An Open Toolkit for Tracking Open Science Partnership Implementation and Impact

E. Richard Gold[*], *Sarah E. Ali-Khan, Liz Allen, Lluis Ballell, Manoel Barral-Netto, David Carr, Damien Chalaud, Simon Chaplin, Matthew S. Clancy, Patricia Clarke, Robert Cook-Deegan, Megan Doerr, Lisa Federer, Steven A. Hill, Neil Jacobs, Antoine Jean, Osmat Azzam Jefferson, Chonnettia Jones, Linda J. Kahl, Thomas M. Kariuki, Sophie N. Kassel, Robert Kiley, Elizabeth Robboy Kittrie, Bianca Kramer, Wen Hwa Lee, Emily MacDonald, Lara M. Mangravite, Elizabeth Marincola, Daniel Mietchen, Jennifer C. Molloy, Mark Namchuk, Brian A. Nosek, Sébastien Paquet, Claude Pirmez, Annabel Seyller, Malcolm Skingle, S. Nicole Spadotto, Sophie Staniszewska, and Mike Thelwall.*

Request for Feedback

The article and associated documents present a toolkit for tracking the implementation and impact of open science (OS) partnerships. The article describes the need for and the collaborative process used to develop the toolkit while the associated documents contain the toolkit itself. We are now seeking comments and suggestions on both the article and toolkit from the larger community. We specifically seek comments from those studying, working with, or engaged in OS and OS-related projects. In particular, we welcome comments relating to the comprehensiveness of our measures and what may be missing. We also seek comments on whether the breadth of the toolkit is too ambitious to be effectively implemented and, if so, what measures should be eliminated. We further invite the community to identify any projects – OS or otherwise – that may be amenable to collecting and sharing data based on the toolkit indicators. The present toolkit will need to be translated into open source tools that, to the extent possible, collect the data automatically. Any assistance in developing these tools would be most appreciated. Comments will be accepted online on GoogleDocs until **January 31st, 2019**. After the comment period closes, our team will revise the article and toolkit, taking into account proposed edits. We then propose to submit the article and toolkit to the Gates Open Platform for publication. While we include the toolkit in this release, please comment directly on the GoogleDocs below:

- Annual report, guides and survey;
- Open science measures to be considered by others;
- Incomplete and rejected open science measures; and
- Origin of the measurement toolkit.

[*] This project is led by Prof. E. Richard Gold at McGill University. We acknowledge and thank the participants listed in both Supplementary files 5 and 6 as well as Valmi Dufour-Lussier, Michael Gardiner, Andrew Medeiros and Matthew Squire for their organizational and research assistance. We also want to acknowledge and thank the Wellcome Trust, the Bill & Melinda Gates Foundation, UK Research and Innovation and Jisc for their support.

Table of Contents

Introduction

We live in the safest, healthiest, richest and most democratic period in history (Roser, 2018) partly due to our ability to secure clean water, deliver vaccines, institute the rule of law, and develop our ideas of equality and democracy. Despite this, there are rising concerns about the way research is collectively organized, ranging from its escalating cost and lower research productivity (DiMasi, Grabowski, & Hansen, 2016; Munos, 2009; Pammolli, Magazzini, & Riccaboni, 2011), to low public trust in researchers to report the truth even if against the interests of sponsors (American Academy of Arts & Sciences, 2017), lack of diversity and community engagement (Puritty et al., 2017; Valantine & Collins, 2015), and a research culture that incentivizes researchers to publish over producing quality research, leading to questionable research practices and irreproducibility (Begley & Ellis, 2012; Nosek, Spies, & Motyl, 2012; Open Science Collaboration, 2015). Researchers, public research organizations, firms, governments, funders and society more broadly are adopting open science (OS) practices and OS partnerships to address these concerns (Ali-Khan, Jean, MacDonald, & Gold, 2018; Dai, Shin, & Smith, 2018).

OS comprises a set of practices, including open education, open research funding, open access publications, open data and materials, preregistration, and the avoidance of restrictive intellectual property. These practices aim to reduce transaction costs, promote data re-use, increase rigor and reproducibility, decrease redundant research, better involve patients, consumers and others, facilitate researcher transparency in sharing processes and results, and improve connections with a larger variety of actors to produce more innovative approaches and solutions over the medium to long terms (Gold, 2016). Nevertheless, there exists no single standard for OS with the result that different organizations, governments, and firms apply OS as a label for their own favored set of practices.

This article contributes to the OS discussion by proposing the creation of an open toolkit and data set, based on internationally developed and open measures, to provide an evidence base through which we can collectively determine if, how, when, and where partnerships based on OS principles and practices can contribute to social and economic welfare in general and research and innovation (R&I) in particular. Already, the Structural Genomics Consortium (SGC) and the Montreal Neurological Institute (MNI) have agreed to use the toolkit to collect and share data. Acknowledging the different definitions of OS, we set out to measure participation in particular practices rather than determine which set of practices constitute OS.

OS Partnerships

While there are different ways of implementing OS, we focus on partnerships (OS partnerships) in which all partners agree to comply with OS practices in conducting their joint work. Public entities, either with other public institutions or jointly with private firms, can create these partnerships by using and combining the policies, contracts, and infrastructure of institutions to increase knowledge flow and reduce redundancy (Fecher & Friesike, 2014). Relevant public institutional policies include conditions for tenure and promotion, research grant practices, sharing by default, preregistration of studies and analysis plans, the avoidance of restrictive intellectual property rights, patient consent, continuing education and training, publication and data release

(Australian National Data Service, 2017). Contracts relate to standardized forms for material transfer, sponsorship, partnership agreements and subject participation. Institutional infrastructure comprises personnel and the physical and electronic infrastructure that support the immediate, free and usable sharing of data, software, policies, and practices (Gold, 2016).

Through these policies, contracts, and infrastructure, those pursuing OS partnerships aim to increase efficiency and reproducibility, and inspire discovery and innovation (Ali-Khan, Harris, & Gold, 2017). Two institutions are prominent exemplars of OS private-public biomedical partnerships: the SGC and the MNI (Dolgin, 2014; Edwards, Bountra, Kerr, & Willson, 2009; Poupon, Seyller, & Rouleau, 2017). These build on years of open source, open access, and open data partnerships in projects such as Linux, the Apache HTTP Server Project, the Human Genome Project, the SNP Consortium (Thorisson & Stein, 2003), and the Open Source Malaria Project, all of which have delivered significant advance in technology and knowledge .

Despite these successful partnerships, many public research organizations, government policy-makers, researchers, and firms remain uncertain about the costs and benefits of OS and their distribution among stakeholders (Dai et al., 2018). The lack of evidence concerning costs and benefits as well as attitudes and experience, hinders experimentation with OS partnerships upon which to build theory around OS and R&I systems (Ali-Khan et al., 2018).

To overcome this lack of evidence, we propose here the use of a measurement toolkit to spur understanding of OS partnerships, their effects and characteristics. The toolkit consists of measures through which to collect data to be reported annually, interview guides for semi-structured interviews, sample surveys to assess implementation of OS practices, and other measures that can be collected by or for OS and non-OS partnerships. These shared quantitative and qualitative data are based on a common coding framework (See Supplementary file 1). The policies comprise communication, patient and public involvement and engagement, intellectual property management, promotion and peer review criteria, skill development and training, sharing, and commercialization models. We propose that the toolkit and resource become adopted as a community-managed and open toolkit around the globe.

A critical contribution of this article is to propose that prospective data on OS partnerships be collected and shared. A prospective approach will strengthen the quality of the data and move us beyond the more common retrospectively created data sets that inevitably leave theoretical holes, rely on surrogate measures, lack historical context, and result in incomplete data sets (Kemp & Prasad, 2017; Schwartz & Sichelman, 2017). The measurement toolkit will enable prospective collection and sharing of data on OS partnerships. As such, this measurement toolkit will provide richer, more in-depth and harmonized data to better study OS partnerships. With greater knowledge of how these partnerships contribute to R&I, we envision that policy-makers and researchers will devise better indicators of success for particular projects or funding programs.

The measurement toolkit was created with quantitative measures and qualitative approaches that research organizations participating in OS and non-OS partnerships could implement for collecting data about their collaborations. Here, we describe how we created these measures through a collaborative process drawing on the expertise of various stakeholders, including researchers, publishers, and funders. We begin with a literature review outlining the

rationale for our methodology and our conceptual approach. We then describe the development of the measures. We end with a call to the larger community to comment upon and improve the proposed measures and to begin implementing them.

Literature Review

Previous studies have focused more on the practice and implementation of OS and less on the measurable effects OS may have on better engagement, research efficiency, communications, and priority setting, as well as new delivery mechanisms and new products and services (Morgan Jones et al., 2014; National Academies of Sciences & Medicine, 2018; Tripp & Grueber, 2011). For example, some initiatives present both quantitative and qualitative indicators to track openness and transparency in publication and data sharing (Smith, 2017; Smith, Gunashekar, Parks, & Chataway, 2016) and stakeholder understanding and engagement with OS (Ali-Khan et al., 2017; Tuomi, 2016). Other studies developed indicators to investigate how organizations implement OS (Lampert, Lindorfer, Prem, Irran, & Serrano Sanz, 2017; Nosek et al., 2015; Smith, 2017; Smith et al., 2016; Tuomi, 2016) and a few studies have evaluated the implementation or impact of specific OS policies or practices (Hardwicke & Ioannidis, 2018; Kidwell et al., 2016). Our project differs from the other studies by developing more comprehensive measures of both social and economic influence, research outcomes, diversity and inclusion, trust, and opportunities for youth and early career researchers. Our measures aim to facilitate researchers' understanding of the nature and extent of the impact of OS.

In addition to earlier studies on OS, other studies have proposed measures of innovation in general, such as the OECD's Oslo Manual (OECD/Eurostat, 2005). These measures, however, do not evaluate the relationship between OS partnerships and outcomes. Further, many of these measures are *ad hoc* to the specific studies and created based on retrospectively created data sets, limiting their use in more generic contexts. Finally, these measures tend to focus on firms using proprietary models, such as open innovation and closed/semi-closed partnerships (Community Innovation Surveys (Mairesse & Mohnen, 2010); OECD and World Bank Innovation Indicators; OECD innovation scoreboards (OECD, 2010, 2017); and The Global Innovation Index (Cornell University, INSEAD, & WIPO, 2017).

Our aim, in this article, is to propose measures that enable hypothesis-driven research on the influence and impact of OS partnerships on a variety of social and economic outcomes, as well as research culture, rigor, diversity, social capital and patient and consumer voice. The set of measures we propose establishes a global basis for collecting and sharing data and will accelerate not only our collective understanding of OS, but provide support and evidence to those contemplating, implementing or monitoring the effects of OS partnerships.

Methodology

We draw on existing methodologies, with the modifications that we discuss below, to develop the set of measures in the proposed measurement toolkit. In particular, we examine the literatures on evaluation of projects, programs, and knowledge transfer. We adopted a three-stage knowledge exchange process to facilitate our development of the toolkit.

The first body of literature assesses whether projects or programs have achieved their anticipated outcomes. This literature relies on logic models to track whether those partnerships deliver outputs that, over the medium and long terms, produce the outcomes promised by those who established the partnership. There are two reasons why logic models are inappropriate for the creation of the measurement toolkit and the set of measures we propose here. First, as noted, logic models are rigid in that they focus on anticipated outcomes within a model rather than exploring foundational questions (Cooksy, Gill, & Kelly, 2001; Treasury Board of Canada Secretariat, 2012). This narrow focus on anticipated outcomes leaves aside effects that "can be realized by paths other than those presumed by program theory" (Weiss, 1997). Second, we aim for the toolkit to aid in developing theory rather than applying an established theory. As Weiss notes, "if theory is taken to mean a set of highly general, logically interrelated propositions that claim to explain the phenomenon of interest, theory-based evaluation [i.e., a logic model] is presumptuous in its appropriation of the word." Weiss writes that logic models derive from an established theory to evaluate whether anticipated outputs actually result from undertaken activities, but not to develop the theory itself (Weiss, 1997).

