Open Science Monitoring

Methodological Note

Date: 23/01/2017

Research and Innovation

Author information and acknowledgements

This study was delivered by **RAND Europe**, with the support of **Deloitte** and **Digital Science & Research Solutions**, including **Altmetric.com** and **figshare**. The project was commissioned by the European Commission Directorate-General for Research and Innovation and received funding under specific contract 2015/RTD/A6/SC/PP-02741-2015.[1]

The team at RAND Europe included: Elta Smith (project leader), Salil Gunashekar (project manager), Sarah Parks, Catherine Lichten, Louise Lepetit, Molly Morgan Jones, Catriona Manville and Calum MacLure. The team at Deloitte included: Valentina Cilli, Maruca De Ramón, Vilma Zotou, Benoit Vandresse and Sebastiaan van der Peijl.

Data sources

We would like to thank the following organisations for generously contributing their data for use on the monitor: 101innovations, Altmetric, arXiv, bioRxiv, Clarivate Analytics, F1000Research, figshare, GenBank, Helmholtz-Centre for Environmental Research – UFZ and German Centre for Integrative Biodiversity Research (iDiv), Nature Publishing Group and Palgrave Macmillan, OpenAIRE, PeerJ preprints, Publons, re3data.org, RePEc, SHERPA/RoMEO, SHERPA/JULIET, Taylor & Francis, and Wiley.

Expert panel

We are grateful to our expert panel members who provided helpful advice and guidance: **Aletta Bonn** (German Centre for Integrative Biodiversity Research (iDiv) and Helmholtz Center for Environmental Research (UFZ), Friedrich-Schiller-Universität Jena, Germany); **Thomas Crouzier** (School of Biotechnology, Royal Institute of Technology (KTH), Sweden); **Ivo Grigorov** (National Institute of Aquatic Resources, Technical University of Denmark); **Peter Kraker** (Postdoctoral Fellow, Know-Center Social Computing Group, Graz University of Technology, Austria); **Eleni Malliou** (National Documentation Center, Greece); and **Paul Wouters** (CWTS-Centre for Science and Technology Studies, Leiden University, The Netherlands).

Acknowledgements:

We would like to thank the following individuals for their contributions to the monitor: Susan Guthrie (RAND Europe), Gavin Cochrane (RAND Europe), Alexandra Pollitt (King's College London), and Marion Bywater who acted as our quality assurance reviewers and provided thoughtful and timely comments on our analyses; Jessica Plumridge (RAND Europe) who designed the graphics presented in the online monitor; Monica Hertzman (RAND Corporation) who provided technical assistance in relation to the design of the website; Steven Wooding (RAND Europe) who provided advice and support throughout the study; Vincent Larivière (Observatoire des Sciences et des Technologies) for his insights and advice; Joanna Chataway (University of Sussex) for her advice and support; Rebecca Ioppolo (RAND Europe) who provided research assistant support; Jack Melling (RAND Europe) who helped in the dissemination activities related to the launch of the monitor; and the many individuals and organisations who participated in an online consultation we conducted as part of the development of the monitor.

For more information about this document or the Open Science Monitor, please contact:

Elta Smith
RAND Europe | Westbrook Centre, Milton Road, Cambridge CB4 1YG, United Kingdom
Telephone: +44 (1223) 353 329 | E-mail: eltas@rand.org

[1] Framework Contract 'Provision of Services in the Field of Research Evaluation and Research Policy Analysis', Lot 2: Data collection and performance indicators to monitor European research policy 2010/S 172-262618

Introduction

This is a methodological note for the study *Open Science: Monitoring Trends and Drivers*.[2] The study was commissioned by the Directorate-General for Research and Innovation (DG RTD) of the European Commission and was delivered by RAND Europe, with the support of Deloitte, Digital Science, Altmetric and figshare.

This note summarises the objectives and approach taken to developing the open science monitor, including the data collection methods; an assessment of the quality, comparability and relevance of the data; and ways to address gaps and further develop the monitor in the future.

Objectives

The Commission is invested in fostering open science by enabling the conditions necessary to realise its full potential. The overarching objective of this study was to inform the Commission's future engagement with open science. There were two main objectives:

- **Objective 1 - Monitor and measure open science trends**: The primary objective of the study was to monitor and measure open science trends in the EU-28 Member States, Associated Countries and other third country scientific partners (Russia, the USA, Canada, China and Japan).
- **Objective 2 - Develop a website to host the monitor**: The monitor should be accessible through a publicly accessible online platform, focusing on policymakers and stakeholders involved in DG RTD's Open Science Policy Platform. It should present a picture of the open science research system at EU level and in relation to third countries. The monitor should also disaggregate information by Member State and/or scientific discipline where possible.

There were four additional objectives:

- **Objective 3 - Develop a methodology for future monitoring**: The study represents a pilot for open science monitoring in the EU. This initiative is expected to extend beyond the life of the study and so a methodology was developed that enables future researchers to reconstruct, adapt and/or change the pilot monitor. This methodological note fulfils the requirements under this study objective.
- **Objective 4 - Assess the drivers of and barriers to open science**: An understanding of open science drivers and barriers is required to enable the Commission to effect changes in open science. The study has built a picture of these aspects of the system, including their impact on open science.
- **Objective 5 - Assess the impacts of open science**: At the time of the study, it was too early in the evolution of open science practices to collect and analyse impact indicator data. Instead, the study team developed case studies to provide preliminary insight into open science outcomes and impacts. The study team developed case studies for each open science characteristic assessed in the monitor.
- **Objective 6 - Evaluate the monitor**: Finally, the monitor itself was evaluated to assess its value and its limits, including its portability, scalability and sustainability over the long-term.

