OPEN SCIENCE MONITOR

UPDATED METHODOLOGICAL NOTE

Brussels, October 8th 2018

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 and on 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 recognized 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 standardized 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.

---

[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

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.

1.1 Objectives

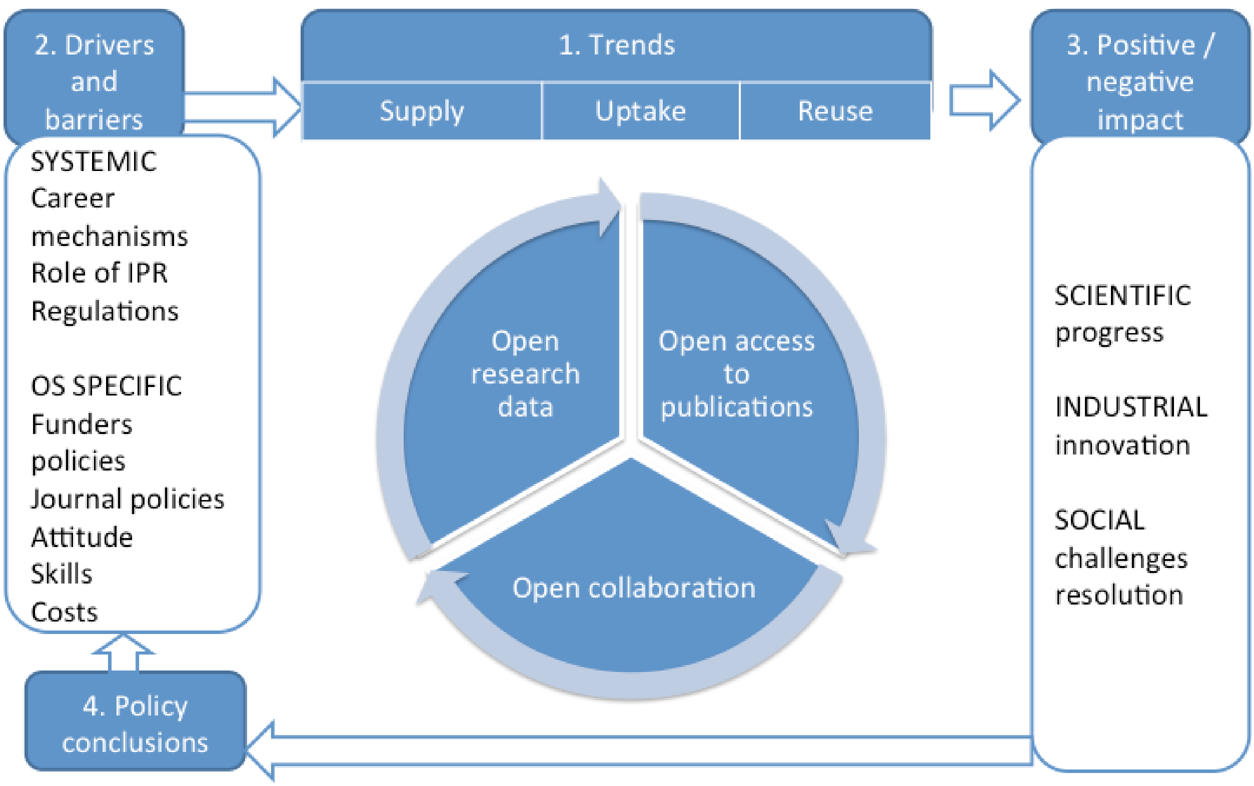
The OSM covers four tasks:

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 scientist 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,

[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)

	• 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).