Though we do not use formal logic models, we nevertheless acknowledge the importance of developing measures that correspond to potential influences and impact of OS partnerships on R&I systems, diversity, social capital and other critical outcomes. We thus constructed a set of potential hypotheses concerning the influence of OS partnerships, without attempting to eliminate contradictions or alternative pathways. We employed a method of knowledge exchange through which stakeholders come together to identify research questions, jointly construct the measures, collect data and share and analyse that data. In such a method, stakeholders collectively refine knowledge – hypotheses and measures – iteratively until "only the most valid and useful knowledge is left" (Graham et al., 2006). By ensuring a diversity of perspectives in co-creating the set of hypotheses, this process also increases communication and the likelihood of research uptake (Kothari, MacLean, Edwards, & Hobbs, 2011).

We are aware that previously developed measures to describe certain environments have become prescriptive rather than descriptive, often without sufficient analysis of how metrics can establish perverse incentives and perverse side effects (Cain, Loewenstein, & Moore, 2005). For example, the use of patent counts and promised licensing revenues from university technology transfer changed from a useful means of comparison to an output measure of performance (Kim, Anderson, & Daim, 2008). Such practices often lead universities to over-patent and engage in poor licensing practices (Ryan & Frye, 2017). Using descriptive measures as targets – such as number of patents held – rather than providing a snapshot of current activities, also raises significant ethical concerns over the use and dissemination of measures. These concerns can be partially countered by proposing a large enough set of measures to make it difficult to cherry-pick only a handful of measures that can be gamed. Further, combining quantitative and qualitative measures also reduces the risk of gaming.

We recognize that it is difficult to track causal links between phenomena and ultimate impact (Council of Canadian Academies, 2013). Beyond the difficulties in establishing causation, OS practice varies based on the setting, problem, available resources and stakeholders. Additionally, internal and environmental features can also lead to multiple pathways and

interactions between measures and impacts. Some of these features are difficult to capture, including informal knowledge transfer, relationship building, trust and education of new trainees and expert personnel (Nicol, 2008). Instead, we expect relationships between OS practices and outcomes to take the form of a contribution chain that acknowledges influence, but shies away from claiming causation.

A Three-Stage Process

We adopted a three-stage process to implement the knowledge exchange. First, we developed a working definition of OS partnerships based on a review of the literature and of partnerships that consider themselves to be open science (Ali-Khan, Jean, & Gold, 2018). Second, we convened global stakeholders in Washington, DC to map out the ways in which OS partnerships might influence innovation and social and economic outcomes. Third, drawing on these influences and potential outcomes, we brought together experts in measurement, evaluation and empirical studies from a variety of disciplines and countries to develop a prospective set of measures that we propose OS partnerships around the world use to construct data sets.

The global stakeholders we convened in the second stage included thought-leaders from developed and developing nations, intergovernmental organizations, researchers, governments, science agencies, funders, members from the philanthropic sector, patient organizers, and members from biotechnology, pharmaceutical, and artificial intelligence industries. After presenting our definition of open science and discussing the example of the MNI, stakeholders together engaged in a series of facilitated discussions asking what success of OS means from the point of view of researchers, governments, industry, philanthropies and patients. The organizers then summarized these discussions and represented them to the group for further discussion and elaboration. Ali-Khan et al. (2018) summarized those discussions, obtained feedback from participants, and published the results. Through these iterative discussions, stakeholders collectively mapped out the different ways that OS partnerships might contribute to innovation and desired or feared social and economic outcomes. Examples of the jointly-created hypotheses included the following: 1) that OS partnerships would simplify and thus increase exchanges of students and postdoctoral fellows between university and industrial labs; 2) that students practicing OS making the transition to tenure track positions would be hindered by not having their own private data set to found their own labs or, alternatively, that these students would benefit by increasing their exposure to a larger network of investigators; and 3) that OS partnerships would increase the quality of data by encouraging researchers to place more emphasis on data quality and reproducibility prior to public exposure or, alternatively, would decrease the quality of data due to the desire and facility of quickly publishing their work and establishing priority.

As these examples illustrate, stakeholders understood the relationship between OS, research, innovation, communities and the public to be complex, and explored different, sometimes contradictory, hypotheses in order to generate, in the third stage, a set of prospective measures that would allow researchers and stakeholders to investigate that relationship. We published the results of that meeting and proposed seven overarching themes for further exploration as follows: 1) Increased quality and efficiency of scientific outputs; 2) Accelerated innovation and impact; 3) Increased trust and accountability of the research enterprise; 4) Increased

equity in research; 5) Better opportunities and recognition of early career researchers and youth; 6) Positive economic impact; and 7) Implementation success (Ali-Khan et al., 2018).

At the third stage, we assembled a group of global experts across diverse fields – including innovation measurement and policy, law, public engagement, bibliometrics, economics, business and sociology – to develop a set of measures to underpin the development of the prospective measurement toolkit. To provide continuity, we included some participants from the Washington Forum in this workshop. Most participants, however, were new to include individuals with different expertise as well as those involved in other major OS measurement and standard-setting initiatives. The latter included individuals who had worked on the European Commission (EC) OS Monitor, the RAND SGC analysis (Morgan Jones et al., 2014), the EC Expert Groups on Indicators and FAIR Data, the TOP Guidelines and the Metric Tide (Wilsdon et al., 2015). We included these individuals to promote alignment and complementary processes between our proposed measures and measurement toolkit with other global OS measurement initiatives.

The goal of this third-stage workshop was to generate prospective measures based on the seven themes produced at the first meeting (Graham et al., 2006). Matching the hypotheses generated in the first workshop to measures enables the testing of hypotheses about the influence of OS partnerships (Canadian Academies of Health Sciences, 2009; Tracz & Lawrence, 2016). Accordingly, we organized participants into groups corresponding to the seven themes identified in the first workshop. These groups developed working documents with a mixture of quantitative (e.g., counts, revenues, patents, students, survey results, etc.) and qualitative (principally semi-structured interview guides) to provide a nuanced set of data through which to study OS partnerships.

Following the third-stage workshop, we reviewed and organized the proposed measures. We eliminated duplicate measures and put aside for future work those that were missing critical information – e.g., lack of data source, coding frame, or clear connection to a hypothesis. We sorted (and in some cases adapted to fit a partnership context rather than a country or region) those measures that could be implemented in the study of individual OS partnerships from those that related to general environmental conditions, such as overall government funding or education levels generally. We also recorded measures proposed at the workshops that were specific to countries, specific databases (e.g., databases of academic articles such as PubMed or Web of Science), or that would require the state to compel information disclosure (e.g., by governmental statistical agencies). We leave these to others to expand and potentially implement in other contexts. We present our outcomes below.

Results

The outcome is a set of measures that can be collected about OS and non-OS partnerships – and potentially individual institutions or projects – which agree to do so, and the resulting data shared openly. This data will not only create a baseline for analysis but will provide insight into the evolution of research and innovation practices. We divided the measures into separate instruments based on the nature of the measures (quantitative or qualitative), source of the data (participants in the partnership, social science group observing the partnership, or other entity). The seven themes we identified crossed these categories, making them less relevant as an

organizing framework of these instruments; nevertheless, we preserved the underlying hypotheses, themes and working group information as metadata to document their origin (See Supplementary file 4).

The measures include the following components (See Supplementary file 1):

1. A form of annual report of quantitative data related to the partnership, such as publications and data sets (including their persistent unique identifiers such as DOIs), number of students, student employment post-graduation, authorship, investments, etc.;
2. A series of semi-structured interview guides to better understand norms, attitudes and understanding across the spectrum of stakeholders involved in the partnership (e.g., do you feel that you derive benefit from your participation in the OS collaboration? What challenges and opportunities does OS present for your business?);
3. A form of survey to identify implementation of OS practices within the partnership; and
4. A select number of other quantitative measures that need to be collected by those with expertise in certain social-science methods. These include, for example, measures that require linking publications with citations in the academic, grey or patent literatures.

Beyond this set, we identified a non-exhaustive set of measures that can be best implemented by governments, intergovernmental organizations, research funders, agencies, or database owners that are not specific to any one OS partnership (See Supplementary file 2). Finally, we recorded incomplete and rejected measures so that the community may draw on these in the future (See Supplementary file 3).

The measures we propose are in plain language and are user-friendly in conformity with best knowledge dissemination practice, thus encouraging user uptake (Kothari et al., 2011). We include definitions, data sources and coding rules in addition to tracing how we developed the measure and underlying hypotheses that lead to it.

In accordance with good practice, the measures we propose are aimed to be transparent and clear in their coding. We also aimed for the necessary data to be cost effective and easy to collect across a spectrum of OS partnerships. As noted in the methodology section, we combined qualitative assessments to support quantitative evaluations. By publishing these measures, definitions and instruments on an open platform that allows comment, transparent updating and review, we have created the opportunity to continuously update the measures, introduce new ones and retire those that prove difficult to collect or share in practice (Wilsdon et al., 2015).

Discussion

We developed the set of measures proposed in this article as a necessary step towards the construction of a global measurement toolkit on OS partnerships, which we see as key to understanding changing research and innovation environments and to the role and impact of OS in particular. We anticipate that partnerships around the world will collect and share data on OS practice and outcomes by drawing on our measures. The resulting measurement toolkit will provide researchers with the ability to validate data and improve the measurement toolkit, and to test hypotheses to develop a grounded theoretical understanding of the contributions, positive and negative, of OS partnerships on research, innovation and social and economic life. Stakeholders

can also draw on the data to better appreciate their own organizations and operations. Decision-makers in government, industry, universities and community groups will be able to draw on this learning to structure future OS partnerships and to eventually develop logic models through which to assess particular partnerships.

The economic and social influence of OS partnerships may take years to materialize and may be subject to a plethora of diverse influences. While we recognize that OS successes do not happen in a vacuum, careful empirical analysis of OS will nevertheless help researchers identify key determinants of values and benefits of OS. This will allow the community to propose mechanisms to enable OS practice and to define the contribution chain between OS activity and outcomes.

We acknowledge certain limitations to the measures we propose and call on other researchers to investigate and propose improvements. First, while our stakeholders included individuals and institutions from developing countries, data for some of the measures will be easier to collect and most relevant to partnerships in industrialized countries. This is because data sources will likely be more available in industrialized countries and sharing mechanisms, motivations, and barriers to implementation may differ across countries. Specifically, we recognize that data collection in lower income countries is constrained by lack of resources, weaknesses in institutional organization, and inability of governments and organizations to collect reliable and appropriate data (Elahi, 2008). Further research is needed to determine the suitability of our proposed measures, to propose additional measures and to investigate ways to access data sources. Second, the indicators predominantly (but not exclusively) relate to the life sciences, with a particular focus on biomedical science. Whether these indicators are as suitable to other fields such as nanotechnology, information technology, health system analysis, environmental sustainability, arts (digital, visual or performance), agriculture, or history, for example, needs to be investigated.

Finally, to mitigate the dangers of misuse of the measures and their associated data, we encourage those who are using the measures to use them openly and transparently. By doing so, the community can better monitor use of the measures and quickly respond with any concerns arising from their use.

Conclusion

Measuring the influence of OS partnerships is important to improving R&I systems because deeper understanding of OS influence will reduce uncertainty about the relative benefits, positive impacts, and negative impacts of OS partnerships. This uncertainty manifests itself in several ways: in a lack of trust in open and public scientific knowledge generation, in a lack of policy frameworks in some countries and by inertia within public research organizations, and in a failure of researchers, public research organizations, communities, or firms to experiment with OS partnerships.

Implementing the set of proposed measures will lead to a data resource to aid in understanding the role of OS partnerships in R&I systems. This data resource might encourage the establishment of OS partnerships by mitigating the uncertainty surrounding OS partnerships, contributing to a better theoretical understanding of OS, and encouraging a shift towards more

openness and inclusivity in science. To fully realize this understanding, diverse communities will need to investigate the benefits and drawbacks of using OS approaches using such evidence-based metrics. By doing so, communities can generate an evidence base regarding beneficial impacts and drawbacks of OS, and share data openly as research data. The data therefore should be FAIR (findable, accessible, interoperable and reusable), and "as open as possible but as closed as necessary" (European Commission, 2016). In order to build a comprehensive data set, it would be advantageous for OS partners to share annual reports and conduct semi-structured interviews and administer the proposed survey at least once every two years. Ideally, we envision that stakeholders will develop an OS partnership that will act as a repository for the data, curate that data, share it and revisit and update, periodically, the measures we propose here. Both the SGC and the MNI have agreed to do so; we invite and welcome other stakeholders to share their data sets should they be willing.