[2] Specific contract 2015/RTD/A6/SC/PP-02741-2015 under the Framework Contract 'Provision of Services in the Field of Research Evaluation and Research Policy Analysis' Lot 2: Data collection and performance indicators to monitor European research policy 2010/S 172-262618.

Approach and method

The study terms of reference set out an ambitious programme of open science characteristic and indicator identification; data collection; contextualisation in terms of drivers, incentives and constraints; analysis of open science impacts; and evaluation of the monitor. The study terms of reference also presented challenges to research design. Monitoring, impact analysis and evaluation were combined in a single study. This required combining methods and research approaches that could be deployed in an effective yet efficient way within the one year timeframe of the study.

Furthermore, when the study was launched to develop the monitor, there was no overarching monitor for open science already in existence, although several EU-level science and technology monitors assessed particular aspects of open science. These included the European Innovation Scoreboard (formerly Innovation Union Scoreboard) (EC 2017a), European Research Area (ERA) progress reports (EC 2014a), the Digital Scoreboard (EC 2017b), and Responsible Research and Innovation trends (ResAGorA 2017). The OpenAIRE platform is an EU-funded infrastructure programme specifically focused on monitoring open access data for research publications (Rettburg & Schmidt 2015).

Jaffe (1999) identified a good indicator as ideally being as precise as possible and unbiased. The relationship between the underlying concept and indicator should be stable over time, comparable, not susceptible to manipulation and subject to aggregation. It is generally not possible to meet all of these criteria, because there are trade-offs in choosing indicators which fulfil the most important objectives of the monitoring system, but it is important to consider all of these aspects in the choice of indicators.

In general, therefore, the monitor developed for this study took as a first principle that successful monitoring of open science requires that the indicators used need to be easily linked to the trend being assessed so that they can be readily understood, and ideally need to be built from well-established data sources.

Each of the steps involved in developing the monitor is discussed below.

Step 1 : Identify open science characteristics

Open science can be regarded as a systemic change in the way research is being conducted, affecting steps throughout the research process, from upstream components, such as idea generation, planning and design, through to the outputs and impacts of research, further downstream. The monitor was designed to assess the different components making up this broader open science system. The study team initially undertook a comprehensive literature review to identify core defining characteristics of open science. These were then presented in an online consultation to invited stakeholders for further validation.[3] Open access publications, open research data, and 'open' communication activities were ranked most important by the majority of respondents. This enabled the team to prioritise and focus on three main characteristics to feature in the monitor.

[3] The consultation involved individuals working in areas related to open science in the EU and third countries, and it primarily included the following two groups: (a) policymakers from within the European Commission and from the Member States and outside the EU, as well as funders, librarians and publishers (invitations were sent to 84 individuals); and (b) researchers, and representatives of academies of science, open data and open access platforms, crowdfunding and crowdsourcing platforms, and citizen science platforms (invitations were sent to 132 individuals).

Based on feedback from participants obtained during the online consultation, the study team determined that another important characteristic in the open science ecosystem – citizen science or civic engagement – should be included in the assessment of open science trends, but would not be monitored in the same way as the other three characteristics. This is because citizen science is part of both the supply and the demand side of open science, and for this reason should be analysed separately. It relates to who is taking part in the scientific process. Moreover, 'measuring' citizen science is much more difficult than is measuring other aspects of open science, for reasons that we describe below.

Characteristic 1: Open access to publications

Among the open science trends, arguably the most fully developed from a policy perspective is open access to research outputs – particularly publications. Open access relates to the accessibility of the research approach undertaken as well as to the outputs of the research. In the preceding Seventh Framework Programme for research and innovation (FP7), which ran over the period 2007-2013, the Commission enabled open access to FP7-funded research outputs with the reasoning that doing so helps optimise the impact of publicly funded research in the EU (EC n.d.).

Horizon 2020, the EU research and innovation framework programme for 2014-2020 – in the spirit of the EU's strategy to boost international cooperation in research and innovation – promotes 'borderless research' that is open to researchers from all over the world, and openness in the sharing of results. It is envisaged that all resulting publications will become openly and freely available, and researchers are encouraged to make available the underlying data (EC 2014b).

A study by the European Commission (2013) found that in the EU-28 Member States, 45% of the papers published from 2008 to 2011 were open access (Archambault et al. 2013). Related studies found that among 48 major science funders, the majority considered both 'gold' and 'green' forms of open access acceptable forms of open access publication (Caruso et al. 2013).[4]

Open access was the characteristic ranked highest among open science characteristics in the online consultation conducted with policymakers and researchers and was also the primary focus of discussions at the 2016 Dutch Presidency Conference on open science (The Netherlands EU Presidency 2016). Open access is viewed as being integral and specific to open science: integral in that it is important for achieving a major goal of open science, which is to accelerate scientific progress and impact for societal benefit, and specific because it is a necessary condition for open science. Open access allows both researchers and citizens to access research findings, extending the flow and use of knowledge within the 'academy' and beyond.

Open access is important to measure in part because this aspect of open science is the most developed, with established policies and momentum behind it to achieve targets for open access to become universal by 2020 (The Netherlands EU Presidency 2016). Articles also remain the primary research output, so measuring the degree of openness in this area

[4] In line with wider policy approaches in this area, the EC identifies two primary routes for open access publication. 'Gold' open access: scientific publishers provide open access to the article immediately upon publication (in this case, the publication costs are borne by the authors, their institution or the funding agency sponsoring the research). 'Green' open access or 'self-archiving': the researcher (or a representative of the researcher) archives the published article (or the final version of the peer-reviewed manuscript) in an online repository before, alongside or after the article's publication (European Commission 2013).

provides insight into the extent to which open science is becoming a 'normal' part of research activities.