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 update 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 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, DOAJ, ROAD, PubMedCentral, CrossRef, OpenAire
P(oa) - # Scopus publications that are Open Access (CWTS method for OA identification)	Scopus, DOAJ, ROAD, PubMedCentral, CrossRef, OpenAire
P(green oa) - # Scopus publications that are Green OA	Scopus, DOAJ, ROAD, PubMedCentral, CrossRef, OpenAire
P(gold oa) - # Scopus publications that are Gold OA	Scopus, DOAJ, ROAD, PubMedCentral, CrossRef,

[3] http://v2.sherpa.ac.uk/juliet/
[4] http://www.sherpa.ac.uk/romeo/index.php?la=en&fIDnum=|&mode=simple

	OpenAire
PP(oa) - Percentage OA publications of total publications	Scopus, DOAJ, ROAD, PubMedCentral, CrossRef, OpenAire
PP(green oa) - Percentage gold OA publications of total publications	Scopus, DOAJ, ROAD, PubMedCentral, CrossRef, OpenAire
PP(gold oa) - Percentage green OA publications of total publications	Scopus, DOAJ, ROAD, PubMedCentral, CrossRef, OpenAire
TCS - Total Citation Score. Sum of all citations received by P in Scopus.	Scopus, DOAJ, ROAD, PubMedCentral, CrossRef, OpenAire

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.

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

[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.

Attitude of researchers on data sharing.	Survey by Elsevier, follow-up of the 2017 report.[6] Other existing surveys will be also included in the monitor.
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

[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.
[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.

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

2.3.2 Open scientific hardware

The few comments received here pointed to the limited importance of open hardware licences, 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	Scopus, Mendeley & Plum

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

Mendeley. P(mendeley)/P(tracked)	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 Bachelor students)	Scopus, Mendeley & Plum Analytics
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 the next round of indicators by January 2019, as planned, together with case studies.

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

Annex 1: Technical report on the identification of Open Access publishing

Thed van Leeuwen & Rodrigo Costas

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

Introduction

In this document the approach for the identification and creation of the Open Access (OA) labels for the *Open Science Monitor* (hereafter referred to as OS Monitor) is presented. As stated in the Terms of reference, CWTS is following the exact same method that has been developed over the last two years, and which has been reported at the Paris 2017 STI Conference (van Leeuwen et al, 2017). In this method we strive for a high degree of reproducibility of our results based upon data carrying OA labels following from the methodology we developed. Our initial developments were based on the Web of Science, but for the OS Monitor the exact same method will be based on Elsevier's Scopus data.

The methodological approach that we propose mainly focuses on adding different OA labels to the Scopus database, using various data sources 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 labeling 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). 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.

Data sources used for establishing OA labels

As main data sources to identify evidence of Open Access for publications covered in the Scopus database for the years 2009 to 2016, we used:

- the DOAJ list (Directory of Open Access Journals) [https://doaj.org/],
- the ROAD list (Directory of Open Access scholarly Resources) [http://road.issn.org/],
- PMC (PubMed Central) [https://www.ncbi.nlm.nih.gov/pmc/],
- CrossRef [https://www.crossref.org/], and
- OpenAIRE [https://www.openaire.eu/]

These five sources serve to label the publications according to the terminology used in the OA development. The first two sources (DOAJ and ROAD) serve to identify and label ***Gold OA***, while the last three sources (PMC, CrossRef and OpenAIRE) serve to identify and label ***Green OA***. In cases where publications published in Gold OA journals were also identified in one of the other sources, we determine the status of the publication as Gold OA. So Gold OA goes over Green OA, as Gold is a more deliberate choice of the authors, often driven by a mandate of publishing in a journal that is fully OA.

All these five sources fulfill the above-mentioned requirements while other popular 'apparent' OA sources such as ResearchGate and SciHub fail to meet these two principle requirements. Thus, it is important to stress here that our approach has a more policy perspective than a utilitarian one (i.e. just identifying publications that are freely available). In other words, our approach aims to inform the number and share of sustainable and legal OA publications (i.e. publications that have been published in OA journals or archived in official and legal repositories), instead of the mere identification of publications whose full text can be retrieved online (regardless the source or the legal status of the access to the publication). For a broader discussion on other types of OA as well as other possibilities of identifying OA we refer the reader to our recent paper Martín-Martín et al. (2018) [https://arxiv.org/abs/1803.06161]

Sources of Open Access evidence

The sources that were mentioned above were fully downloaded (as provided by the original sources) using their public Application Programming Interfaces (API). The metadata obtained has been parsed and incorporated into an SQL environment in the form of relational databases.