References

Ali-Khan, S. E., Harris, L. W., & Gold, E. R. (2017). Motivating participation in open science by examining researcher incentives. *eLife, 6*, e29319. doi:10.7554/eLife.29319

Ali-Khan, S.E., Jean, A. & Gold, E.R. (2018). Identifying the challenges in implementing open science [version 1; referees: awaiting peer review]. *MNI Open Research*, 2(5). doi: 10.12688/mniopenres.12805.1

Ali-Khan, S. E., Jean, A., MacDonald, E., & Gold, E. R. (2018). Defining Success in Open Science [version 2; referees: 2 approved]. *MNI Open Research, 2*(2). doi:10.12688/mniopenres.12780.2

American Academy of Arts & Sciences. (2017). The Public Face of Science. Reference Source

Australian National Data Service. (2017). *ANDS Guide: Institutional policies and procedures*. Reference Source

Begley, C. G., & Ellis, L. M. (2012). Raise standards for preclinical cancer research. *Nature, 483*, 531. doi:10.1038/483531a

Cain, D. M., Loewenstein, G., & Moore, D. A. (2005). The Dirt on Coming Clean: Perverse Effects of Disclosing Conflicts of Interest. *The Journal of Legal Studies, 34*(1), 1-25. doi:10.1086/426699

Canadian Academies of Health Sciences. (2009). *Making an Impact: A Preferred Framework and Indicators to Measure Returns on Investment in Health Research*. Reference Source

Cooksy, L. J., Gill, P., & Kelly, P. A. (2001). The program logic model as an integrative framework for a multimethod evaluation. *Evaluation and Program Planning, 24*(2), 119-128. doi:10.1016/S0149-7189(01)00003-9

Cornell University, INSEAD, & WIPO. (2017). The Global Innovation Index 2017: Innovation Feeding the World. Reference Source

Council of Canadian Academies. (2013). *Innovation Impacts: measurement and assessment*. Reference Source

Dai, Q., Shin, E., & Smith, C. (2018). Open and inclusive collaboration in science. doi:10.1787/2dbff737-en

DiMasi, J. A., Grabowski, H. G., & Hansen, R. W. (2016). Innovation in the pharmaceutical industry: New estimates of R&D costs. *Journal of Health Economics, 47*, 20-33. doi:10.1016/j.jhealeco.2016.01.012

Dolgin, E. (2014). Patent-free pact pushes the boundaries of precompetitive research. *Nature Medicine, 20*, 564. doi:10.1038/nm0614-564

Edwards, A. M., Bountra, C., Kerr, D. J., & Willson, T. M. (2009). Open access chemical and clinical probes to support drug discovery. *Nat Chem Biol, 5*(7), 436-440. doi:10.1038/nchembio0709-436

Elahi, A. (2008). Challenges of data collection in developing countries - the Pakistani experience as a way forward. *Statistical Journal of the IAOS, 25*(1, 2), 11-17.

European Commission. (2016). *H2020 Programme: Guidelines on FAIR Data Management in Horizon 2020*. Reference Source

Fecher, B., & Friesike, S. (2014). Open Science: One Term, Five Schools of Thought. In S. Bartling & S. Friesike (Eds.), *Opening Science: The Evolving Guide on How the Internet is Changing Research, Collaboration and Scholarly Publishing* (pp. 17-47). Cham: Springer International Publishing.

Gold, E. R. (2016). Accelerating Translational Research through Open Science: The Neuro Experiment. *PLOS Biology, 14*(12), e2001259. doi:10.1371/journal.pbio.2001259

Graham, I. D., Logan, J., Harrison, M. B., Straus, S. E., Tetroe, J., Caswell, W., & Robinson, N. (2006). Lost in knowledge translation: Time for a map? *Journal of Continuing Education in the Health Professions, 26*(1), 13-24. doi:10.1002/chp.47

Hardwicke, T., & Ioannidis, J. P. A. (2018). *Mapping the Universe of Registered Reports*. Berkeley Initiative for Transparency in the Social Sciences. BITSS Preprints. Reference Source

Kemp, R., & Prasad, V. (2017). Surrogate endpoints in oncology: when are they acceptable for regulatory and clinical decisions, and are they currently overused? *BMC Medicine, 15*(1), 134. doi:10.1186/s12916-017-0902-9

Kidwell, M. C., Lazarević, L. B., Baranski, E., Hardwicke, T. E., Piechowski, S., Falkenberg, L.-S., . . . Nosek, B. A. (2016). Badges to Acknowledge Open Practices: A Simple, Low-Cost, Effective Method for Increasing Transparency. *PLOS Biology, 14*(5), e1002456. doi:10.1371/journal.pbio.1002456

Kim, J., Anderson, T., & Daim, T. (2008). Assessing University Technology Transfer: A Measure of Efficiency Patterns. *International Journal of Innovation and Technology Management, 5*(4), 495-526.

Kothari, A., MacLean, L., Edwards, N., & Hobbs, A. (2011). Indicators at the interface: managing policymaker-researcher collaboration. *Knowledge Management Research & Practice, 9*(3), 203-214. doi:10.1057/kmrp.2011.16

Lampert, D., Lindorfer, M., Prem, E., Irran, J., & Serrano Sanz, F. (2017). New Indicators for Open Science: Possible Ways of Measuring the Uptake and Impact of Open Science. *Journal for Research and Technology Policy Evaluation, 44*, 50-56.

Mairesse, J., & Mohnen, P. (2010). Using innovation surveys for econometric analysis. In B. H. Hall & N. Rosenberg (Eds.), *Handbook of the Economics of Innovation* (Vol. 2, pp. 1129-1155): United Nations University.

Morgan Jones, M., Castle-Clarke, S., Brooker, D., Nason, E., Huzair, F., & Chataway, J. (2014). The Structural Genomics Consortium: A Knowledge Platform for Drug Discovery: A Summary. *Rand Health Q, 4*(3), 19.

Munos, B. (2009). Lessons from 60 years of pharmaceutical innovation. *Nat Rev Drug Discov, 8*(12), 959-968. doi:10.1038/nrd2961

National Academies of Sciences, E., & Medicine. (2018). *Open Science by Design: Realizing a Vision for 21st Century Research*. Washington, DC: The National Academies Press.

Nicol, D. (2008). Strategies for dissemination of university knowledge. *Health Law Journal, 16*, 207-235.

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., . . . Yarkoni, T. (2015). Promoting an open research culture. *Science, 348*(6242), 1422-1425. doi:10.1126/science.aab2374

Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific Utopia:II. Restructuring Incentives and Practices to Promote Truth Over Publishability. *Perspectives on Psychological Science, 7*(6), 615-631. doi:10.1177/1745691612459058

OECD. (2010). Measuring Innovation: A New Perspective, OECD Publishing, Paris. https://doi.org/10.1787/9789264059474-en

OECD. (2017). OECD Science, Technology and Industry Scoreboard 2017. doi:10.1787/9789264268821-en

OECD/Eurostat. (2005). Oslo Manual: Guidelines for Collecting and Interpreting Innovation Data, 3rd Edition. *The Measurement of Scientific and Technological Activities*, OECD Publishing, Paris. doi.org/10.1787/9789264013100-en

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251). doi:10.1126/science.aac4716

Pammolli, F., Magazzini, L., & Riccaboni, M. (2011). The productivity crisis in pharmaceutical R&D. *Nature Reviews Drug Discovery, 10*, 428. doi:10.1038/nrd3405

Poupon, V., Seyller, A., & Rouleau, G. A. (2017). The Tanenbaum Open Science Institute: Leading a Paradigm Shift at the Montreal Neurological Institute. *Neuron, 95*(5), 1002-1006. doi:10.1016/j.neuron.2017.07.026

Puritty, C., Strickland, L. R., Alia, E., Blonder, B., Klein, E., Kohl, M. T., . . . Gerber, L. R. (2017). Without inclusion, diversity initiatives may not be enough. *Science, 357*(6356), 1101-1102. doi:10.1126/science.aai9054

Roser, M. (2018). Our World in Data. Reference Source

Ryan, C., & Frye, B. L. (2017). An Empirical Study of University Patent Activity. *NYU Journal of Intellectual Property & Entertainment Law*. doi:10.2139/ssrn.2915243

Schwartz, D. L., & Sichelman, T. M. (2017). Data Sources on Patents, Copyrights, Trademarks, and Other Intellectual Property. In P. S. Menell, D. L. Schwartz, & B. Depoorter (Eds.), *Research Handbook on the Law & Economics of Intellectual Property*: Edgar Elgar Publishing.

Smith, E. (2017). Open Science Monitoring. *RAND Europe*. Reference Source

Smith, E., Gunashekar, S., Parks, S., & Chataway, J. (2016). A framework to monitor open science trends in the EU. *OECD Blue Sky III Forum, Informing science and innovation policies: Towards the next generation of data indicators*. Reference Source

Thorisson, G. A., & Stein, L. D. (2003). The SNP Consortium website: past, present and future. *Nucleic Acids Research, 31*(1), 124-127.

Tracz, V., & Lawrence, R. (2016). Towards an open science publishing platform. *F1000Research, 5*, 130. doi:10.12688/f1000research.7968.1

Treasury Board of Canada Secretariat. (2012). *Theory-Based Approaches to Evaluation: Concepts and Practices*. Government of Canada. Reference Source

Tripp, S., & Grueber, M. (2011). Economic Impact of the Human Genome Project. *Battelle Memorial Institute, 58*, 1-58.

Tuomi, L. (2016). The accelerator of open science: The impact of the Finnish Open Science and Research Initiative (ATT). *Profitmakers Ltd.*, 4-59.

Valantine, H., & Collins, F. (2015). National Institutes of Health addresses the science of diversity. *Proceedings of the National Academy of Science of the United States of America, 112*(40), 12240-12242.

Weiss, C. H. (1997). How Can Theory-Based Evaluation Make Greater Headway? *Evaluation Review, 21*(4), 501-524. doi:10.1177/0193841x9702100405

Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., Hill, S., . . . Johnson, B. (2015). *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*. Reference Source

Supplementary File 1: Measurement Toolkit

Foreword

This document sets out the measurement toolkit developed in *An Open Toolkit for Tracking Open Science Partnership Implementation and Impact* in order to build a data resource through which to study and, with that knowledge, build assessment tools for open science collaborations. We recommend that partnerships complete and share the results of the Annual Report (Part A) on a periodic basis, which we suggest being once per year. A group independent from the collaboration's management – to ensure confidentiality of results – ought to administer the semi-structured interviews (Part B) to a representative sample of stakeholders each period. We suggest that the collaboration ought to administer the survey (Part C) at the beginning of the collaboration and periodically thereafter. Finally, we suggest either the collaboration's administration or an independent group ought to develop the measures in Part D during the same period as for the annual report and after having been given access to the results of the annual report.

We envision that this toolkit be implemented through information technology, rather than through manual data entry, with standard nomenclature (e.g., as to departments and institution names). Two OS organizations, the Structural Genomics Consortium and the Montreal Neurological Institute have agreed to draw upon the toolkit to collect and share data.

A. Open Science Collaboration Annual Report

Section One: Name of Collaboration

1. List the principal academic, community, industrial and governmental partners of the collaboration for the reporting period. For each partner, provide the following details:
   1.1. The organizational identifiers; and
   1.2. The sector (academic, government, industrial, philanthropic, community, etc.).