While this characteristic is important to monitor, the study team also believe that it should not be over-emphasised, as open science is much more than simply open access, despite the fact that open access tends to dominate discussions at this early stage in the evolution of open science. Additionally, open access may be achieved in a relatively short period of time – potentially in fewer than five years – and it may therefore quickly become 'obsolete' as an area to monitor. This consideration is built in to the monitor as a form of 'future proofing'.

Characteristic 2: Open research data

Open access to research data is less well developed than open access to publications, both conceptually and in terms of policy, but it is viewed by stakeholders as a critical component of open science. Horizon 2020 therefore encourages researchers to make available their underlying research data (EC 2014b). The Commission also launched an Open Research Data Pilot within Horizon 2020, as a step towards increasing the practice of data sharing and re-use of data, and to support the realisation of the societal and economic benefits those data could bring (European Commission 2013). Under the pilot, researchers provide a draft 'data management plan' as part of their proposal, to be finalised in the first six months of the project. Although the plan can be prepared voluntarily by all participants, it is mandatory in the Horizon 2020 Work Programme (2014-2015) for a number of areas of Horizon 2020 that together account for about 20 per cent of the total budget for that period.[5]

Open research data relates both to the transparency of data collection and to enabling the replication of research results. Open data was ranked very highly in the online consultation for this study and was also discussed widely at the 2016 Dutch Presidency Conference. Like open access, this characteristic was considered to be integral to open science, although with more caveats. For example, compared with open access to publications, open access to data is more relevant to specific disciplines in terms of the opportunities it provides (at least currently) and requires more metadata to make it useful, such as guidance on where it came from and how it was used.

Furthermore, the study team wanted to distinguish between 'open data' and 'open *useful* data.' That is, simply posting data on a website, for example, with no description of it, or even putting genome sequences in a sequence database but not explaining how they were generated, is not particularly helpful from the perspective of evaluating or reusing data. Accessibility is important to open science (a 'first step'), but it was clear from the consultation that this quality is necessary but not sufficient; the ability to understand, use and reuse data is essential. For this reason, the team chose a case study approach to illustrate one of the trends associated with the use of open research data. This enabled in-depth analysis of specific open research data platforms to illustrate not only that open research data platforms are (increasingly) used, but also that the approaches being used to ensure that the data hosted on the platforms are *useful* to others.

[5] These include future and emerging technologies, research infrastructures (e-Infrastructures), leadership in enabling and industrial technologies (Information and Communication Technologies), several societal challenges, and science with and for society.

Characteristic 3: Open scholarly communication

Communication activities are an important part of open science. Open science communication relates to the engagement of different stakeholders and activities at various stages of the research cycle and touches upon aspects of collaboration, accessibility and transparency. Communication is 'enabled' in some sense by open access and open research data: that is, as research data and findings become more open, there is more for people to communicate about.

Communication activities related to open science help to characterise the way in which 'conversations' about scientific research take shape. Communication activities can take place across disciplines as well as between science and society. Unlike open access and, to some extent, open research data, communication activities related to open science are a more 'bottom-up' characteristic of the movement, driven largely by the research community rather than by policy. Communication spans discussion on social media (e.g. Twitter, ResearchGate and Academia.edu), comments on published papers (e.g. PubPeer), discussion of preprints (e.g. bioRxiv), communication while projects are ongoing, and 'live' project or proposal work (e.g. the 'opening science' book (Bartling & Friesike 2014)), or proposals for the open data prize (Martone et al. 2016).

The definition of what constitutes communication activities and collaboration was blurred in the online consultation. On reflection, it is difficult to draw the boundary between these two concepts. For example, platforms that enable more open communication help promote collaboration. And while there are aspects of communication that are more readily 'measureable' as being 'open', collaboration is harder to characterise in this way.

Characteristic 4: Citizen science and public engagement in research

Many aspects of open science are about new ways that researchers are interacting with one another and sharing information, but another aspect relates to how a different set of research players – citizens – are taking a more active role in science. This dimension of open science relates to who is taking part in the research process. This involvement aligns with EU values for citizen engagement. For instance, *The Future of Europe Is Science*, a 2014 report published by the Commission's Science and Technology Advisory Council (STAC), emphasises the importance of involving citizens in decision making, not just 'small elites', and incorporating citizens' views in science and technology policies (STAC 2014).

While citizen involvement in science is both an aim and an enabler of open science (as are the other three characteristics), there are many dimensions of citizen science to consider in the open science 'ecosystem'. The biggest issue is definitional: what exactly does 'citizen science' mean? It can refer to citizens 'doing science', for example, through crowdsourcing. Or it can mean greater understanding of science by the public by virtue of greater access to information about the research process (e.g. ability to use open research data or download open access articles) or the ability to understand science and engage with scientists (i.e. more 'open' communication through blogs, social media, priority setting for funding, etc.). The public is also engaging in policymaking through, for example, agenda setting for research systems.

More importantly, there is no way to clearly distinguish between 'scientists' and 'citizens'. Considering its importance to open science and the issues described here, the study team believed that including citizen science in the monitor would by necessity limit the number of dimensions that could be presented, which is unduly restrictive and potentially problematic for the long-term monitoring and understanding of the open science system. Therefore, citizen science is presented more qualitatively than are the other characteristics, with information highlighting developments across several of the dimensions discussed here.

