Report of the G7 Open Science - Research on Research Sub-Working Group

Matthew Cloutier, Marin Dacos

Report of the G7 Open Science

"Research on Research" Sub-Working Group

Prepared for the G7 Open Science Working Group

May 2023 - DOI : [10.52949/32]

Issue

The G7 Open Science Working Group (OSWG) was established to *share open science policies, explore supportive incentive structures, and identify good practices for promoting increasing access to the results of publicly funded research, including scientific data and publications*[1] [...] To inform the deliberations of the G7 OSWG, three sub-working groups were created to explore key Open Science topics and formulate policy recommendations:

- Interoperability and Sustainability of Infrastructures
- Research Assessment and Incentives
- Research on Research

This report pertains to the topic of Research on Research on Open Science. The mandate of the G7 OSWG Research on Research Sub-Working Group consisted of:

- Examining questions at the intersection of Research on Research and Open Science, which are relevant to the adoption of Open Science practices.
- Helping to mobilize existing Research on Research knowledge to build more evidence-based Open Science policies.
- Identifying knowledge gaps in Research on Research that should be filled, drafting a first attempt of Research on Research agenda, to define more efficient Open Science policies.
- Proposing actions based on lessons learned, to inform decision-making and accelerate the Open Science journey.

The report was informed by two workshops held by the G7 OSWG Research on Research Sub-Working Group, which convened Open Science experts from across the world to discuss topics such as data-sharing, reproducibility, and research evaluation (October 2021), as well as academic, societal, and economic impacts of Open Science (June 2022). It was also informed through Research on Research itself, as well as discussions and knowledge-sharing between all three G7 OSWG Sub-Working Groups (Research on Research, Interoperability and Sustainability of Infrastructures, and Research Assessment and Incentives). The report was also informed by exchanges with individual researchers (see the list of all contributors at the end of the report). It consists of an overview of the Research on Research landscape on Open Science, as well as research recommendations on behalf of the G7 OSWG Research on Research Sub-Working Group.

---

[1] Communiqué of the G7 Ministers of Science and Technology 2016 (utoronto.ca)

After more than 10 years of Open Science policies, 20 years of Open Science debates[2] and 30 years after the first digital Open Science initiative (ArXiv), there is still much work to be done to achieve a fully Open Science-minded academic landscape. Some studies have demonstrated that close to 50% of contemporary research publications are open access globally[3]; however, that proportion fluctuates widely depending on fields of research and countries. With regards to Open or FAIR data[4] (findable, accessible, interoperable, and reusable), there are no conclusive studies providing an estimate of the proportion of data globally that is open and/or FAIR. However, the proportion is likely close to 20%, which corresponds more or less to the so-called "big sciences"[5]. In addition, results are still low on issues such as reproducible research[6] and publication bias[7], and research assessment reform is only in its first stages[8] to facilitate the Open Science transformation. This clearly demonstrates that the path to Open Science has just begun, despite three decades of individual, collective, and political efforts. If the Open Science paradigm is difficult to achieve due to the radical nature of the transformation it implies, the slowness of the associated cultural changes must be understood as resulting from the significant challenge in addressing the cultural changes with adequate, fine-tuned, and contextualized Open Science policies. In other words, if the big picture is clearly drafted, the steps needed to achieve it in a reasonable time are not all clearly identified. Additional efforts to understand obstacles and efficient solutions are needed. In this respect, evidence-based Open Science policies will be of great help, and research on research could play a key role[9] to address the Open Science challenge.

---

[2] In 1997, the report *Bits of Power: Issues in Global Access to Scientific Data* of the National Research Council already stated the importance of open research data.1. National Research Council. *Bits of Power: Issues in Global Access to Scientific Data*. (National Academies Press, 1997). doi:10.17226/5504.

[3] Piwowar, H. *et al. The State of OA: A large-scale analysis of the prevalence and impact of Open Access articles*. https://peerj.com/preprints/3119 (2017) doi:10.7287/peerj.preprints.3119v1.

[4] We use this definition of scientific data: "The recorded factual material commonly accepted in the scientific community as of sufficient quality to validate and replicate research findings, regardless of whether the data are used to support scholarly publications. Scientific data do not include laboratory notebooks, preliminary analyses, completed case report forms, drafts of scientific papers, plans for future research, peer reviews, communications with colleagues, or physical objects, such as laboratory specimens." https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html

[5] Borgman, C. L. *Big Data, Little Data, No Data: Scholarship in the Networked World*. (The MIT Press, 2015).