Section Two: Project Outputs

2. List all projects falling within the collaboration for the reporting period. For each project, specify the following:
   2.1. Whether the project is new, ongoing or closed;
   2.2. Whether a project plan exists. If a project plan does exist, include the project plan as an appendix to the annual report or, if it is public, provide its persistent identifier (e.g. DOI);
   2.3. Whether the project was born open, became open during the project's process, became open upon the project's completion, became open after embargo, switched from open to close, or was never open (with open being understood as available to all who desire access with minimal restrictions, e.g. clickwrap agreement); and
3. For each project listed in (2), indicate whether the project includes each of the following:
   3.1. Open governance that is available through online strategic and organizational meetings, open minutes, and transparent governance rules;
   3.2. Design processes to create, revise, and comment on projects that are openly available;
   3.3. Project proposals that are openly available;
   3.4. Project and collaboration budgets that are openly available;
   3.5. Output management plans that are openly available;
   3.6. Materials generated through the project that are openly shared to all that ask, except where there is a limited supply of physical materials;
   3.7. Outputs generated by the project that are openly available without further restriction on use, except in well-defined and publicly justified cases, e.g. to protect the privacy of patient or donor information, or the precise location of nesting sites of rare species;
   3.8. Open infrastructure through which one can access and comment on outputs, etc.
   3.9. Review of projects that is openly available;
   3.10. Clear, open and transparent research processes, such as open lab books, open research meetings, etc. that are openly available;
   3.11. Preregistration of data collection initiatives that is openly available;
   3.12. Ethics reviews and reasoning that are openly available; and
   3.13. (For closed projects or closed aspects of open projects) Provides rationale for why they are closed using a controlled vocabulary in addition to or instead of details.
4. List all publications, including preprints and outreach materials, arising out of the collaboration during the reporting period. For each publication, provide the following details:
   4.1. Persistent identifier if available, such as DOI;
   4.2. Full citation including authors, title, journal, source, etc.;
   4.3. Accessibility;
   4.4. Whether the journal in which the article is published conforms to TOP guidelines;
   4.5. From which project this publication results; and
   4.6. The standard for machine readability to which the document conforms (e.g., JATS)
5. List all data sets arising from the collaboration in the reporting period. Provide the following information:

5.1. Persistent identifier if available;
5.2. Full citation;
5.3. Accessibility; and
5.4. The standard for machine readability to which the document conforms (e.g., JATS)

6. List any project in the reporting period from question (2) which did not yet result in a publication or in a published data set listed in questions (4) or (5).

Section Three: Measure of Scale

7. List all external awards, prizes and grants that recognize or directly support OS that were awarded or granted to researchers in the collaboration during the reporting period. For each of these awards, prizes or grants, provide the following details:
   7.1. Persistent identifier if available;
   7.2. Title of award, prize or grant;
   7.3. Nature of award (award, grant, prize, etc.);
   7.4. Name of awardee, with number of years following the awardee's highest degree from the date of convocation of the degree, and name of degree;
   7.5. Organization providing award, prize or grant;
   7.6. Nature of that organization (government agency, industrial, philanthropic, etc.);
   7.7. Period covered by the award, prize or grant; and
   7.8. Value of the award, prize or grant.

Section Four: Quality of Outputs

8. List all retractions arising out of the collaboration during the reporting period. For each type of output (publications, data), provide the following:
   8.1. The total number of outputs;
   8.2. The total number of retractions; and
   8.3. The summary statistics of the reasons for these retractions, using a controlled vocabulary.
9. List of all corrections arising out of the collaboration during the reporting period. For each type of output (publications, data), provide the following:
   9.1. The total number of outputs;
   9.2. The total number of corrections; and
   9.3. The summary statistics of the reasons for these corrections, using a controlled vocabulary.

Section Five: Diversity and Youth Engagement

10. Calculate the number of projects listed in (2) that have at least one non-academic (i.e., not hired to conduct research at a public research organization) stakeholder. Calculate the percentage of projects that include at least one non-academic stakeholder out of all projects.
11. List all early career researchers (ECRs), i.e. PhDs candidates, postdocs, and individuals who have received a PhD within the past 5 years, including postdoctoral fellows, who worked with the collaboration over the preceding five years. For each, provide the following details:
    11.1. Period during which the ECRs worked in the collaboration;
    11.2. For each ECR who has left the collaboration, whether the ECR had ever worked or interned outside an academic environment (e.g., in industry, government or civil society) after leaving the collaboration.

Section Six: Efficiency of Outputs

12. List all legal instruments (e.g. contracts and memoranda of understanding) entered into or renewed in respect to the collaboration during the reporting period. For each contract, provide the following information:
    12.1. Persistent identifier if available;
    12.2. Type of instrument (Material Transfer Agreements, research, sponsorship, etc.);
    12.3. Whether the instrument is new or is a renewal;
    12.4. Number of days from the time that the initial instrument negotiations began (ex: request for contract initiated or request to renew) to the execution of the instrument;
    12.5. For each Material Transfer Agreement, the number of days from initial contact to actual transfer of materials; and
    12.6. Whether and to what extent the instrument is open (no claim to intellectual property rights, levels of commitment to open data and open publication, ability to re-share the materials under the same conditions).
13. For each type of contract, calculate the percentage of those contracts that are open.
14. List all new, ongoing or terminated start-ups and spin-outs arising from the collaboration in the reporting period. For each, provide the following details:
    14.1. Name of firm;
    14.2. Location (city) of the firm's head office and locations (cities) of the firm's other offices;
    14.3. Status of the firm (new, ongoing, or terminated);
    14.4. Relationship between the firm and the partnership (owned by one or more partners, owned by a researcher within the collaboration, etc.)
    14.5. Number of FTEs employed by the firm at year end;
    14.6. A description of the field of operation of the firm;
    14.7. Whether the firm is for profit, not-for-profit, or charitable.
15. For terminated start-ups or spin-outs listed in (14):
    15.1. For each of them, calculate the number of months from incorporation to termination.
    15.2. Calculate the average number of months that the terminated firms survived.

Section Seven: Extended Reach

16. For each item listed in (4) publications – or (5) data – that resulted in a first citation within the reporting period:
    16.1. Calculate the number of months between publication and first citation; and
    16.2. Calculate the average number of months from earliest publication to the first citation for both open access publications and all other publications.
17. List all current financial or in-kind contributions to the collaboration by industry or philanthropy during the reporting period other than those listed in (7) to the OS collaboration. For each, provide the following details:
    17.1. Persistent identifier if available;
    17.2. The grantor of the investment and the nature of the grantor (firm, foundation, etc.);
    17.3. The value of the investment, specifying cash and in-kind contributions separately; and
    17.4. The period covered by the investment (start and end date).
18. List all for-profit and non-profit firms or organizations that actively partnered with the collaboration during the reporting period. For each, provide the following:
    18.1. Persistent identifier if available;
    18.2. Name of Firm;
    18.3. Whether the firm is for-profit or non-profit;

18.4. The address of the firm's head office;
18.5. If the firm has an office in the region, indicate its address and how many employees are employed locally;
18.6. If available, the firm's annual revenues;
18.7. Field of operation; and
18.8. The contributions of the firm to the collaboration.

Section Eight: Open Science Engagement

19. List whether the project that results from collaboration during the reporting period has a policy in respect of the following:
    19.1. Sharing the research prioritization process (prioritizing certain research questions or methodologies over others);
    19.2. Sharing proposals;
    19.3. Sharing how funding is allocated;
    19.4. Transparency, openness, or inclusion on governance;
    19.5. Sharing budgets;
    19.6. Transparency, openness, or inclusion on research design;
    19.7. Transparency, openness, or inclusion on execution of the research;
    19.8. TOP Guidelines 2 (Data transparency), 3 (Analytic methods (code) transparency), 4 (Research materials transparency, and 5 (Design and analysis transparency);
    19.9. Open access;
    19.10. Openness of peer review;
    19.11. Openness of how ethics are applied in research decision-making; and/or
    19.12. Openness of rationale for exceptions to open behaviors.
20. Indicate whether the collaboration has a policy of non-open and non-standard, non-open and standard, open and non-standard, or open and standard licensing.
21. Indicate whether the collaboration:
    21.1. Has no data preservation, some preservation or a preservation policy;
    21.2. Dedicates resources for long-term preservation of data; and
    21.3. Has its data stored in certified repositories.
22. Select all that describes the collaboration's level of participation:
    22.1. Closed to observation;
    22.2. Observable by invitation (please note whether the invitations issued were public or private);
    22.3. Observable by anyone;
    22.4. Closed to contribution;
    22.5. Contribution by invitation;
    22.6. Contribution by anyone (please note whether the contributions are or can be anonymous or identified);
    22.7. Allows for passive engagement; and/or
    22.8. Allows for active engagement.
23. Indicate whether the collaboration provides training on OS to the following:
    23.1. Undergraduate students
    23.2. Graduate students
    23.3. Postdoctoral fellows
    23.4. Continuing professional development for faculty and staff, e.g. clinicians, full-time researchers, research administrators, librarians, legal counsels.

B. Semi-Structured Interview Guides

Description

This is a semi-structured interview guide that is meant to be administered annually by open science (OS) collaborations. The purpose of the interview guide is to gather substantive qualitative measures of the benefits and costs of OS. The guide is designed to include a wide set of OS stakeholders, including full-time academic staff, early career researchers, individuals from the private sector, research participants, and ethics review board members and/or administrators. The interview results will be used for a variety of purposes including, at an aggregated level, to assess the OS partnership, to study OS partnerships in general, to assess quantitative measures of OS impact and so on.

General Instructions about Consent and Meeting Research Ethics Requirements

Please ensure that, in addition to obtaining consent for use of the raw data by those administering the survey and sharing anonymized or aggregated data generally, that the raw data can be shared with other groups who are operating under a similar protocol and who have obtained ethics approval, even if these other groups are in a different jurisdiction. Also ensure that the nature of the ethics approval and the process that led to it is as openly documented as possible.

General Questions for All Stakeholders

1. Definition
   - 1.1 What does open science (OS) mean to you?
   - 1.2 What is the minimal level of openness (e.g., open data, open access publications, avoidance of restrictive intellectual property rights, open grants and reviews, etc.)

2. Transparency of Research Output
   - 2.1. Does the OS partnership provide you with information that is useful to your organization and members in a timely and accessible manner? Please give examples, if any, of successful information sharing.
   - 2.2. What can the OS partnership do to improve knowledge sharing internally and externally?
   - 2.3. Is the OS partnership structured so that you and your organization can provide input on your information needs, information research questions, priorities, etc.? If not, why not? If yes, how does the OS partnership achieve this?

3. Public Appreciation and Understanding of Research
   - 3.1. Do the collaboration's partners have a plan to enable public understanding of the research being conducted and of the results?
   - 3.2. If so, what do you believe to be the effectiveness of this plan?
   - 3.3. How would you improve this plan?

4. Institutional Attitude to Transparency
   - 4.1. Please describe your perspective on retractions or scientific publications or data sets. Are retractions a sign that the system is working or not working? Should we aim to eliminate or at least reduce retractions?
   - 4.2. Has the uptake of OS and greater openness in the research process had an effect on the way you think about or handle research errors or retractions? If so, in what ways?
   - 4.3. Has OS contributed to greater transparency in the research and innovation process? If so, please describe how so. If not, also please describe why not.
   - 4.4. Do you have any examples?

4.5. Do you find any changes in the way your peers view retractions and errors? What do you believe is the influence of OS on these views?
4.6. Do you believe that research transparency has a positive or negative effect on public trust in the scientific-research endeavor? Please explain.

5. Institutional Support for Staff Engaged in OS
   5.1. Do you engage in OS practice?
   5.2. If so, do you feel that your institution encourages and supports your efforts to engage in OS practice? In which ways do you feel supported?
   5.3. What could the institution do better to support your engagement in OS practice? Specifically, does your institution's tenure and promotion policies encourage this engagement?

6. Validation of Quantitative Measures

The following question would be posed to interviewees after providing the results of the collaboration's annual quantitative audit.

   6.1. Do the annual results of the quantitative data collected by the collaboration (such as institutional H-index, publication counts, number of open datasets, patent counts) accurately reflect the research impact of the collaboration's work? If so, in which ways? If not, what is missing or inaccurate?