Step 2: Choose indicators to monitor trends for each characteristic

Approach to indicator selection

The terms of reference for the study provided an indicative list of potential indicators. The study team first reviewed the list in terms of data availability (**Appendix A, Table 4**) and proposed additional and alternative indicators based on the team's assessment of the indicators against their relevance to the open science characteristics selected for monitoring, data availability and other considerations which are described in the rationale for indicator selection below.

Online focus group participants were then asked to assess the revised list of indicators in terms of their relevance to the proposed characteristics and data availability (**Appendix A, Table 5**) and to propose additional potential indicators (**Appendix A, Table 6**) as well as potential data sources (**Appendix A, Table 7**). The final selection of indicators and associated data source(s) for each characteristic was chosen by the study team in consultation with the European Commission, particularly taking into consideration the feedback from the online consultation and an external panel of experts assembled for this study, and data availability. The indicators presented in the monitor and their associated data sources are provided in **Table 1**. Additional details regarding the data sources, caveats and limitations are provided with the visualisations for each indicator on the monitor itself.

Table 1: Indicators selected for each characteristic and their data sources

Characteristic	Indicator(s)		Source(s)
Open access to publications	Open access publications	Percentage of publications from each year that are open access	OpenAIRE (numerator data) Clarivate Analytics (denominator data)
		Percentage of publications made available by open access journals	Clarivate Analytics
		Rate of green open access publications compared with journal publications	OpenAIRE (numerator data) Clarivate Analytics (denominator data)
	Journal policies on open access		SHERPA/RoMEO
	Funder policies on open access		SHERPA/JULIET
	Researcher attitudes towards open access (multiple sources)		101innovations; Nature Publishing Group and Palgrave Macmillan; Taylor & Francis
Open access to publications + Open scholarly communication	Alternative publishing platforms	Articles published before peer review	F1000Research
	Preprints (aggregated indicator)	Number of preprints	arXiv, bioRxiv, PeerJ Preprints, RePEc
Open research data	Funder policies on data sharing		SHERPA/JULIET
	Research data repositories	Number of data repositories	re3data.org
		Exemplar research data repositories (case studies)	figshare, GenBank
	Researcher attitudes towards data sharing (multiple sources)		figshare, Wiley
Open scholarly communication	Journal policies on open peer review		Publons
	Open peer reviews	Percentage of peer reviews that are published	Publons
		Percentage of publications in PeerJ that use open peer review	PeerJ
	Use of altmetric platforms		Altmetric
	Corrections and retractions	Corrections and retractions recorded in Web of Science	Clarivate Analytics

Rationale for indicator selection

The monitor brings together the four characteristics as a way to demonstrate how the open science ‘system’ is evolving in Europe. Accordingly, indicators that adhere to a set of core principles are included for each characteristic.

First, indicators must be normalised so that they can be compared and have meaning. Raw counts will generally not be useful unless they are placed on a scale which shows whether the count is relatively big or small in its context. For this reason, where possible the monitor focuses on proportions rather than counts. For example, in the context of open access, the monitor shows the proportion of papers that are available open access, rather than the absolute number of papers.

The indicators must also be contextualised in terms of the quality and importance of data. For some indicators, simple counts are not sufficient, and multiple ways of measuring openness are therefore required. For example, in the context of open research data, the issue is not about tracking how much data is open, but how much is usefully open. While it would be interesting to show information on the downloads and views of data from open data repositories such as figshare and GenBank, the study team felt it would more helpful to demonstrate usefulness of such types of platforms on the monitor by complementing quantitative data with qualitative information through the use of exemplar case studies of these two repositories. The case studies help answer questions related to how these data repositories and databases fit into the broader context of open science, their coverage and future plans within the evolving open science ecosystem.

As far as possible, the indicators align with the three core open science characteristics in terms of the number of indicators per characteristic, the types of indicator and the levels of disaggregation. In general, there are roughly the same numbers of indicators for each characteristic. It is inevitable, however, that there is some overlap of indicators across characteristics. For example, to illustrate open scholarly communication, the study team included the proportion of research outputs published on preprint servers as one of the indicators (as this relates to researchers disseminating their research online in an open and transparent way prior to publication). However, this indicator is very relevant to open access, and so it is linked to both characteristics on the monitor. The number of research outputs that are published before peer review is also linked to both open scholarly communication (as the F1000Research model involves open peer review and also relates to researchers disseminating their research online in an open and transparent way prior to publication) and open access (as these research outputs are also provided open access).

Indicators were also selected across the three characteristics to cover a range of similar 'types'. Under open access and open research data, for example, the monitor displays information derived from external surveys that illustrate researcher attitudes towards open access and data sharing, respectively. Data are also shown related to funders with policies on open access and data sharing. Finally, where possible, indicators for open science trends across all three characteristics are disaggregated by subject/discipline, by country and over time.

Indicators were selected to provide insight into open science across the research process – that is, from inputs (e.g. funding policies) through to activities and outputs (e.g. publications) – to develop a picture of how open science is evolving for each characteristic. Critically, however, at this stage there was no way to assess any causal or even correlative aspects of the trends from inputs through to outputs due to a lack of established datasets across the different features of open science. It was still too early in the evolution of open science to demonstrate these kinds of connections; rather, our aim was to show the way open science is taking shape across different stages of the research process and for different types of activity.

Step 3: Choose case studies

There is seldom a clearly defined path from underpinning research to outcomes and impacts, but impact is ultimately what most people wishing to understand the open science system would want to know about. Open science may be beneficial for many reasons, including reducing duplication of research efforts and, equally, enabling the replication of research results where needed. It may also result in faster, more timely research results and ultimately, greater overall efficiency of the research process.

