Sherpa Romeo[4]
Number of publishers/journals that have adopted the TOP Guidelines (including the level of adoption actual implementation where possible)	Cos.io
P - # Scopus publications that enter in the analysis	Scopus, Unpaywall
P(oa) - # Scopus publications that are Open Access (CWTS method for OA identification)	Scopus, Unpaywall
P(green oa) - # Scopus publications that are Green OA	Scopus, Unpaywall
P(gold oa) - # Scopus publications that are Gold OA	Scopus, Unpaywall
PP(oa) - Percentage OA publications of total publications	Scopus, Unpaywall
PP(green oa) - Percentage gold OA publications of total publications	Scopus, Unpaywall
PP(gold oa) - Percentage green OA publications of total publications	Scopus, Unpaywall

[3] http://v2.sherpa.ac.uk/juliet/

[4] http://www.sherpa.ac.uk/romeo/index.php?la=en&fIDnum=|&mode=simple

2.2 Open research data

Several comments received were useful to identify new data sources to measure open data publication, and have been added.

There were several criticisms of using Elsevier to gather data through the survey, but no valid alternatives of comparable quality and cost/efficiency were proposed. Moreover, data from Elsevier survey will be openly released, as last year. The detailed methodology for developing the Elsevier survey is presented in the annex 4.

Several comments pointed to the need for measuring new additional aspects, such as "number of papers based on openly available raw data". However, no concrete proposals were made about sources. We will follow up with those commenting to obtain further detail.

Indicator	**Source**
Number of Funders with policies on data sharing (with caveat that it is skewed towards western countries)	Sherpa Juliet
Number of Journals with policies on data sharing	Vasilevsky et al, 2017[5]
Number of open data repositories	Re3data
% of paper published with data	Bibliometrics: Datacite
Citations of data journals	Bibliometrics: Datacite
Attitude of researchers on data sharing.	S2016 and 2018 survey by Elsevier, follow-up of the 2017 report.[6]
Sharing of research data: % of researchers that have directly shared research data from their last project, by recipient.	S2016 and 2018 survey by Elsevier.
Benefits of sharing research data: % of researchers per benefit.	S2016 and 2018 survey by Elsevier.
Consequences of sharing data: Contact made with researchers outside their research team after sharing data, % of researchers by type of organisation.	2018 survey by Elsevier.
Making research data available: Effort required to make research data reusable by others, % of researchers per amount of effort.	S2016 and 2018 survey by Elsevier.
Management of research data: % of researchers that take steps to manage their research data and/or archive it for potential reuse by themselves or others.	S2016 and 2018 survey by Elsevier.

[5] Vasilevsky, Nicole A., Jessica Minnier, Melissa A. Haendel, and Robin E. Champieux. "Reproducible and Reusable Research: Are Journal Data Sharing Policies Meeting the Mark?" PeerJ 5 (April 25, 2017): e3208. doi:10.7717/peerj.3208.

[6] Berghmans, Stephane, Helena Cousijn, Gemma Deakin, Ingeborg Meijer, Adrian Mulligan, Andrew Plume, Sarah de Rijcke, et al. "Open Data : The Researcher Perspective," 2017, 48 p. doi:10.17632/bwrnfb4bvh.1.

Attitudes of researchers: % of researchers that agree with statement.	S2016 and 2018 survey by Elsevier.
Number and/or total size of CC-0 datasets.	Base-search.net
Number of OAI-compliant repositories.	Base-search.net
Number of repositories with an open data (https://opendefinition.org/) policy for metadata.	OpenDOAR, "commercial" in metadata reuse policy. https://opendefinition.org/

2.3 Open collaboration

Indicator	**Source**
Membership of social networks on science (Mendeley, ResearchGate, f1000)	Scientific social networks

2.3.1 Open code

Several comments addressed this issue, mainly by suggesting new data sources to be used. They are tentatively included here for discussion. Several suggestions did not include sources and are not listed here for the time being, pending additional analysis.