7. Awareness of OS within your Institution.
   7.1. Have you heard of OS? If so, what does it mean to you?
   7.2. As for as you know, does your institution practice OS? If so, in which ways?
   7.3. How did you hear about OS at your institution?

Questions for Early Career Researchers

(PhDs Candidates, Postdocs, and Individuals Who Have Received a PhD within the Past 5 Years)

8. Attitudes of Early Career Researchers to OS
   8.1. Do you practice OS? If so, in which ways do you practice it (consider open grants, open peer review, open budgets, open access publications, open data sets, open laboratory books, open materials exchange, open reagents, etc.)?
   8.2. If you practice OS, what motivated you to do so? How motivated are you: slightly, moderately, or significantly? What demotivates you from practicing OS?
   8.3. Which factors are most important to you when assessing potential employers? How important is the employer's adherence to OS principles in assessing these factors?

9. New Pathways for Young Investigators
   9.1. Do you feel supported in your career by the institution you work for/ are affiliated with? If so, please describe reasons for this. If not, also please describe why not.
   9.2. Has your institution developed novel pathways to help you succeed in an OS environment? If so, in what ways?
   9.3. What additional ways could your institution help you succeed?
   9.4. Has the growing adoption of OS practice had a positive or negative effect on your attitude towards your research? Why?

10. Skill Diversity of ECRs Working in OS
   10.1. Over the course of your graduate studies, to what extent did you practice OS in your research? Do you have any examples of your engagement with OS from that period?

10.2. To what extent do you feel that your experience in practicing OS gave you any of the following skills: increased empathy, more varied data analysis skills, greater understanding of other's perspectives, greater ability to be a lateral thinker, better data curation skills, more transparent research processes, and collaboration skills? How so?

Questions for Individuals from the Private Sector

11. Growth of Business Models that Use and Support OS

11.1. Does your business draw on any OS outputs? If so, which ones? Please describe the process by which you accessed these outputs.

11.2. What proportion of your activities are based on OS? How important are these activities to your firm's success?

11.3. What challenges and opportunities does OS present for your business? These may include reliance on open access publications, open data sets, product development, identification of markets, identification of partners, quality control, etc.

11.4. Please describe your business model in respect of your OS activities.

11.5. How "open" is your business model? How, if at all, do you protect intellectual property?

11.6. Have you invested (time/money/in-kind/know-how) in an OS initiative?

Questions for Research Participants

12. Conditions that Contribute to Trust

OS collaborations are partnerships between different institutions, whether between institutions in the public sector or between institutions in each the public and private sectors, aiming at sharing knowledge and ideas without restrictive rights.

12.1. Have you heard about OS?

12.2. What do you know of OS? How would you define OS?

12.3. How did you hear about this? Do you feel that you are sufficiently informed about OS?

12.4. How are you involved in OS?

12.5. Do you feel that you derive benefit from your participation in the OS collaboration? If so, in which ways? These may include greater understanding of your contribution, greater knowledge to guide your own activities, financial or other tangible reward, greater networking opportunities, etc.

Questions for Ethics Review Board Members and/or Administrators

13. Ethics Committee Preparedness

13.1. Have you encountered OS in the context of your ethics committee work? If so, how did OS come up?

13.2. What issues, challenges or opportunities has the ethics committee encountered in handling applications that involve OS?

13.3. To what degree, if any, has OS had an impact on the way you approach project evaluation? In which ways? Do you see this impact as constructive and beneficial or otherwise? Please explain why.

13.4. Do you believe that your participation in evaluating research ethics applications arising from OS collaborations has altered the way you evaluate ethical concerns? If so, in which ways?

13.5. In your view, does the increase in OS practices necessitate any changes in the way you conduct ethics reviews? If so, how?

13.6. Do members of ethics committees need greater training on OS? If so, on what topics and in which ways?

C. Survey for Measurement of Open Science Engagement

Description

Open science (OS) collaborations aim to reduce transactions costs, increase sharing, and build better connections with communities. This survey is designed to identify best practices for these collaborations and to assess the ways in which the collaboration is open.

General Instructions for Selecting Survey Participants

Administer to a representative sample of individuals at stakeholder organizations within the collaboration.

Beneficial Elements

1. Do you believe these things are beneficial? Click all that apply.

		Always	Partly	Never
Open Research Grant Application				
1.1.	Open research proposals	☐	☐	☐
1.2.	Open reviews of research proposals	☐	☐	☐
1.3.	Open funding decisions and funding allocations	☐	☐	☐
Open Methodology				
1.4.	Open governance of projects through online meetings, open minutes, and transparent governance rules	☐	☐	☐
1.5.	Project and collaboration budgets available online	☐	☐	☐
1.6.	Open design processes to create, revise, and comment on projects	☐	☐	☐
1.7.	Clear, open and transparent research processes, such as open lab books, open research meetings, etc.	☐	☐	☐
1.8.	Preregistration of data collection initiatives	☐	☐	☐
1.9.	Open output management plans	☐	☐	☐
1.10.	Availability and use of open infrastructure through which to access and comment on outputs, etc.	☐	☐	☐
Open Outcomes				
1.11.	Materials generated by the collaboration are openly shared to all that ask, except where there is a limited supply of materials	☐	☐	☐
1.12.	Where materials are in limited supply, the existence of a clear set of criteria and open governance structure to decide to whom to send materials	☐	☐	☐

1.13.	Outputs generated by the collaboration are openly available without further restriction on use, except to protect the privacy of patient or donor information	☐	☐	☐
1.14.	Outputs, including materials, are subject to open annotations	☐	☐	☐
1.15.	Publications are open access, with open license, open citations and machine actionable full text	☐	☐	☐
1.16.	The outcomes of the collaboration are not subject to intellectual property rights that restrict free and open use and reuse	☐	☐	☐
1.17.	All tools and software are openly accessible and reusable	☐	☐	☐
1.18.	Reporting standards are openly shared	☐	☐	☐
1.19.	Review of projects and of the collaboration are openly available	☐	☐	☐
1.20.	Ethics reviews and reasoning are openly available	☐	☐	☐
1.21.	Any exceptions to openness are transparently and openly shared	☐	☐	☐

Your Own Activities

2. Do you intend to engage in the following activities because they are relevant to you or your role? Click all that apply.

		Always	Partly	Never
Open Application				
2.1.	Open research proposals	☐	☐	☐
2.2.	Open reviews of research proposals	☐	☐	☐
2.3.	Open funding decisions and funding allocation	☐	☐	☐
Open Methodology				
2.4.	Open governance of projects through online meetings, open minutes, transparent governance rules	☐	☐	☐
2.5.	Project and collaboration budgets available online	☐	☐	☐
2.6.	Open design processes to create, revise, and comment on projects	☐	☐	☐
2.7.	Clear open, and transparent research processes, such as open lab books, open research meetings, etc.	☐	☐	☐
2.8.	Preregistration of data collection initiatives	☐	☐	☐
2.9.	Open output management plan	☐	☐	☐

2.10.	Availability and use of open infrastructure through which to access and comment on outputs, etc.	☐	☐	☐
Open Outcomes				
2.11.	Materials generated by the collaboration are openly shared to all that ask, except where there is a limited supply of materials	☐	☐	☐
2.12.	Where materials are in limited supply, the existence of a clear set of criteria and open governance structure to decide to whom to send materials	☐	☐	☐
2.13.	Outputs generated by the collaboration are openly available without further restriction on use, except to protect the privacy of patient or donor information	☐	☐	☐
2.14.	Outputs, including materials, are subject to open annotations	☐	☐	☐
2.15.	Publications are open access, with open license, open citations and machine actionable full text	☐	☐	☐
2.16.	The outcomes of the collaboration are not subject to intellectual property rights that restrict free and open use and reuse	☐	☐	☐
2.17.	All tools and software are openly accessible and reusable	☐	☐	☐
2.18.	Reporting standards are openly shared	☐	☐	☐
2.19.	Review of projects and of the collaboration are openly available	☐	☐	☐
2.20.	Ethics reviews and reasoning are openly available	☐	☐	☐
2.21.	Any exceptions to openness are transparently and openly shared	☐	☐	☐

Open Practice

3. Do you believe that the OS collaboration to which this questionnaire refers carries through on the following elements? Click all that apply.

		Always	Partly	Never
Open Application				
3.1.	Open research proposals	☐	☐	☐
3.2.	Open reviews of research proposals	☐	☐	☐
3.3.	Open funding decisions and funding allocation	☐	☐	☐
Open Methodology				

3.4.	Open governance of projects through online meetings, open minutes, and transparent governance rules	☐	☐	☐
3.5.	Project and collaboration budgets available online	☐	☐	☐
3.6.	Open design processes to create, revise, and comment on projects	☐	☐	☐
3.7.	Clear, open and transparent research processes, such as open lab books, open research meetings, etc.	☐	☐	☐
3.8.	Preregistration of data collection initiatives	☐	☐	☐
3.9.	Open output management plan	☐	☐	☐
3.10.	Availability and use of open infrastructure through which to access and comment on outputs, etc.	☐	☐	☐
Open Outcomes				
3.11.	Materials generated by the collaboration are openly shared to all that ask except where there is a limited supply of materials	☐	☐	☐
3.12.	Where materials are in limited supply, the existence of a clear set of criteria and open governance structure to decide to whom to send materials	☐	☐	☐
3.13.	Output generated by the collaboration are openly available without further restriction on use, except to protect the privacy of patient or donor information	☐	☐	☐
3.14.	Outputs, including materials, are subject to open annotations	☐	☐	☐
3.15.	Publications are open access, with open license, open citations and machine actionable full text	☐	☐	☐
3.16.	The outcomes of the collaboration are not subject to intellectual property rights that restricts free and open use and reuse	☐	☐	☐
3.17.	All tools and software are openly accessible and reusable	☐	☐	☐
3.18.	Reporting standards are openly shared	☐	☐	☐
3.19.	Review of projects and of the Collaboration are openly available	☐	☐	☐
3.20.	Ethics reviews and reasoning are openly available	☐	☐	☐
3.21.	Any exceptions to openness are transparently and openly shared	☐	☐	☐

D. Additional Measures of Open Science

We list here measures that require some analysis, such as identifying the citations (including in patents) to outputs. The list that follows requires, as explained below, expansion.

Patent Citation

1. Citation intensity: Citation intensity weighted by patent family and normalised by research discipline, or technology sector. Citation intensity means the number of third party patents citing artifacts (academic publications, other publications, blogs, grant applications, laboratory books, data sets, materials, policies) derived from the OS collaboration. Citation intensity is a granular measure and can be assessed at the individual researcher level, department or institutional levels and at a different time periods. The Lens.org provides the In4M tool to calculate this number.
2. Patent Citations to Literature: The number of open access publications and data sets referenced within patents in the reporting period. This can be calculated as the percentage of all artifacts to date arising from the OS collaboration which are cited in patent literature. An alternative measure is the ratio between the average number of citations in patent documents during the reporting period in patents to the collaboration's artifacts.

Note: While only including measures on patents, it would be useful to develop indicators similar to the ones above for policy documents.

Community and Diversity

3. Equity of Knowledge Production: The percentage of funds and in-kind support made available within the OS collaboration to researchers, firms or communities in non-high-income countries with respect to overall funds. An alternative measure could include comparing how OS and non-OS projects involve marginalized groups within the research process.
4. Community Engagement: Analysis of project documentations to track the collaboration's community engagement and extent of communication and benefit-sharing with communities. Code 0 if there is no community engagement plan; code 1 if the project plan describes a community engagement plan; and code 2 if the project reports indicate the plan is being followed.

Supplementary File 2: Open Science Measures to Be Considered by Others

We set out here measures that participants in the workshops identified that may be useful to other actors in constructing their own measures.

Measures to Be Implemented by Research Funders

1. Existence of Output Management Plans: The funder should report on the following policies:
   1.1. Does the funding require plans for output sharing and management that are open access;
   1.2. Does the funder assess the degree to which those being funded follow those plans?
   1.3. If so, report on the level of compliance as a percentage of all those being funded.
   1.4. Does the funding organization favor open science projects in their grant applications?