Some of the most prominent examples of open science in action have emerged from large-scale public health crises. For example, data sharing of genome analyses to tackle the Ebola epidemic was widely seen to have enabled geneticists and evolutionary virologists to work together to confirm the origin and transmission mode of the virus, as well as estimated routes of infection and predicted rates of mutation. This information supported crisis management efforts by local and international public health organisations through showing them where to focus their relief efforts and enabling them to develop practical advice to limit the spread of infections. Data sharing was also considered to be helpful for both the public and private sectors to more quickly design new therapies, diagnostics kits and vaccines. Similar efforts have been undertaken in relation to the Zika virus outbreak and to combat malaria.[6]

Evidence has emerged, at least anecdotally, that other benefits of open science are beginning to be realised. Scientists have used online forums and social media to rapidly debunk high-profile studies found to have flaws, such as the 2010 announcement of a discovery of arsenic-based life (Hayden 2012) and the more recent claim that views on same-sex marriage can be changed by a brief conversation with a gay canvasser (Bohannon 2015).

The Joint Research Centre (JRC) completed reports reviewing ongoing citizen science and smart cities activities (Craglia & Granell 2014) and citizen science and 'do-it-yourself' science projects (Nascimento et al. 2014). The JRC found that citizen science projects go beyond data collection, producing such impacts as awareness raising and community building and that it is important that such initiatives emphasise the importance of trust and of giving something back to the community as opposed to focusing solely on taking away information. Citizen science activities can be educational and provide a means of gathering valuable observations or computational contributions (in the case of online games). As citizens take more active roles in science, however, issues relating to bias and conflicts of interest arise that need to be addressed, as they do for 'traditional' scientists (*Rise of the Citizen Scientist* 2015).

There are many indications of changes in the way science is being done, who is doing it, and what they are studying. These examples suggest that these changes may have widely beneficial effects on science and, ultimately, on society, although there will be challenges and potentially negative effects as well. But it is too early in the evolution of open science to measure impact in a comprehensive way for any of the trends we monitor in this project. At this stage, a qualitative, case study approach was chosen to illuminate specific examples of open science in action.

Case studies were selected by the study team based on a review of the literature and feedback from policymakers, researchers and other stakeholders involved in the online consultation run at the beginning of the study. The final selection of case studies was made based on the following criteria:

- Ensuring coverage across the four open science characteristics highlighted by the monitor (i.e. open access, open research data, open scholarly communication and citizen science).
- Providing coverage of open science initiatives across a variety of disciplinary perspectives.
- Availability of published information about the initiative to inform desk research.

The case studies chosen for the monitor and their associated characteristics are presented in **Table 2**.

[6] For example: Ebola Data (as of 23 January 2017: www.eboladata.org) and the Open Source Malaria project (as of 23 January 2017: www.opensourcemalaria.org).

Table 2: Case studies and related characteristics

Case study	Open access to publications	Open research data	Open scholarly communication	Citizen science
Data FAIRport		✓		
Foldit				✓
Geo-Wiki				✓
Polymath			✓	
Research Excellence Framework				✓
Reproducibility Project	✓	✓		
Sloan Digital Sky Survey		✓		
Social Science Open Access Repository	✓			
Structural Genomics Consortium		✓		
Zenodo	✓	✓		
Zooniverse				✓

Step 4: Develop indicators and visualisations

The indicators used in the monitor were chosen in part because they were well aligned to the characteristic and could be constructed from existing data sources. The first step in developing the indicators therefore was to contact potential data providers to request the use of their data. Data were collected from a range of sources, including: other EU projects, surveys, open science companies and open science resource providers (the latter generally being non-profit organisations). The data sources are listed in **Table 1**.

Potential data sources were identified by respondents to the online consultation, and also through the literature search. With each data provider we discussed the indicator we wished to build with their data, and the caveats or limitations of their data for that purpose. Notes including details of the data provider, and caveats and limitations of the data, are provided with each indicator on the website.

The majority of the data were provided in Excel or csv format. In some cases the data were already freely available online, so the data were downloaded. The study team checked each data file and cleaned it where possible/appropriate (for example, ensuring one spelling is used for each country, checking years are within sensible ranges and understanding why they might be outside these ranges). There are small known inaccuracies with some of the data, which are noted in the caveats/limitations.

In all cases, subject data was mapped to the following re3data.org subject categories:[7]

- Life sciences
- Natural sciences
- Engineering sciences
- Humanities
- Social and behavioural sciences

This categorisation was chosen because it was the most suitable to apply to the data sources used. Where there are multiple data sources for one indicator, the data files were cleaned to ensure they contained the same country names (if applicable) and then merged.

In most cases, data visualisations were prepared using the Tableau Public software.[8] This software enables interactive visualisations to be built, which allow users to explore the data in-depth and ask their own questions. In some cases, the data are static and are not amenable to an interactive format. For these indicators, charts were prepared in Excel. All data providers were contacted to review the interactive visualisations of their data for validation and additional inputs.

For as many indicators as possible, we chose data which could be split by country, by subject, and over time. Data on policies were not available as time series data; it was only possible to find data for one specific point in time. However, in future, a second data point could be used to provide initial time series data, which could be expanded in subsequent years. For some data sources (e.g. OpenAIRE), subject data were not readily available, but these could also be added in the future.

For all indicators, the figures shown are estimates based on the best available data. They each help indicate the current state of play for different aspects of open science, and illustrate relevant trends. They should only be considered as indicators for open science trends in general and interpreted in this context. Caution should be taken in making normative or relative interpretations of particular countries or subjects regarding any aspects of open science due to that aspect having a higher or lower value for an indicator, as in many cases there are likely to be contextual explanations for the differences observed.