Indicator	**Source**
Number of code projects with DOI	Mozilla Codemeta
Number of scientific API	Programmableweb
% of journals with open code policy	Stodden 2013[7]
Software citations in DataCite	Datacite
Number of code projects in Zenodo	Zenodo
Add: number of software deposits under an OSI-approved license.	Base
Number of Software papers in Software Journals	(e.g. JORS https://openresearchsoftware.metajnl.com/ and others)
N. of users in reproducibility platforms such as CodeOcean	CodeOcean

[7] Stodden, V., Guo, P. and Ma, Z. (2013), "Toward reproducible computational research: an empirical analysis of data and code policy adoption", PLoS One, Vol. 8 No. 6, p. e67111. doi: 10.1371/ journal.pone.0067111.

2.3.2 Open scientific hardware

The few comments received here pointed to the limited importance of open hardware licenses, because of the fragmentation across the EU. That indicator has then been removed.

Indicator	Source
Number of projects on open hardware repository	Open Hardware repository[8]

2.3.3 Citizen science

The very few comments received did not include additional sources and are therefore not included for the time being.

Indicator	Source
N. Projects in Zooniverse and Scistarter	Zooniverse and Scistarter
N. Participants in Zooniverse and Scistarter	Zooniverse and Scistarter

2.3.4 Altmetrics

The feedback received in this case was highly critical of the dependence on Plum Analytics and Mendeley. Based on the feedback received, the consortium will keep the indicators as such, but perform additional checks and analysis using alternatives to Plum Analytics, such as Altmetric.com, as suggested by the comments. For what concerns Mendeley, it is the only source currently available providing open data about readership and will therefore continue to be used. The indicators will be reassessed once the data become available.

Indicator	Source
P(tracked) - # Scopus publications that can be tracked by the different sources (e.g. typically only publications with a DOI, PMID, Scopus id, etc. can be tracked).	Scopus & Plum Analytics
P(mendeley) - # Scopus publications with readership activity in Mendeley	Scopus, Mendeley & Plum Analytics
PP(mendeley) - Proportion of publications covered on Mendeley. P(mendeley)/P(tracked)	Scopus, Mendeley & Plum Analytics
TRS - Total Readership Score of Scopus publications. Sum of all Mendeley readership received by all P(tracked)	Scopus, Mendeley & Plum Analytics
TRS(academics) - Total Readership Score of Scopus publications from Mendeley academic users (PhdS, Professors, Postdocs, researchers, etc.)	Scopus, Mendeley & Plum Analytics
TRS(students) - Total Readership Score of Scopus publications from Mendeley student users (Master and	Scopus, Mendeley & Plum Analytics

[8] https://www.ohwr.org

Bachelor students)	
TRS(professionals) - Total Readership Score of Scopus publications from Mendeley professional users (librarians, other professionals, etc.)	Scopus, Mendeley & Plum Analytics
MRS - Mean Readerships Score. TRS/P(tracked)	Scopus & Plum Analytics
MRS(academics) - TRS(academics)/P(tracked)	Scopus & Plum Analytics
MRS(students) - TRS(students)/P(tracked)	Scopus & Plum Analytics
MRS(professionals) - TRS(professionals)/P(tracked)	Scopus & Plum Analytics
P(twitter) - # Scopus publications that have been mentioned in at least one (re)tweet	Scopus & Plum Analytics
PP(twitter) - Proportion of publications mentioned on Twitter. P(twitter)/P(tracked)	Scopus & Plum Analytics
TTWS - Total Twitter Score. Sum of all tweets mentions received by all P(tracked)	Scopus & Plum Analytics
MTWS - Mean Twitter Score. TTWS/P(tracked)	Scopus & Plum Analytics

3 Next steps

The consortium will deliver new set of case studies by the end of July 2019.

The consortium will continue revising the methodology with the community, through an open Linkedin group.

Annex 1: Methodological note on the implementation of *Unpaywall* data into the Open Access labelling of the Open Science Monitor

Thed van Leeuwen & Rodrigo Costas

Centre for Science and Technology Studies (CWTS), Leiden University, the Netherlands

Introduction

In this document the methodological approach for the identification and creation of the Open Access (OA) labels for the *Open Science Monitor* is presented.

CWTS has been working on an OA evidence over the last three years, and which has been reported at the Paris 2017 STI Conference (van Leeuwen et al, 2017). In this method we strived for a high degree of reproducibility of our results based upon data carrying OA labels following from the methodology we developed, based upon freely and open data sources (DOAJ, ROAD, CrossRef, PMCentral and OpenAIRE). This was applied on Elsevier's Scopus publication data.