Measures for Governments or Intergovernmental Agencies

2. Employment in STEM fields: Number of employees employed at spin-offs of OS collaborations in the region, country, or internationally during the reporting period (one year). This could be the sum of the employment reported in the annual reports or could be based on surveys of SMEs. Provide a breakdown of employee type, such as executive, scientific directors, database owners, PhD scientist, masters, and undergraduate research and technical staff, administration, marketing, and finance.

Supplementary File 3: Incomplete and Rejected Open Science Measures

1. Diversity in research subjects: The researcher should report on the expected population of research subjects related to research conducted in the OS partnership for data collection. The expected population should be compared to the sample in the dataset.[3]
2. Diversity in research subjects: The researcher should report whether the OS partnership or the researcher's institution includes a diversity statement or language pertaining to diversity in a code of conduct.[4]
3. Diversity of authors: Diversity of authors' disciplines within a paper. This can be calculated as the number of authors from different disciplines within a single paper.[1]
4. Duplication in clinics: The quantity of duplicated studies of the same trial/data in a same institution/clinic.[2,3]
5. Linking of datasets to international data repositories: Tracking the international collaborations between datasets and international data-warehouses. Code 1 if the data is included in a collaboration with an international data-warehouse; code 0 if the data is not included in a collaboration with an international data-warehouse.[1]
6. Diversity of users/readers: Diversity of users/readers of a paper. This can be calculated by looking at country of origin of users' login/registration data. An alternative measure would be to look at the country of origin of papers citing the paper at hand.[3]
7. Open Data Use: Tracking views and downloads of OS data.[1,3]
8. Funding for secondary use of data/experiments reproduction: The total money given by funding agencies for secondary use of data or for experiment reproduction.[3]
9. Tracking Open Access Patterns: The percentage of publications, year-over-year, that are published in open-access forums. While the Harvard OATP tracking project is one potential source for this information, work is still needed to determine further data sources and methods of data scrubbing so that this effort could be conducted on a consistent basis.[2]
10. Machine readable data agreements: The percentage of the data that is scrubbable by machines without the need for human checks.[2]
11. Use in Metadata: Measure that asks whether the data was aggregated in meta-analysis. Code 1 for OS data that was used or added to an aggregate data; code 0 for OS data that was not used or added to an aggregate data. Bibliography scrubs (publication analysis) provide the tool to collect this data.[6]
12. Diversity of contribution and meaning data: The percentage of data made open that is FAIR compliant. First pass statement or data analysis would serve as tools to collect this data.[2,3]
13. Traceability of data: Measure that asks whether the providence of the data is traceable and whether this data has first been pass-vetted by an institution (like TOSI, for example). Code 1 for data that has been vetted by an institution; code 0 for data that has not been vetted.[2]
14. Knowledge product references in policy: Percentage of knowledge products coming from OS outputs in assessed OS collaborations that are referenced in policy documents. This would involve tracking informed policy decision making that is made using citations of literature and evaluate what percentage of that literature is available as open access articles. Code 1 for policy documents that cites an open access article; code 0 for policy documents that does not cite an open access article. Policy documents and Altmetrics would provide the tool to collect this data.[2,3]
15. Entropy in grant: Shannon entropy of successful grants.[3]
16. Recognition of OS practices in academic career incentives for Early Career Researchers (ECRs): The extent to which adherence to OS practices is associated with career success. This can be calculated by the percent of individuals' research which are in a FAIR format that is associated with funding success,

securing tenure, or awards. Code 1 for research that is FAIR compliant and receives funding success, tenure, or awards; code 0 for research that is FAIR compliant but does not receive success, tenure, or awards. FAIR status of works and Uber Dimensions would serve as tools to collect this data.[1,2]

17. Students who intern or work in non-academic roles: The number of students that end up working in non-academic professions in the industry. Code 1 for a student that ends up working outside of academia in the industry; code 0 for a student that remains within academia. University information would serve as tools to collect this data.[5]
18. OS increase in youth engagement with research: The number of students using OS platforms to participate in and learn about research. UNESCO data on youth in science would serves as a tool to collect this data.[3]
19. DRMs requiring Open Data that are being used for secondary research: Percentage of data sets where DRMs require open data to be used for secondary research. DRM and DOI citations would serve as tools to collect this data.[5]
20. Number of materials cited in funding application: Percentage of open artefacts compared to closed artifacts. Granting funding applications would serve as a tool to collect this data. It is not clear how this relates to specific OS collaborations, but the measure could be used by national funding organizations to track OS uptake.[3]
21. Rejected patents based on duplication (non-novelty): Percentage of all patent applications rejected for novelty or non-obviousness out of all the patents filed in a given year. Patent registries would serve as a tool to collect this data.[2,3]
22. Student destinations: The location of the firms that are hiring students. University-level tracking would serve as a tool to collect this data.[3]
23. Economic value of OS researchers: A set of economic measures that researchers associate with "highly open" research. This includes using data to identify highly valuable patents citing researchers' work, start-ups founded by the individual, and salary of the individual. Publication datasets, patent view, Lens.org, and university records would serve as tools to collect this data.[1,2,3]
24. OS community measure set: The proportion of scientists with an "open index" above some threshold. Code 1 if the scientist is above the threshold; code 0 if the scientist is below the threshold.[3]
25. Success rates at key funding decision points for ECR/HQP who practice OS: A comparison of success rates in funding awards of ECRs who practice OS to those who do not. This can be tracked over time to examine how practicing OS has an impact on important career rewards. Self-reported surveys assessing researchers' OS practice, as well as information from funding bodies, tenure committees, and prize panels would serve as tools to collect this data.[3]
26. Success rates at career advances decision point for ECR/HQP who practice OS: A comparison of success rates in career advancement of ECRs who practice OS to those who do not. Success rates are to be tracked over time to examine how practicing OS impacts important career rewards. Information on applications and awards from funding bodies, tenure committees, and prize panels would serve as tools to collect this data.[3]
27. Public opinion of OS initiatives: Qualitative analysis of public sentiment of popular press coverage and social media coverage related to OS initiatives. Social media sites, internet archive sites, and interviews with project facilitators/managers regarding trust would serve as tools to collect this data.[3]
28. OS increases youth engagement with research: Qualitative analysis of the extent to which OS enables and increase the number of youth who are engaged with research principles. This can be calculated by performing a longitudinal study that would track a group of students from various countries and how they interact with science in a closed environment or open environment. Leverage data from UNESCO and the OECD, both of which collect education data that could serve as tools to collect this data.[1]

29. Changes in the practice of research to be more open (proportion of ECRs involved in OS initiatives, as well as their role, and influence in the project): Qualitative analysis of cultural change in the practice of research that is attributable to ECRs with OS experience. This can be calculated by looking at the percentage of public involvement in research designed and run by ECRs compared to total research.[2]
30. Number of start-ups arising from individual OS collaborations: A count, at the regional level, of start-ups emerging from OS collaborations with industry. This can be calculated by counting, on an annual basis, all new start-ups emerging from individual collaborations on OS partnerships. Universities, regional start-up communities, and industry start-up resources would serve as tools to collect this data.[5]
31. OS patent measure set: The value of patented products and services built on and by OS. Kogan et al's (2017) article as well as Lens.org and PatentsView.org would serve as tools to collect this data.[3]
32. Reproducibility and data re-use: The percentage of data/research emanating from the OS collaboration that it is suitable for re-use or is reproducible. This can be calculated by tracking the proportion of publications which appears in outlets that have adopted and implement the TOPS guidelines. Code 1 if a published output that is published in an outlet that has adopted the TOPS guidelines; code 0 if a published output that is not published in an outlet that has adopted the TOPS guidelines.[3]
33. Nature of transaction with industry: Qualitative analysis of the types of interactions (open vs closed) with industry. This can be calculated by surveying the transactions and categorizing the resources values (in kind vs monetary).[3]
34. Increased attractiveness of firms to the local area: The percentage of firms partnering with OS researchers in local areas. Standard bibliometrics and news releases would serve as tools to collect this data.[5]
35. Increased innovation originating from ECRs who are practicing OS: A comparison between ECRs working in OS and ECRs working in closed systems. One would compare the participation in patenting process and innovation articles and whether the patent/article includes an ECR as a contributor. Lens.org and In4M data would serve as tools to collect this data.[3]

---

[1] This measure was classified as incomplete as it is not fully defined.
[2] This measure was classified as incomplete as clear data sources are still required.
[3] This measure was classified as incomplete as a clear coding framework is still required.
[4] This measure was suggested by individual participant after the workshop and thus not yet vetted.
[5] This measure was classified as redundant.

Supplementary File 4: Origin of the Measurement Toolkit

A. Open Science Collaboration Annual Report

1. N/A
2. This measure originated from the Washington workshop discussions on 1) greater public trust, appreciation and understanding of science and the research process, 2) how OS will allow for a greater number of people to engage the development of research and products. This measure was also developed through the hypothesis that OS will increase trust and accountability because it will democratize the research process by creating more ways for community led research and ways for communities to participate in the research process. In London, the groups developing measures for 1) increased trust and accountability of the research enterprise, 2) accelerated innovations and impact and 3) implementation success this measure. The McGill team added this list of projects together with project plans and refined the measure by eliminating a de minimus level of involvement requirement, substituting a list of outside stakeholders that third parties could evaluate.
3. This measure originated from the Washington workshop discussion on increasing the number of researchers engaging in sharing activities, and increasing the use of open processes and tools across the entire research workflow. In London, the group on implementation success developed this measure in assessing OS engagement through transparency.
4. This measure is designed to provide a record of publications to support other measures and reporting. This measure originated in discussions at the Washington workshop around reduced rates of scientific misconduct and retraction, as well as discussions around greater use, re-use and re-combination of datasets. In London, the group developing measures on increased quality and efficiency of scientific outputs developed this measure.
5. This measure is designed to provide a record of data to support other measures and reporting.
6. This measure is designed to provide a record of publications to support other measures and reporting. This measure originated in discussions at the Washington workshop around reduced rates of scientific misconduct and retraction as well as discussions around greater use, re-use and re-combination of datasets. This measure is one of those developed by the group on increased quality and efficiency of scientific outputs in London.
7. This measure originated in discussions at the Washington workshop related to the hypothesis that OS will lead to an acceleration of discovering and identifying innovative ideas. Then, in London, the group developing measures on increased quality and efficiency of Scientific Outputs translated the initial measure into a coding of 0 [no awards], 1 [1 award], or 2 [2 or more awards], focused particularly at junior researchers. It was modified by the McGill team to list all awards and focus on all researchers, while still retaining the ability to determine which awards were directed towards junior researchers.
8. This measure originated in discussions at the Washington workshop on quality of outputs. Delegates considered two contradictory hypotheses: 1) that because of increased expected scrutiny, researchers will perform greater validation of data prior to disclosure; 2) that due to lower initial peer review of open access publications, data will be less reliable. In London, the group developing measures on increased quality and efficiency of scientific outputs developed this measure. The McGill team restricted the measure to the specific OS collaboration rather than being a measure of all OS-derived publications that were retracted out of all publications.
9. This measure originated in discussions at the Washington workshop on quality of outputs. Delegates considered two contradictory hypotheses: 1) that because of increased expected scrutiny, researchers will perform greater validation of data prior to disclosure; 2) that due to lower initial peer review of

open access publications, data will be less reliable. In London, the group developing measures on increased quality and efficiency of scientific outputs developed this measure. The McGill team restricted the measure to the specific OS collaboration rather than being a measure of all OS-derived publications that were corrected out of all publications.