Step 5: Develop case studies

The case studies were developed primarily through desk research of primary and secondary information about each initiative. The case studies were enriched and validated through interviews and correspondence with individuals involved directly in the initiatives where possible. Each case study adhered to a common template designed by the study team, focusing on the background to the initiative, the open science elements, and the outputs, outcomes and impacts observed.

Step 6: Assess drivers and barriers

Drivers and barriers were identified by the study team based on results from the online consultation conducted as part of the development of this monitor and a review of articles

[7] Subject classifications that contained the category 'multidisciplinary' were also kept as a sixth category. Further details of what falls within each subject category can be found on the re3data.org website. As of 23 January 2017: http://www.re3data.org/browse/by-subject/

[8] See Tableau Public website. As of 23 January 2017: https://public.tableau.com/s/

and reports produced in the past five years (2011-2016) covering open science issues (either in general or particular aspects of it) and presenting evidence on factors that have aided or hindered developments. From these sources, a list of drivers and barriers was created, which was then organised according to whether each item was primarily a political, economic, social or technological (PEST) factor. A fifth category, 'values and norms in science', was also created to encompass an important set of factors that are specific to the science context. Many of the factors apply to all aspects of open science. Where they are specific to certain aspects, this has been indicated. The results are presented on the monitor as a series of five tables, one for each PEST factor category, along with a list of the sources used.

Step 7: The potential for future monitoring

Despite their centrality to the system, open science is more than open access, open research data, open scholarly communication and citizen science. These characteristics were chosen as the focus of our analysis for the first monitoring effort because they represent the core features of open science at present and they are the most well-developed and well-understood aspects of the system. For these reasons, the chosen characteristics also lend themselves more readily to measurement in a monitor, or in the case of citizen science, there is enough information available to present a compelling picture of its importance.

There are many other characteristics of open science that could be included in a future monitor. These include: open code, open software and research infrastructure, open evaluation, open educational resources and open innovation. Each of these is to a large extent enabled by open access, open research data and open scholarly communication activities. But while all of them are important to consider as part of the wider open science ecosystem, they are not as prominently featured. The aspects of open science that were not chosen for this pilot of the monitor should not be excluded entirely from the monitoring effort. We propose that the website could include a tool for crowdsourcing ideas for future monitoring approaches. This could include a feature that allows for comment on the importance of these other characteristics, which is likely to change over time, and therefore identify whether and when they should be monitored and how.

The study team also identified indicators that would, in principle, be better at monitoring open science trends than those that were ultimately chosen for this monitoring exercise. Identifying appropriate indicators for the project required making trade-offs between indicators that are based on already existing data that were feasible to collect and indicators that are relevant to the characteristics that were chosen for monitoring at this stage (**Table 3**).

Table 3: Future updates to the monitor and related indicator(s)

Characteristic	Potential updates or new indicator(s)
Open access	Number of preprints – while we have an estimate for this, this estimate could be improved, in particular for social sciences, by adding other preprint servers, such as the Social Sciences Research Network
Open research data	Number of data DOIs produced
	Use of/citation of data DOIs
Open scholarly communication	Number of post-publication peer review comments on sites which allow this
Open tools and code	Number of code DOIs produced
	Use of/citation of DOIs for code

For this pilot version of the monitor, we have used files provided by data providers. The data on the monitor are therefore static, although the datasets are open and downloadable. In future, the monitor would be most useful if it was dynamic and 'self-updating'. For some data sources the data is already provided in a form where it would be possible with additional effort to make the visualisations dynamic; for others more work would be required, or alternative sources may be needed.

References

Archambault, E., D. Amyot, P. Deschamps, A. Nicol, L. Rebout & G. Roberge. 2013. *Proportion of Open Access Peer-Reviewed Papers at the European and World Levels – 2004-2011*. Montreal: Science-Metrix [produced for the European Commission Directorate-General for Research & Innovation]. As of 23 January 2017: www.science-metrix.com/pdf/SM_EC_OA_Availability_2004-2011.pdf

Bartling, S. & S. Friesike, eds. 2014. *Opening Science: The Evolving Guide on How the Internet Is Changing Research, Collaboration and Scholarly Publishing*. Berlin: SpringerOpen. As of 23 January 2017: http://www.openingscience.org/get-the-book/

Bohannon, J. 2015. 'Science Retracts Gay Marriage Paper without Agreement of Lead Author LaCour'. Science.com/ScienceInsider. 28 May. doi:10.1126/science.aac4659

Caruso, J., A. Nicol & E. Archambault. 2013. *Open Access Strategies in the European Research Area*. Montreal: Science-Metrix. [produced for the European Commission Directorate-General for Research & Innovation]. As of 23 January 2017: http://www.science-metrix.com/pdf/SM_EC_OA_Policies.pdf

Craglia, M. & C. Granell, eds. 2014. *Citizen Science and Smart Cities*. [European Commission Joint Research Centre Technical Report JRC90374.] Luxembourg: Publications Office of the European Union. As of 23 January 2017: http://publications.jrc.ec.europa.eu/repository/bitstream/JRC90374/lbna26652enn.pdf

EC. n.d. *Open Access Pilot in FP7*. The EU Framework Programme for Research & Innovation. As of 23 January 2017: http://ec.europa.eu/research/science-society/document_library/pdf_06/open-access-pilot_en.pdf