From mid of 2018 onwards, another source for OA tagging became prominent, namely the *Unpaywall* database (https://unpaywall.org/). CWTS is (still) working on integrating this data source into the current analysis, which means a combination of basic research on this implementation, in which we compared the previously developed methodology with the new one, e.g. comparing our methodology and the numbers of publication labelled with OA tags, with the *Unpaywall* data (see also Martín-Martín et al, 2018). The inclusion of *Unpaywall* in the methodology requires us to conduct research to better understand what OA evidence data *Unpaywall* provides, whether all types of OA evidence align with our criteria of building OA evidence, and whether there is any potential conceptual issues related to some typologies of OA provided by *Unpaywall* (for example, we wonder whether the 'Bronze' OA typology disclosed by *Unpaywall* can be considered as a sustainable form of OA, cf. Martín-Martín et al, 2018).

The methodological approach that we propose mainly focuses on adding different OA labels to the publications covered in the Scopus database, using *Unpaywall* to establish this OA status of scientific publications. It is important to highlight that two basic principles for this OA label are ***sustainability*** and ***legality***. By sustainability we mean that it should, in principle, be possible to reproduce the OA labelling from the various sources used, repeatedly, in an open fashion, with a relatively limited risk of the sources used disappearing behind a pay-wall, and particularly that the reported publications as OA will change their status to closed. The second aspect (legality) relates to the usage of data sources that represent legal OA evidence for publications, excluding rogue or illegal OA publications (i.e. we do not consider OA publications made freely available in platforms such as ResearchGate or Sci-hub legal, and probably also not sustainable). While the former criterion is mainly oriented to a scientific requirement, namely that of reproducibility and perdurability over time, the latter criteria is particularly important for science policy, indicating that OA publishing aligns with policies and mandates.

In the re-loading of the publication counts in the Open Science Monitor, *Unpaywall* data form the source for tagging Scopus publications with labels on Open Access availability. This is a change from the previous OA tagging method, when it comes to the variety of OA tags, and the procedure, while the criteria developed for the tagging of publications with OA labels remain intact.

In the implementation of *Unpaywall* data, we were expecting to be capable of distinguishing next to Gold and Green, which we used so far, also Hybrid OA as a further form of compliant OA publishing. However, we identified in the current *Unpaywall* category of 'Hybrid', some form of mixing up of Gold with true Hybrid (which means APC-based publishing in an otherwise tool access journal) occurs, which blurs the perspective on the Hybrid category. Consequently, also the category Gold will be affected, as the figures presented here will be likely to a lower estimate of the real situation. We are currently working on sorting this out, in order to create better defined categories of OA types, distinguishing Gold, Green and Hybrid.

A fourth OA category in *Unpaywall* is 'Bronze', which is basically a form of openly available publishing initiated by the publishers, in which the copyright status is not clear. As in our criteria this is not a sustainable form of OA we opt for not considering as a separate OA category, although it will be included in the overall consideration of OA publications.

In this data delivery, and hence the re-loading of the OSM website, we take four different analytical approaches: overall (all publications), countries, fields, and fields & countries combined. We do this in two ways, one for the full period, the other for the trend analysis from 2009-2017.

The following OA indicators are calculated:

- *Total number of publications*: this is the overall number of publications, which is used as the denominator for the calculation of shares of OA.
- *Total (and share of) OA*: the overall number (and share) of OA available publications (covering all types of OA recorded by *Unpaywall* - namely Gold, Green, Hybrid and Bronze). With this we intend to be fully in line with the *Unpaywall* data in the disclosure of OA availability.
- *Green OA* and *Gold OA*: in this data delivery we report publication counts (and shares) of Green and Gold OA publications separately. This means that we do not apply any preference approach (e.g. giving priority to Gold over Green when a publication can be labelled as both). The rationale behind this choice is based on the idea that different types of OA have different interests depending on the different stakeholders (e.g. readers, authors, academic institutions, funders, etc.), and at this stage we opt for leaving them in their most original form, so both types of OA can be fully informed.