DOAJ

A first source we used was the DOAJ list of OA journals. This list was linked to the Scopus database on the basis of the regular ISSN code, as well as the eISSN code available in both the DOAJ list as well as in the Scopus database. This resulted in a recall of 1,028,447 publications labeled in Scopus as being OA, via the regular ISSN code, while the eISSN code resulted in 95,162 additional publications.

ROAD

A next source used to add labels to the Scopus database is the ROAD list. ROAD has been developed with the support of the UNESCO, and is related to ISSN International Centre. The list provides access to a subset of the ISSN Register. This subset comprises bibliographic records which describe scholarly resources in OA identified by an ISSN: journals, monographic series, conference proceedings and academic repositories. The linking of the ROAD list is based upon the ISSN code, as well as the eISSN code available in both the Scopus as well as in the ROAD list. This resulted in a total of 524,082 publications being labeled as OA, while the eISSN code resulted in 938,787 additional publications.

CrossRef

A third source that was used to establish an Open Access label to Scopus publications was CrossRef, based upon the DOI's available in both systems. This led to the establishment of a total of 37,119 publications as being licensed as OA according to CrossRef.

PubMed Central

A fourth source used is the PubMed Central database. This is done in two ways; the first based upon the DOI's available in both the PMC database as well as in the Scopus database. This resulted in total in 1,974,941 publications being labeled as OA in the Scopus environment. The second approach was based upon the PMID code (where PMID stands for PubMedID) in the PMC database as well as in the Scopus database. This resulted in a total of 1,102,937 publications being labeled as OA in the Scopus database.

OpenAIRE

A fifth and final data source used to add OA labels to the Scopus database is the OpenAIRE database. OpenAIRE is a European database that aggregates metadata on OA publications from multiple institutional repositories (mostly in Europe), including also thematic repositories such as **arxiv.org**. The matching is done in two different ways: the first one based upon a matching by using the DOI's or PMIDs available in both OpenAIRE and in Scopus (resulting in 2,326,442 publications); and second, on a fuzzy matching principle of diverse bibliographic metadata both in Scopus and OpenAIRE (including articles' titles, publication years and other bibliographic characteristics) (resulting in total in 2,976,620 publications) (the methodology is similar to the methodology for citation matching employed at CWTS – Olensky et al. 2016.

In comparison with the previous studies in which our methodology of labeling OA was applied to Web of Science (WoS), the implementation of the methodology on the Scopus database offers with respect to the DOAJ and ROAD lists the advantage that Scopus also contains the eISSN codes, contrary to WoS. This results in a relative larger number of publications covered by the methodology related to DOAJ and ROAD, hence the numbers of publications as well as the share of publications in Gold OA are higher as compared to results obtained for the WoS database.

The fuzzy matching algorithms underlying the linking of OpenAIRE to Scopus have been revised, and made more accurate in comparison to the previous version of the algorithm. So this probably leads to higher recall as well. Due to the fact that this is applied first in WoS and now in Scopus, with both databases differing in coverage and also time periods, it is impossible to state what the exact difference is.

A new source of Open Access evidence: Unpaywall data

More recently, a new source for OA evidence appeared on the scene, the former OADOI, nowadays *Unpaywall* database (https://unpaywall.org/). CWTS is working on integrating this data source into the current analysis, and plan to include this in a next run of the analysis, within the OS Monitor with (potentially) additional OA publishing information. For now we have conducted a few analyses, comparing our methodology and the numbers of publication labeled 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 data *Unpaywall* actually disclose, whether all types of OA evidence actually fit into our criteria of building OA evidence, and whether there are other potential conceptual issues related to some typologies of OA provided by *Unpaywall* (e.g. it is not totally clear whether the Bronze OA typology disclosed by *UnPayWall* can really be considered a sustainable form of OA, cf. Martín-Martín et al, 2018).

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: 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