10. This measure originated in discussions at the Washington workshop related to how OS will allow for a greater number of people to engage the development of research and products. In London, the group developing measures on accelerated innovations and impact identified this measure. The McGill group refined the measure by eliminating a de minimus level of involvement requirement, substituting a list of outside stakeholders that third parties could evaluate.
11. This measure originated in discussions at the Washington workshop on whether OS collaborations reduce barriers for graduate students moving between academia and industry. In London, the group developing measures on better opportunities and recognition of early researchers and youth developed this measure. The McGill team refined the coding so as to reveal less of patients' personal information and to ensure a consistent coding. The McGill team also added non-industrial positions (i.e., civil society).
12. This measure originated in discussions at the Washington workshop related to faster generation and greater translation of valuable knowledge to policy and practice. In London, the group developing measures on accelerated innovations and impact developed this initial measure. The McGill group refined the method of counting (by days) and adding the reporting of the type of contract.
13. This measure originated in discussions at the Washington workshop related to faster generation and greater translation of valuable knowledge to policy and practice. In London, the group developing measures on accelerated innovations and impact developed this initial measure.
14. This measure originated in discussions at the Washington workshop around the ease of launching start-ups. In London, the group developing measures on positive economic impact developed this measure. The McGill team added the one-year period and the restriction to the OS collaboration.
15. This measure originated in discussions at the Washington workshop around the ease of launching start-ups. In London, the group developing measures on positive economic impact developed this measure.
16. This measure originated in discussions at the Washington workshop on the contributions of OS to greater use, re-use, and re-combination of dataset. In London, the group developing measures on increased quality and efficiency of scientific outputs developed this measure. Initially, the coding was to measure whether OS publications were higher/lower than non-OS publications. The McGill team changed the coding to an average number of months to first citation for each OS and non-OS publications arising from the OS collaboration.
17. This measure originated in discussions at the Washington workshop on whether OS collaborations will lead to increased resources available to universities through access and collaborations with industry. In London, the group developing measures on positive economic impact developed an initial measure. The McGill team merged measures for philanthropy and industry, and provided details as to the reporting.
18. This measure originated in discussions at the Washington workshop on faster generation and greater translation of valuable knowledge to policy and practice. In London, the group developing measures on accelerated innovations and impact developed this measure. The McGill group changed the measure from a count to a list and set out the details.
19. This measure originated in discussions at the Washington workshop on effective comprehensive tracking of the range of scientific outputs, and greater incorporation of OS into standard research workflows. In London, the group developing measures on implementation success developed this measure to assess OS engagement through policy.

20. This measure originated in discussions at the Washington workshop on researchers increasingly engaging in sharing activities such as publishing open access. In London, the group developing measures on implementation success developed this measure of assessing OS engagement through licensing.
21. This measure originated in discussions at the Washington workshop on researchers increasingly engaging in sharing activities. In London, the group developing measures on implementation success developed this measure of assessing OS engagement through time. Drawing from the FAIR guidelines, the McGill team changed the coding to include an assessment of the resources dedicated to preservation as well as the use of certified repositories.
22. This measure originated in discussions at the Washington workshop on the increased use of open processes and tools across the entire research workflows. In London, the group developing measures on implementation success developed this measure of assessing OS engagement through inclusivity.
23. This measure originated in discussions at the Washington workshop about greater levels of education and sharing of best practices of OS. In London, the group developing measures on implementation success developed this measure of assessing OS engagement through training.

B. Semi-Structured Interview Guides

1. N/A
2. This measure originated in discussions at the Washington workshop around whether OS collaborations improve the quality and frequency of communication to stakeholders. In London, the group developing measures on increased trust and accountability of the research enterprise translated the initial OS success outcome identified in Washington into the above questions. The McGill team refined the questions.
3. This measure originated in discussions at the Washington workshop around public trust, appreciation and understanding of science and the research process. In London, the group developing measures on increased trust and accountability of the research enterprise translated the initial OS success outcome identified in Washington into the above question. The McGill team refined the questions.
4. This measure originated in discussions at the Washington workshop around standard and consistent data management as well as the hypothesis the OS will lead to an increase in reliability, accountability, and reproducibility of scientific outputs. In London, the group developing measures on increased quality and efficiency of scientific output translated the initial OS success outcome identified in Washington into the above question. The McGill team refined the questions.
5. This measure originated in discussions at the Washington workshop that OS will possibly lead to the encouragement of researchers to further re-use and re-combination of datasets. In London, the group developing measures on increased quality and efficiency of scientific outputs translated the initial OS success outcome identified in Washington into the above question. The McGill team refined the questions.
6. This measure originated in discussions at the Washington workshop around standard and consistent data management and the hypothesis that the practice of OS will result in an increase in the reliability, accountability, and reproducibility of scientific outputs, because a larger community will be able to examine the results. In London, the group developing measures on increased quality and efficiency of scientific outputs translated the initial OS success outcome identified in Washington into the above question. The McGill team refined the questions.
7. This measure originated in discussions at the Washington workshop around the hypothesis that OS will increase trust and accountability because it will democratize the research process by creating more ways for community led research and for communities to participate in the process from the beginning. In London, the group developing measures on increased trust and accountability of the research enterprise

translated the initial OS success outcome identified in Washington into the above questions. The McGill team refined the questions.

8. This measure originated in discussions at the Washington workshop on how OS allows for neutral or better and more diverse opportunities for students, postdocs, and young researchers to launch their careers. In London, the group developing measures on better opportunities and recognition of early career researchers and youth translated the initial OS success outcome into the above questions.
9. This measure originated in discussions at the Washington workshop around more diverse opportunities for students, postdocs, and young researchers to launch their careers and the hypothesis that OS, through open access and open publications, will allow ECRs to gain visibility and get more cited to the increased accessibility of their work. In London, the group developing measures on better opportunities and recognition of early career researchers and youth translated the initial OS success outcome identified in Washington into the above questions. The McGill team refined the questions.
10. This measure originated in discussions at the Washington workshop around more diverse opportunities for students, postdocs, and young researchers to launch their careers and the hypothesis that OS, through open access and open publications, will allow ECRs to gain visibility and get more cited to the increased accessibility of their work. In London, the group developing measures on better opportunities and recognition of early career researchers and youth translated the initial OS success outcome identified in Washington into the above questions. The McGill team refined the questions.
11. This measure originated in discussions at the Washington workshop around the hypothesis that over time, OS will allow for better partnerships with industry to develop where both research and industry will benefit from collaboration. In London, the group developing measures on positive economic impact translated the initial OS success outcome identified in Washington into the above questions. The McGill team refined the questions.
12. This measure originated in discussions at the Washington workshop around greater public trust of science and the research process, as well as the hypothesis that OS will help drive the public drive the initiative and have a sense of empowerment, leading to high consent rates. In London, the group developing measures on increased trust and accountability of the research enterprise translated the initial OS success outcome identified in Washington into the above questions. The McGill team refined the questions.
13. This measure originated in discussions at the Washington workshop around how OS will lead to more effective oversight of research by ethics committees. In London, the group developing measures on increased trust and accountability of the research enterprise translated the initial OS success outcome identified in Washington into the above questions. The McGill team refined the questions.

C. Survey Measurement for Open Science Engagement

1. This measure originated in discussions at the Washington workshop on researchers engaging in sharing activity, as well as the long-term outcome of researcher's attitudinal shift in favour of sharing research outputs. In London, the group developing measures on implementation success began development of this measure. The McGill team adapted the London framework into the current survey format.
2. This measure originated in discussions at the Washington workshop on researchers engaging in sharing activity, as well as the long-term outcome of researcher's attitudinal shift in favour of sharing research outputs. In London, the group developing measures on implementation success began development of this measure. The McGill team adapted the London framework into the current survey format.
3. This measure originated in discussions at the Washington workshop on researchers engaging in sharing activity, as well as the long-term outcome of researcher's attitudinal shift in favour of sharing research outputs. In London, the group developing measures on implementation success began development of this measure. The McGill team adapted the London framework into the current survey format.

D. Measures of Open Science

1. This measure originated in discussions at the Washington workshop around how OS can contribute to the generation of "sticky" knowledge that can then be used to generate local growth. In London, the group developing measures on positive economic impact developed the initial measure into the coding framework above.
2. This measure originated in discussions at the Washington workshop on how OS will accelerate the generation and translation of valuable knowledge to practice. In London, the group developing measures on accelerated Innovation and Impact translated the initial measure into this form.
3. This measure originated in discussions at the Washington workshop of the hypothesis that OS will allow for greater diversity in its participants, collaborators, and leaders to participate in research. This includes an increase in the amount of developing countries that are meaningfully involved in research and innovation. In London, the group developing measures on increased equity in research translated the initial measure into the current form. The McGill team refined the coding frame to be clearer.
4. This measure originated in discussions at the Washington workshop on greater public trust, appreciation and understanding of science and the research process, as well as the hypothesis that OS will increase trust and accountability because it will democratize the research process by creating more ways for community led research and for communities to participate in the process. In London, the group developing measures on increased trust and accountability of the research enterprise translated the current coding framework.

E. Open Science Measures to be Considered by Others

1. This measure originated in discussions at the Washington workshop on improved data management practice, including marketing and stewardship. In London, the group on developing measures on increased trust and accountability developed this measure. The McGill team changed the measure from an interview to an annual report.
2. This measure originated in discussions at the Washington workshop on increased number of data professionals. In London, the group developing measures on implementation success developed this measure.

F. Incomplete and Rejected Open Science Measures

1. This measure originated in discussions at the Washington workshop on increased equity in research, and in particular the subcategory on greater diversity in research. Greater diversity includes participants, collaborators, and leaders. In London, the group developing measures on increased equity in research began development of this measure.
2. This measure was suggested by one of the delegates during the review of the toolkit.
3. This measure originated in discussions at the Washington workshop on greater diversity and the number of accessible, useable, and interoperable datasets. In London, the group developing measures on increased quality and efficiency of scientific outputs began development of this measure.
4. This measure originated in discussions at the Washington workshop on increased quality and efficiency of scientific outputs, and in particular on reduced quantity of research on the same targets. In London, the group developing measures on increased quality and efficiency of scientific outputs began development of this measure.
5. This measure originated in discussions at the Washington workshop on use, re-use and re-combination of datasets. In London, the group developing measures on increased quality and efficiency of scientific outputs in London began development of this measure.

6. This measure originated in discussions at the Washington workshop on greater diversity and the number of accessible, useable, and interoperable datasets. In London, the group developing measures on increased quality and efficiency of scientific outputs in London began development of this measure.
7. This measure originated in discussions at the Washington workshop on greater diversity and the number of accessible, useable, and interoperable datasets. In London, the group developing measures on increased quality and efficiency of scientific outputs began development of this measure.
8. This measure originated in discussions at the Washington workshop on greater use, re-use and re-combination of datasets. In London, the group developing measures on increased quality and efficiency of scientific outputs began development of this measure.
9. This measure originated in discussions at the Washington workshop on greater diversity and the number of accessible, useable, and interoperable datasets. In London, the group developing measures on increased quality and efficiency of scientific outputs began development of this measure.
10. This measure originated in discussions at the Washington workshop on researchers engaging in sharing activity. In London, the group developing measures on implementation success began development of this measure.
11. This measure originated in discussions at the Washington workshop on standard and consistent data management and curation of datasets. In London, the group developing measures on increased quality and efficiency of scientific outputs began development of this measure.
12. This measure originated in discussions at the Washington workshop on greater diversity and number of accessible, useable, and interoperable data sets with detailed metadata. In London, the group developing measures on increased quality and efficiency of scientific outputs began development of this measure.
13. This measure originated in discussions at the Washington workshop on increased reliability and reproducibility of scientific outputs. In London, the group developing measures on increased quality and efficiency of scientific outputs began development of this measure.
14. This measure originated in discussions at the Washington workshop on accelerated innovation and impact. In London, the group developing measures on accelerated innovation and impact began development of this measure.
15. This measure originated in discussions at the Washington workshop on greater diversity and number of accessible, useable and interoperable datasets. In London, the group developing measures on increased efficiency and quality of scientific began development of this measure.
16. This measure originated in discussions at the Washington workshop on neutral or better and more diverse opportunities for students, postdocs, and young researchers to launch their careers. In London, the group developing measures on better opportunities and recognition of early career researchers and youth began development of this measure.
17. This measure originated in discussions at the Washington workshop on reduced barriers to graduate students moving between academia and industry. In London, the group developing measures on better opportunities and recognition of early researchers and youth began development of this measure.
18. This measure originated in discussions at the Washington workshop on neutral or better and more diverse opportunities for students, postdocs, and young researchers to launch their careers. In London, the group developing measures on better opportunities and recognition of early career researchers and youth began development of this measure.
19. This measure originated in discussions at the Washington workshop on greater use, re-use and re-combination of datasets. In London, the group developing measures on Increased quality and efficiency of scientific outputs began development of this measure.