Finally, regarding the underlying publication data, we have restricted the analysis to only those publications having a DOI in Scopus, since currently *Unpaywall* only provides OA labels to publications with DOIs. The inclusion of all publications with and without DOIs could lead to an underestimation of OA prevalence, since those publications without DOIs would increase the denominator, while they cannot be tracked for OA. Future developments will be also oriented towards providing OA evidence for those publications without DOIs, thus expecting to increase the OA analytical landscape. Furthermore, we have only worked with articles and reviews, and future developments will also consider the incorporation of other document types.

References

van Leeuwen TN, Meijer I, Yegros-Yegros, A & Costas R, Developing indicators on Open Access by combining evidence from diverse data sources, Proceedings of the 2017 STI Conference, 6-8 September, Paris, France (https://sti2017.paris/) (https://arxiv.org/abs/1802.02827)

Martín-Martín, A., Costas, R., van Leeuwen, T., & Delgado López-Cózar, E. (2018). Evidence of Open Access of scientific publications in Google Scholar: a large-scale analysis. SocArXiv papers. DOI: 10.17605/OSF.IO/K54UV

Olensky, M., Schmidt, M., & Van Eck, N.J. (2016). Evaluation of the Citation Matching Algorithms of CWTS and iFQ in Comparison to the Web of Science. *Journal of the Association for Information Science and Technology*, 67(10), 2550-2564. doi:10.1002/asi.23590.

Annex 2: Answer to comments

Below, the comments received online are group under headings, based on their content. At the end of each answer, the relevant comments ids are listed in parenthesis. The full comments with ids are available online.

Open access

Only open sources should be used, not proprietary data since open data sources already exist.

There are today no open data sources that offer the richness of metadata provided by proprietary sources. Crossref in particular lacks several fields that are crucial to the work of the Open Science Monitor. The full explanation of the differences and the necessity to use proprietary data is provided in the slides presented in the workshop.
(1431,1428,1289,1288,1280,1281,1305,1315,1340,1346,1379,1381,1479,1480,1487,1265,1272,1341,1309,1342,1426,1455,1268,1343,1290,1468,1472,1422,1424,1449)

Scopus is biased and has a conflict of interest because it's owned by Elsevier

It is the consortium developing the indicators, while Elsevier only provides underlying data for some indicators. In particular, it is CWTS that attributes the open access tag. Scopus has biases, as all other sources have, and they are known and treated transparently by the consortium, but it remains a fundamental and high-quality instrument for bibliometric analysis. The role of the consortium is precisely to develop robust indicators taking into account the limitations of the different sources.
(1344,1382,1456,1345)

Data are not accessible for replication because they are based on a proprietary database

Scopus can be made available to individual academic researchers to assess or replicate the OSM methodology, under the standing policy of Elsevier to permit academic research access to Scopus data. Requests outlining data requirements and research scope should be submitted through the project email (opensciencemonitor@lisboncouncil.net).
(1430,1397,1440)

Multiple sources should be used to ensure robustness

To address this comment, the consortium will carry out and publish an ad hoc additional analysis carrying out the same analysis based on Web of Science.
In addition, the consortium will use Unpaywall data alongside Scopus data in the database used to create the headline indicators.
(1266,1311,1396,1466,1484,1492,1267,1467,1269,1275,1325)

Sources have insufficient coverage (in terms of journals, disciplines, countries, monographs).

With regard to Scopus, to widen the scope and capture, the consortium has obtained access to Unpaywall data, which has a larger footprint and will be integrated in the analysis in addition to Scopus.

(1392,1429,1438,1439,1442,1454,1469,1483,1306,1432,1470,1481,1308,1471,1352,1444,1445,1459,1489,1490)
With regard to Sherpa, unfortunately this limitation is unavoidable. The information about the biased coverage will be included in the presentation of the indicators. (1420)

Indicator should not take into account impact factor and related issues.
The consortium agrees. The indicators related to "highly cited" journals have been removed. (1326,1347,1457,1493,1348,1349,1441,1286,1287,1312,1316,1328,1350,1401,1458,1473)