———. 2013. ‘Commission Launches Pilot to Open Up Publicly Funded Research Data.’ 16 December. [Press release.] As of 23 January 23: http://europa.eu/rapid/press-release_IP-13-1257_en.htm

———. 2014a. ‘ERA Progress Report 2014.’ European Research Area/ERA Progress. As of 23 January 2017: http://ec.europa.eu/research/era/eraprogress_en.htm

———. 2014b. *HORIZON 2020 in brief.* Luxembourg: Publications Office of the European Union. doi:10.2777/3719

———. 2017a. ‘European Innovation Scoreboard.’ Growth/Internal Market, Industry, Entrepreneurship and SMEs. As of 23 January 2017: http://ec.europa.eu/growth/industry/innovation/facts-figures/scoreboards_en

———. 2017b. ‘Digital Scoreboard.’ Digital Single Market/Digital Economy & Society. As of 23 January 2017: https://ec.europa.eu/digital-single-market/en/digital-scoreboard

Hayden, E. 2012. ‘Study Challenges Existence of Arsenic-based Life’. Nature/News. 20 January. doi:10.1038/nature.2012.9861

Jaffe, A.B. 1999. *Measuring Knowledge in the Health Sector*. Paper presented at the OECD/NSF high-level forum, Measuring Knowledge in Learning Economies and Societies, Washington, DC. As of 23 January 2017: http://www.oecd.org/edu/innovation-education/1855508.pdf

Leonelli, S., D. Spichtinger & B. Prainsack. 2015. ‘Sticks *and* Carrots: Encouraging Open Science at Its Source’. Geography and Environment 2: 12–16. doi:10.1002/geo2.2

Martone, M., P. Murray-Rust, J. Molloy, T. Arrow, M. MacGillivray, C. Kittel, S. Kasberger et al. 2016. ‘ContentMine/Hypothes.is Proposal.’ *Research Ideas and Outcomes* 2: e8424. doi:10.3897/rio.2.e8424

Nascimento, S., A.G. Pereira & A. Ghezzi. 2014. *From Citizen Science to Do It Yourself Science: An annotated account of an on-going movement*. [JRC Science and Policy report.] Luxembourg: Publications Office of the European Union. As of 23 January 2017: http://publications.jrc.ec.europa.eu/repository/bitstream/JRC93942/ldna27095enn.pdf

ResAGorA. 2017. ‘What Is RRITrends?’ RRI Trends. As of 23 January 2017: http://www.rritrends.res-agora.eu/

Rettburg, N., & B. Schmidt. 2015. ‘OpenAIRE: Supporting a European Open Access Mandate’ *College and Research Libraries News* 76 (6): 306–10. As of 23 January 2017: http://crln.acrl.org/content/76/6/306.full

Rise of the Citizen Scientist. 2015. ‘Rise of the Citizen Scientist’. *Nature* 524 (7565): 265. [Editorial.] doi:10.1038/524265a

STAC (Science and Technology Advisory Council). 2014. *The Future of Europe Is Science*. Luxembourg: Publications Office of the European Union. doi:10.2796/28973

The Netherlands EU Presidency 2016. 2016. *Amsterdam Call for Action on Open Science*. As of 23 January 2017: http://english.eu2016.nl/documents/reports/2016/04/04/amsterdam-call-for-action-on-open-science

Appendix A

Table 4: Indicators proposed in the study specification and associated data collection challenges anticipated

Indicators	Data collection assessment
Number of scientific blogs Interactions on blogs, micro-blogs and wikis for research Number of active research networks in the social web Number of content-sharing platforms and cloud services Number of platforms for open annotation of data, corpora, literature Number of data-sharing platforms Number of platforms of open lab books and to share workflows Number of citizen science platforms Number of platforms to share data, share of funding Number of platforms to publish preprint documents Number of platforms to share references	Data not available and/or infeasible to collect
Number of code downloads from SourceForge11 Number of new open code on SourceForge	Data available globally, but difficult to select data specific to science/research
Share of funders funding systematically open access (OA) to publications Share of funders funding systematically OA to data Share of research performing organisation making available online, free of charge (publicly funded) scientific research data Number of journals with open peer review	Data could be collected manually through desk research and surveys
Article reference to blogs; use of social media Size of the Linked Open Data Cloud10 Size of citizen science community Number of datasets made available with official publications Citations of datasets in official publications (e.g. Data Citation Index by Thomson) Open data (see data gathering and analysis) Number of preprint papers published Open access (indicators to be used to monitor OA in the ERA)	Some partial data are available but would require collating

Table 5: Indicators assessed by the online consultation participants

Characteristic	Indicators
open access to publications	Number of platforms/repositories
	Proportion of papers citing open access resources
	Proportion of papers available open access
	Number of funders with policies on open access
open research data	Proportion of papers making data openly available
	Citations of datasets
	Number of funders with policies on data
	Number of studies that reuse existing data
open scholarly communication	Number of platforms/repositories
	Number of mentions of research outputs per platform
	Number of research outputs published on preprint servers or other platforms prior to or instead of in traditional journals
citizen science	Number of platforms/repositories
	Number of crowdfunded projects
	Number of publications with non-researchers as co-authors
	Statistics about public perceptions of science
collaborative research and publications	Number of papers authored by large consortia
	Number of collaborative research initiatives
	Number of funding calls requiring collaboration
	Proportion of scientific publications that have co-authors from more than one country
	Number of participants in collaborative research initiatives