20. This measure originated in discussions at the Washington workshop group on increased access to meaningful, understandable, and useable health information. In London, the group developing measures on accelerated innovation and impact began development of this measure.
21. This measure originated in discussions at the Washington workshop on reduced rates of scientific misconduct around retractions. In London, the group developing measures on increased quality and efficiency of scientific outputs began development of this measure.
22. This measure originated in discussions at the Washington workshop on reducing barriers to graduate students moving between academia and industry. In London, the group developing measures on better opportunities and recognition of early career researchers and youth began development of this measure.
23. This measure originated in discussions at the Washington workshop on positive economic impact of OS. In London, the group developing measures on positive economic impact began development of this measure.
24. This measure originated in discussions at the Washington workshop on researchers engaging in sharing activity. In London, the group developing measures on implementation success began development of this measure.
25. This measure originated in discussions at the Washington workshop on neutral or better and more diverse opportunities for students, postdocs, and young researchers to launch their careers. In London, the group developing measures on better opportunities and recognition of early career researchers and youth began development of this measure.
26. This measure originated in discussions at the Washington workshop on neutral or better and more diverse opportunities for students, postdocs, and young researchers to launch their careers. In London, the group developing measures on better opportunities and recognition of early career researchers and youth began development of this measure.
27. This measure originated in discussions at the Washington workshop on increasing public trust in science. In London, the group developing measures on trust and accountability began development of this measure.
28. This measure originated in discussions at the Washington workshop on neutral or better and more diverse opportunities for students, postdocs, and young researchers to launch their careers. In London, the group developing measures on better opportunities and recognition of early career researchers and youth began development of this measure.
29. This measure originated in discussions at the Washington workshop in on neutral or better and more diverse opportunities for students, postdocs, and young researchers to launch their careers. In London, the group developing measures on better opportunities and recognition of early career researchers and youth began development of this measure.
30. This measure emerged in discussions at the Washington workshop about faster generation and translation of valuable knowledge to practice. The measure also emerged from the hypothesis that because of OS's greater openness, there might be less risk for participants to explore new possibilities, thus making it possible to identify innovative ideas earlier in the research process and accelerate discoveries. In London, the group developing measures on accelerated innovation and impact translated the initial measure into the coding framework above.
31. This measure was derived in discussions at the Washington workshop on how OS is likely to create an ecosystem that will attract experts and partners, allowing mutually beneficial partnerships between researchers and industry to develop. In London, the group developing measures on positive economic impact translated the initial measure into this form.
32. This measure originated in discussions at the Washington workshop around the hypothesis that OS will encourage re-use and re-combination of datasets as well as increase reliability, accountability, and

reproducibility. In London, the group developing measures on increased quality and efficiency of scientific outputs developed this measure.

33. This measure originated in discussions at the Washington workshop on greater transparency across the research process. In London, the group developing measures on increased trust in and accountability of the research enterprise began development of this measure.
34. This measure originated in discussions at the Washington workshop addressing the hypothesis that OS will lead to faster generation and greater translation of valuable knowledge into policy and practice. In London, the group developing measures on accelerated innovation and impact translated the initial measure into this form.
35. This measure originated in discussions at the Washington workshop on neutral or better and more diverse opportunities for students, postdocs, and young researchers to launch their careers. In London, the group developing measures on better opportunities and recognition of early career researchers and youth developed this measure.

Supplementary File 5: Washington Leadership Forum Participant List

Dr. Sarah Ali-Khan, Research Associate, Faculty of Law, McGill University

Dr. Lluis Ballell-Pages, Director, External Opportunities, GlaxoSmithKline, Tres Cantos Open Lab for Diseases of the Developing World

Dr. Patricia Brennan, Director, National Library of Medicine, Interim Associate Director for Data Science, National Institutes of Health, US Department of Health and Human Services

Dr. Katja Brose, Science Program Officer, Chan Zuckerberg Science Initiative

Ms. Rachel Bruce, Head of Open Science, Department for Business, Energy & Industrial Strategy, UK Government

Mr. David Carr, Program Manager - Open Research, Wellcome Trust

Dr. Simon Chaplin, Director of Culture and Society, Wellcome Trust

Me Mylène Deschênes, Ethics and Legal Advisor to Chief Scientist of Quebec, Fonds de recherche du Québec

Dr. Aled Edwards, Chief Executive Officer, Structural Genomics Consortium

Ms. Ashley Farley, Associate Officer of Knowledge & Research Services, Bill & Melinda Gates Foundation

Dr. Richard Gold, James McGill Professor, Faculty of Law, McGill University

Ms. Jennifer Hansen, Senior Officer, Knowledge & Research, Bill & Melinda Gates Foundation

Dr. Jason Karamchandani, Associate Professor, Department of Pathology McGill University

Dr. Michael Hawrylycz, Investigator, Allen Institute for Brain Science

Dr. Nadia Khelef, Senior Advisor for Global Affairs, Institut Pasteur

Mr. Robert J Kiley, Head of Open Research Development, Wellcome Trust

Ms. Elizabeth Kittrie, Strategic Advisor for Data and Open Science, US National Institutes of Health

Mr. Manoj Kumar, Head of Entrepreneurship and Innovations, Tata Trusts

Dr. Alexandre Le Bouthillier, Founder and COO, Imagia

Dr. Matthew Lucas, Executive Director, Corporate Strategy and Performance, Social Sciences and Humanities Research Council of Canada

Dr. Thomas Maina Kariuki, Director of the Alliance for Accelerating Excellence in Science in Africa

Dr. Lara Mangravite, President, Sage Bionetworks

Ms. Jessica Mankowski, Manager, Knowledge Translation Strategies, Canadian Institutes of Health Research

Dr. Sanjay Mehendale, Additional Director General, Indian Council of Medical Research

Dr. Mark Namchuk, SVP Research, Non-Clinical and Pharmaceutical Development, Alkermes

Ms. Thea Norman, Senior Program Officer, Bill & Melinda Gates Foundation

Mr. James O'Leary, Chief Innovation Officer, Genetic Alliance

Dr. Sébastien Paquet, Lead Applied Research Scientist & Culture Hacker, Element AI

Mr. Ben Pierson, Senior Program Officer, Bill & Melinda Gates Foundation

Dr. Claude Pirmez, Senior Researcher, Oswald Cruz Institute

Ms. Casey Selwyn, Fellow, Global Health Program**,** Bill & Melinda Gates Foundation

Ms. Annabel Seyller, Chief Operating Officer of the Open Science Experiment, The Montreal Neurological Institute

Dr. Carthage Smith, Senior Policy Analyst, Organization for Economic Co-Operation and Development (OECD)

Dr. Jeff Spies, Chief Technology Officer and Co-founder, Centre for Open Science

Dr. David Sweeney, Executive Chair Designate of Research England and Director, Research and Knowledge Exchange Higher Education Council for England

Dr. Michiel van Den Hauten, Head/Deputy Director of Research and Science Policy at the Ministry of Education, Culture and Science

Dr. Kate Williams, Scientific Director, Krembil Foundation

Supplementary File 6: London Workshop Participant List

Dr. Sarah Ali-Khan, Research Associate, Faculty of Law, McGill University

Dr. Liz Allen, Director of Strategic Initiatives, F1000

Dr. Eric Archambault, CEO and Founder, Science-Metrix

Dr. Manoel Barral Netto, Vice-President of Education, Information and Communication

Ms. Sarah Bowman, Director of Strategic Engagement, Trinity College Dublin

Ms. Rachel Bruce, Head Open Science, Department for Business, Energy & Industrial Strategy, UK Government

Dr. Tania Bubela, Dean, Simon Fraser University

Mr. Dave Carr, Program Manager, Open Research, Wellcome Trust

Mr. Damien Chalaud, Senior Director of Communications, Montreal Neurological Institute and Hospital

Dr. Matthew Clancy, Structure, Productivity and Technology Branch Resource and Rural Economics Division, US Department of Agriculture

Dr. Patricia Clarke, Program Manager Policy and EU Funding, National Delegate for H2020 Health, HRB Post-award and Evaluation Unit, Health Research Board of Ireland

Ms. Ève-Lyne Comtois-Dinel, PhD Candidate, Sociology, University of Québec à Montréal and TÉLUQ, Intern OECD

Dr. Bob Cook-Deegan, Professor, School for the Future of Innovation in Society, University of Arizona

Mr. Adam Dinsmore, Insight Research Analyst, Wellcome Trust

Ms. Megan Doerr, Principal Scientist, Governance, Sage Bionetworks

Dr. Aled Edwards, Director and CEO, Structural Genomics Consortium, Chair of Board M4K Pharma, University of Toronto

Dr. Keri Facer, Professor of Educational and Social Futures

Ms. Ashley Farley, Associate Officer of Knowledge & Research Service, BMGF

Ms. Lisa Federer, Research Data Informationist, US National Institutes of Health (NIH) Library

Dr. Maryann Feldman, Professor of Public Policy, University of North Carolina

Dr. Fernando Galindo-Rueda, Senior Economist, Organization for Economic Co-Operation and Development (OECD)

Dr. Richard Gold, James McGill Professor, Faculty of Law, McGill University
Ms. Jennifer Hansen, Senior Officer, Knowledge & Research, Bill & Melinda Gates Foundation

Mr. Steven Hill, Director of Research, Research England and UK Research and Innovation

Mr. Neil Jacobs, Head of Scholarly Communications Support, Jisc

Dr. Osmat Jefferson, Director of Applications Development, The Lens; Professor of Science, Technology & Law, Queensland University of Technology; Principal Scientist, Cambia

Dr. Chonnettia Jones, Director of Insight and Analysis, Wellcome Trust

Dr. Linda Kahl, Senior Counsel (IP), The Lens

Dr. Jason Karamchandani, Director, MNI Biobank, Neuropathologist, Montreal Neurological Institute, McGill University

Mr. Robert Kiley, Head of Open Research, Wellcome Trust

Dr. Bianca Kramer, Information and Collection Specialist, Life Sciences and Medicine, Utrecht University Library, Utrecht University

Mr. Manoj Kumar, Head of Entrepreneurship and Innovations, Tata Trusts

Dr. Wen Hwa Lee, Director, Disease Foundation Network, Strategic Alliances, Structural Genomics Consortium

Dr. Devika Madalli, Professor, Indian Statistical Institute

Dr. Marc Malandro, Vice President of Operations, Chan-Zuckerberg Initiative

Mr. Christopher Manuel, Corporate Performance Analyst, Canadian Institutes of Health Research

Ms. Elizabeth Marincola, Senior Advisor for Science Communications and Advocacy, African Academy of Sciences

Dr. Daniel Mietchen, Data Scientist, Data Science Institute, University of Virginia

Dr. Jenny Molloy, Shuttleworth Foundation Research Fellow, Department of Plant Sciences,

University of Cambridge

Dr. Molly Morgan Jones, Senior Analyst, Innovation and Technology Policy, RAND Europe

Dr. Brian Nosek, Executive Director, Centre for Open Science, Professor, University of Virginia

Dr. Claude Pirmez, Senior Researcher, Oswaldo Cruz Institute

Dr. Ismael Ràfols, Associate Faculty, University of Sussex, Universitat Politecnica de Valencia, Centre for Science and Technology Studies Leiden University

Dr. Robin Simpson, Lead Analyst, Department for Business, Energy and Industrial Strategy (BEIS), UK

Dr. Malcolm Skingle, Director, Academic Liaison, GlaxoSmithKline

Dr. Carthage Smith, Senior Policy Analyst, OECD

Dr. Sophie Staniszewska, Professorial Fellow, Patient and Public Involvement and Experiences of Care, University of Warwick

Dr. Mike Thelwall, Professor of Data Science, School of Mathematics and Computer Science, University of Wolverhampton

Dr. Steve Wooding, Lead, Research and Analysis, Centre for Science and Policy, University of Cambridge