New indicators and sources
The consortium received many useful proposals, but only few of them immediately actionable. Most proposals need additional effort, and some are not deemed relevant. To enable this effort as well as additional collaboration on any indicator, the consortium will set up additional collaboration spaces, beyond the one-off consultation about the methodology. (1327,1298,1329,1515,1545,1546,1548,1549,1282,1283,1297,1330,1355,1402,1403,1406,1463,1474,1390,1465)
Some comments were out of scope, based on the tender requirements.
(1351,1443,1303,1357,1359,1398,1446,1495,1499,1509,1510,1513,1400,1291,1399,1496,1497,1498,1299,1285,1360,1384)

Open research data

Alternative provider to Elsevier for the survey because of negative perception and conflict of interest

There were several criticisms of using Elsevier to gather data through the survey, but no valid alternatives of comparable quality and cost/efficiency were proposed. In this case too, the consortium is responsible for the definition of the survey and the construction of the indicator. The survey was already carried out in 2017, with positive reception by the community, and continuity is a value added of the analysis.
Moreover, full anonymised data from the survey will be openly released, just as in 2017.
(1270,1314,1356,1361,1378,1417,1425,1523)

Use alternative surveys such as Figshare's
The consortium already includes the results of other surveys, such as Figshare's 2017 survey, in the dashboard. When available, new data will be added.
(1356,1523)

Sources have insufficient coverage
With regard to Sherpa, unfortunately this limitation is unavoidable. The information about the biased coverage will be included in the presentation of the indicators. (1421)

New indicators and sources
The consortium received many useful proposals, but only few of them included immediately usable data sources.

(1292,1300,1301,1211,1296)
Most proposals need additional effort. To enable this effort as well as additional collaboration on any indicator, the consortium will set up additional collaboration spaces, beyond the one-off consultation about the methodology.
(1318,1460,1504,1505,1506,1507,1508,1332,1333,1522,1358,1385,1414,1518)
Some suggestions were relevant for other sections.
(1388,1415,1517)
Other suggestions were deemed out of scope or not relevant enough.
(1503,1270,1314,1361,1378,1417,1425,1302,1331,1304)

Open collaboration

Alternative provider to Elsevier for the survey because of conflict of interest
Similar answer to the previous comments also in this case. It is the consortium responsible for processing the data and building the indicators. Sources are assessed purely on merit. In particular, Plum offer high value data that are needed for the monitor.
(1310,1364,1393,1365,1408,1370,1371,1372,1373,1374,1278,1279,1313,1319,1323,1375,1416,1488)

Need to avoid using proprietary data
Proprietary data are used where no open data are available, and there are no open data available on altmetrics. The alternative would be using Altmetric.com, which is also proprietary.
On a different note, Mendeley provides reading statistics as open data, which are useful to elaborate indicators, although obviously limited in scope. Appropriate disclaimers will be included in the dashboard.
(1215,1380,1383,1389,1411,1257,1258)

Use multiple sources
With regard to altmetrics, the obvious alternative is altmetric.com – which requires a license. The consortium will investigate the feasibility of the license, in order to carry out ad hoc "robustness checks" for the analysis.
With regard to readership data, Mendeley is the only provider of open data on this.
(1394,1407,1435,1256,1337,1447,1461,1464,1476,1485,1338,1263,1262,1410,1255)

Remove some indicators because not valid
The consortium agrees to remove some indicators of limited validity, in particular the indicators related to open code since GitHub and other repositories does not provide a way to define coding projects related to science.
(1334, 1254, 1386, 1320)
With regard to readership and social media, the indicators are considered important and useful. They will be reassessed at the time of the analysis.
(1409,1448,1486,1259,1367,1260,1277,1336,1368,1261,1335,1369,1529)

New indicators and sources

Some comments included new indicators and sources, which will be included in the methodology.

(1213,1294,1387)

Other comments had interesting proposals but no feasible sources. Ongoing collaboration will take place to better define the indicators and the sources.

(1434,1433,1321,1322,1363,1524,1527,1395,1528,1228,1339,1362,1391,1437,1477,1520,1530,1521,1295,1212,1239,1225,1376,1462)

Finally, some comments were deemed out of scope or contained suggestions for indicators not relevant enough.