Table 6: Indicators suggested by the online consultation participants

Characteristic	Indicators
open access to publications	Number of important OA articles (for instance with citation number over 200, or with strong altmetric score)
	Funder mandates for open access (including open data) publication, recognizing preprints as well
	Aggregate impact factor or citation numbers of papers available open access
	New impact measures that go beyond citation
	Proportion of grants with funding allocated for open access
	The average cost per article (to get some sense of the effectiveness of the emerging market)
	Quality of open access publications
	Number/proportion of papers made available under each of the major approaches to open access
	Number/proportion of OA articles in accredited digital preservation facilities
open research data	Percentage of data having a DOI
	Some indication (increasing %) of research outputs that are using datasets in labs/research groups/places that are NOT the originators of the research dataset in first place (indicating reuse ability and the speed in this happening)
	Number of articles that reuse and cite open data
	Reproducibility of research
	Quality of data preservation and reuse support services
open scholarly communication	Some measure of volume of communication traffic linked to open data repositories
	Number of post-publication peer reviews or comments
	Proportion of research conducted in the open
	Number of papers pointing to previous and/or ongoing discussions (or data processing, etc.) that already contained original research material, which then led to the published paper
	Number of learned societies, academic libraries, etc. that help to maintain, moderate, capture suitable platforms
	Some measure of the time between submitting an article/piece/dataset to a repository/open science platform and its visibility (it should be as immediate as possible so others can use and build upon where appropriate – if open science is enabling, this should be narrowing from now on
	Number of comments per paper
	Active sharing of information within networks (e.g. count and size of networks, hub between different networks)
	Level of engagement with a platform (e.g. unique users)
	A measure of the openness of peer review
	Number of journals using open or other forms of transparent peer review
	Number of active researchers who use these services to communicate and collaborate
citizen science	Amount raised through crowdfunding
	Proportion of projects conducted in collaboration with citizens
	Number of citizen science projects monitored via social media traffic
	Number of projects initiated by non-experts and similar organisations
	Total amount per country that research funder provide for citizen science (CS) projects
	Number or proportion of citizens participating in CS projects

Characteristic	Indicators
	Number or proportion of researchers collaborating with citizens
	Degree of and variance in types of participation
	Engagement of study participants (medical/translational research) – patient organisation participation, participant boards, feedback mechanisms
	Number of studies responding to citizens' expectations based on evidence or on citizens' consultation
	Indication (perhaps using social media) of alternative metrics of non-academic usage and citation of academic work (e.g. citation of academic work in policy documents)
	Mentions of research outputs in media consumed by society
	Number/percentage of projects involving a citizen science component
	Level of participation in trials
	Time taken to recruit people to trials
	Mention of research in 'public places' (media)
	Measure of diversity of people within science
collaborative research and publications	Use of Web 2.0 tools for collaboration
	Number of cross-disciplinary research projects and papers
	Number of publications authored by multiple institutions and producing open data
	Number of papers which allowed for usage of public or semi-public collaboration platforms – e.g. GitHub, Wikiversity – in preparation for the final paper
	Number of learned societies, academic libraries, etc., which help to maintain, moderate, capture suitable platforms
	Proportion of funding assigned to collaborative projects
	Statistics about scientist perceptions and experience of collaboration
	Number of proposals requiring collaboration across disciplines and groups
	Number of transdisciplinary consortia
	Proportion of transnational consortia
open tools	Number of non-duplicated and reproducible tools
	Number of papers, projects, products that were enabled by tools
	Proportion/quantity of grants and funding to create common Open Access/Open Source research tools
	Proportion of research papers that can be reproduced
	Number of papers where code is available from a software repository
	Measure of different 'profession types' within science, such as technical staff
	Relative number of impressions generated by a research output
	Proportion/quantity of grants and funding to create common open access/open source research tools
open peer review, incentives and rewards	Percentages of reviewers who reveal their identity to the authors, or use Publons to keep track of their peer review activities
	Share of journals with levels of open evaluation
	Proportion of research outputs that are reviews in an open evaluation process
	Proportion of funders incentivising use of research metrics by requesting them in grant applications
	Proportion of graduates schools in a university offering accredited open science training
	Proportion of graduates in a university that attended an accredited open

Characteristic	Indicators
	science course
	Promotion criteria weighting open science
open educational resources	Share of research results reused in open educational resources
open innovation	Citations with patents
	Number of publications with authors from public and private sector

Table 7: Data sources for indicators suggested by the online consultation participants

Data sources proposed
Altmetric
bitbucket
CASRAI (Consortia Advancing Standards in Research Administration Information)
CORE< BASE
CRIS (Current Research Information Systems) and euroCRIS
CrossRef
Danish National Research Database Open Access Indicator
DataCite
DGEEC (Direção-Geral de Estatísticas da Educação e Ciência)
Digging into Data
DOAJ (Directory of Open Access Journals)
ERA (European Research Area) reports and country-related reports
Europe PMC (PubMed Central)
figshare
GitHub metrics (Git is the name of a version control system (VCS))
openAIRE (open Access Infrastructure for Research in Europe)
openDOAR (open Directory of Open Access Repositories)
openMInTeD (open Mining Infrastructure for Text and Data)
PAGES (USA)
Publons
ROAR (Registry of Open Access Repository)
ROARMAP (Registry of Open Access Repository Mandates and Policies)
RRI (Responsible Research and Innovation) indicators
SHARE (SHared Access Research Ecosystem)
SHERPA/Juliet (Securing a Hybrid Environment for Research Preservation and Access/Juliet)
SHERPA/RoMEO (Securing a Hybrid Environment for Research Preservation and Access/Rights MEtadata for Open archiving)
Snowball Metrics
SourceForge
ÜberResearch dimensions (in general for funding)
WoS (Web of Science)
Zenodo