(1257,1258,1404,1214,1293,1405,1436,1451,1475,1525,1526,1264,1377,1412,1450,1452,1453,1511,1512,1531,1532,1533,1534,1535,1536,1537,1538,1539,1540,1541)

Annex 3: Methodological approach for the case studies

Research Overview

The study employs a multiple case study of a *thematic sampling* of 30 research projects across different disciplines and countries worldwide to support the assessment about drivers, barriers and impact of open science (Eisenhardt, 1989). The projects are selected across the open science trends: open access, open data, open peer review, open science hardware, open code, and reproducible science, citizen science, and open collaboration. The selection has emphasised open science trends where there is a lack of bibliometric data or where the quantitative data available is anecdotal.

In order to make the selection, the study has dynamically generated a database of open science projects, that is, research projects that have adopted at least one of the trends or projects dedicated to supporting the development of open science. The notion of a research project is recognised as a unit of analysis by researchers, institutions and funders alike. The database is continuously enriched by the study members based on desk research, community members recommendations and by mining the data of open science platforms.

The selection of cases has been made in three sequential phases (M1, M6, M12) to provide the case analysis to the open science monitor in cascade, while in parallel the project database is being completed. The project database includes descriptive data about the research projects: types of trend, discipline, country, duration, and others. The list is dynamically used for identifying the case studies to be carried out.

From the preliminary list of cases in the project database, we have selected a cross-section of 13 *instrumental cases* (Stake 1995) along the open science trends that facilitated our understanding of drivers, barriers and impacts (i.e., to science, industry, and society) of open science. Three major ***types*** of cases have been performed:

i. *Policy cases*, which refer to case studies devoted to studying open science policies at different governmental levels (i.e., national, regional and funder policies) across Europe.

ii. *In-depth case studies,* which are exploratory studies of open science projects that have combined multiple data collection methods, including secondary data analysis, semi-structured interviews and/or study visits (observations). The average length of the analysis provided is around 7000 words.

iii. *Informative case studies*, which are descriptive case studies about open science projects that have been carried through desk research by analyzing available secondary data. The average length of the analysis provided is around 3000 words.

Data Collection

The data collection process focused on a diverse set of primary and secondary data. For the in-depth case studies, primary data has included until January 2019 13 semi-structured interviews and direct observation from one study visits for one of the case studies (i.e., White Rabbit).

Interviews for the cases were chosen on the initial recommendation of the leader of the open science project appearing in the public sources, with subsequent recommendations from the interviewees. The goal was, in general for the in-depth cases, to interview a representative cross-section of the project team.

Secondary sources included information retrieved from publications, projects repositories, wikis, websites, blogs and other social media information referring to the project, amongst others.

Table 1 Number of interview respondents per case

Title Case Study	**Nº Interviews**
White Rabbit	3
Open Targets	3
Pistoia Alliance	3
UK policy	1
Finland policy	1
UK Research Softare Engineers	1
Datacite	1
NL policy	0
Comparing Wos and Scopus	0
Total	13

Data Analysis

The study has employed an embedded design (Yin, 2009), focusing on each open science project at three levels: (1) management team in the project; (2) project characteristics; (3) organization characteristics and policies. The analysis of cases includes:

1) An *analysis per case,* as an entity itself, where researchers have provided background information about the case, which includes a literature review related to the open science trend where the project is located, and some contextual information about the project itself; drivers, which uncovers the motivations and main driving forces making the project possible; barriers, which describe the major bottlenecks and challenges encountered by the project team in the course of the project; and impact or direct effects of the project towards the scientific community, business ecosystem and social benefits of the project at large.

2) An *aggregated analysis* of cases (i.e., cross-analysis). The multiple case designs allow replication logic, by treating the cases as a series of experiments and focus on

identifying unique patterns of each case and finding patterns across different cases. Identification of similarities and differences between cases help us to categorise the different mechanisms that operate in the different projects and relationships. While the analysis per case has been done by a single research team of one of the organizations involved in the Open Science Monitor (i.e., CWTS, ESADE, and LC), the cross-analysis has been performed by a mixed research team combining members of the three organizations. The tasks for the cross-analysis were distributed among the mixed research team, and different group discussions supported the analysis, which has been performed in several iterations.

Two cross-analysis have been envisaged: the first one in January 2019 (included in present report), which includes 13 cases; and the last one scheduled for M24, which will include all 30 case studies. The preliminary results of the cross-analysis will be shared with the Open Science Monitor Advisory board, composed by 15 members across the open science community. The research team will revise the analysis according to the feedback provided by the experts in the advisory group.

Status: Overview of case studies performed

At this stage of the Open Science Monitor (January 2019), the following case studies have been performed:

Table 2 Overview of case studies (up until January 2019)

Trend *Type of case*	**Open access**	**Open data**	**Open collaboration**		
			Open Code	Open hardware	Open review
Policy case	Netherlands				
	Finland				
	United Kingdom				
			Research Software Engineers (UK)		
In-depth case		Open Targets		White Rabbit	
		Pistoia Alliance			
		DataCite			
	Web of Science & Scopus				
Informative case	F1000	Reana			F1000
		Yoda			

Annex 4: Methodological approach for the survey

Study design

A cross sectional study – an online survey delivered by email to respondents using the Confirmit survey platform. The survey provides a snapshot of the current scientific environment and attitudes of researchers, open data and its reuse in research; as we had collected data on this topic previously in 2016 we were able to compare to previous data collected.

An e-mail was sent to active researchers requesting them to participate in a survey on open data. The survey was conducted using the Confirmit survey platform. Researchers were selected from Scopus database of published researchers. The survey invitation was sent out on 27th September 2018, with a reminder sent a week later.

Participants

Researchers were randomly selected from the Scopus database of published researchers, with the sample profiled so country and subject area speciality were tracked to ensure sufficient responses. These were measured against the known distribution of researchers according to the OECD and UNESCO. The respondents were able to change their answers at any time before submitting the filled-out questionnaire, but not after. Emails sent to the participants contained unique links to access the survey.

Study size

The survey was sent to just over 40,000 researchers, and the reminders were sent to the non-respondents. 1029 responses were received (2.5% response rate).

Informed consent and ethics approval

Participants were informed in the invitation letter and in the survey description about the purpose of the study, the research team behind the survey, the median time (15 minutes) needed to complete the survey (based on our survey pilot data), that the data will be made publicly available, no identifying information will be shared, and that by filling out the questionnaire they will be giving their consent to participate in the research.

Incentives

No incentives, except the option to be alerted of the study results, were offered to the participants.

Data sources/ measurement

Respondents were required to answer all the survey questions or choose the 'do not know/not applicable' option.

Respondents could contact the investigators, if they encounter any, technical or other difficulties via email (reply email address, but also shown within the invitation)

Storage

All data was stored on the Confirmit platform during data collection. The anonymised survey data is available on Elsevier's data storage platform (Mendeley Data) and on our projects' data repository site. Responses are confidential and stored in a secure environment.

Bias

As per any large anonymous online survey, it is possible we may experience response bias, i.e. respondents' opinions differing systematically from those of non-respondents as they are interested in the topic of the survey. Survey responses were weighted to be representative of the researcher population (UNESCO counts of researchers). All results in the report are weighted; base sizes are unweighted.

Statistical methods

The variables will be presented as absolute number and percentages and (exact) 95% confidence intervals (CI).

Annex 5: Participants to the experts' workshop and members of the advisory group

Andreas Pester, Researcher, Carinthia University of Applied Sciences

Barend Mons, Scientific Director, GoFair

Beeta Balali Mood, Consultant, Pistoia Alliance

David Cameron Neylon, Senior Scientist, Science and Technology Facilities Council Didcot

Emma Lazzeri, National Open Access Desk, The Italian National Research Council - Institute of Information Science and Technologies (CNR-ISTI)

George Papastefanatos, ESOCS, Research Associate Management of Information Systems Research Center "Athena"

Heather A. Piwowar, Cofounder, ImpactStory / Unpaywall

Jason Priem, Cofounder, ImpactStory / Unpaywall

Marin Dacos, Open Science Advisor to the Director-General for Research and Innovation, French Ministry of Higher Education, Research and Innovation

Michael Robert Taylor, Head of Metrics Development, Digital Science & Research Solutions Limited

Paolo Manghi, Technical Manager, Institute of the National Research Council of Italy

Paul Wouters, Professor of Scientometrics, Director Centre for Science and Technology Studies, Leiden University

Rebecca Lawrence, Managing Director, F1000Research open for science

Roberta Dale Robertson, Open science policy senior analyst, JISC

Žiga Turk, Professor, University of Lubljana