[6] Schweinsberg, M. et al. Same data, different conclusions: Radical dispersion in empirical results when independent analysts operationalize and test the same hypothesis. Organizational Behavior and Human Decision Processes 165, 228–249 (2021). DOI: https://doi.org/10.1016/j.obhdp.2021.02.003
Reproducibility of scientific results in the EU: scoping report. (2020). https://data.europa.eu/doi/10.2777/341654

[7] Fanelli, D. Negative results are disappearing from most disciplines and countries. *Scientometrics* 90, 891–904 (2012). https://doi.org/10.1007/s11192-011-0494-7
Vinkers, C. H. *et al.* The methodological quality of 176,620 randomized controlled trials published between 1966 and 2018 reveals a positive trend but also an urgent need for improvement. *PLOS Biology* 19, e3001162 (2021). DOI: https://doi.org/10.1371/journal.pbio.3001162

[8] Towards a reform of the research assessment system: scoping report. (Publications Office of the European Union, 2021). https://data.europa.eu/doi/10.2777/707440

[9] Ioannidis, J. P. A. Meta-research: Why research on research matters. PLOS Biology 16, e2005468 (2018). https://doi.org/10.1371/journal.pbio.2005468

Definition, Landscape and Results of Research on Research

Research on Research can be used to study, mobilize, improve, and accelerate the Open Science journey. In recent years, the importance of Research on Research on Open Science has been demonstrated by the number of global initiatives. For example, a 2021 report has identified different approaches and initiatives in the field: "*Some are theoretically and methodologically rooted in traditional academic fields such as sociology, economics, political science, philosophy, or information science (bibliometrics and scientometrics). Others with more data-intensive approaches come from computational social sciences or biomedical fields and have evolved in a favourable context for open science in terms of public policies. While the former currents are based on well-established pre-existing academic fields and methods, such as science and technology studies (STS) or scientometrics, the latter have appeared more recently, and have adopted a prescriptive, change-oriented focus as well as a normative commitment to foster better and more open science.*"[10] The report also refers to the evolving nature of Research on Research, notably being fuelled by contemporary debates "*such as reproducibility, evidence-based practices, integrity and inclusivity in research, and some community-issued warnings about not "reinventing the wheel."* According to the authors, new alliances "*are forming between research centres and laboratories, funding institutions, policy-makers and data providers in order to support public policy-makers with evaluation tools and research protocols to guide decision-making and action*." There are thus basically two kinds of researchers that are specialized in Research on Research: both Humanities and Social Sciences (HSS) and Science, Technology, Medicine (STM) researchers that specialized in Research on Research itself; and researchers not specialized in Research on Research, but who invest their time and research effort in their own discipline, in order to improve the efficiency, transparency, and reproducibility of their own research.

Depending on its history, its methodologies and problematic, Research on Research encompasses different connected research fields: sociology of science, science and technology studies (STS), meta-research, metascience, and science of science. This research community publishes in dedicated Research on Research journals, such as Journal of Infometrics, Scientometrics, and Quantitative Science Studies. However, the research community also publishes in journals dedicated to a different specific research field (for example, in PLOS journals) or in journals where Research on Research plays an important role without being the focus of the publication (for example, the *Journal of the Association for Information Science and Technology, Revue d'anthropologie des connaissances*).

Research on Research has been structured around the creation of research centres or specialized initiatives dedicated to creating a community of researchers. Examples include the Association for Interdisciplinary Meta-Research and Open Science (AIMOS), the Center for Open Science (COS), the Centre for Science and Technology Studies (CWTS), the CSI (Centre de Sociologie de l'Innovation), the Meta-Research Innovation Center at Stanford (METRICS), the Canada Research Chair on the Transformations of Scholarly Communication,

[10] Gruson-Daniel, C. & Anderson-González, M. Étude exploratoire sur la « recherche sur la recherche » : acteurs et approches. 66 p. (Comité pour la science ouverte, 2021). doi:10.52949/24.

the planned Centre for Research on Research on Open Science (LabSo)[11]. There are, however, many other growing initiatives such as National Institute for Informatics[12] (NII, Tokyo), National Institute of Science and Technology Policy[13] (NISTEP, Tokyo), the QUEST Center for Responsible Research (Berlin), ELICO (Lyon), Institut Francilien Recherche Innovation Société (IFRIS, Paris), Center for Science of Science and Innovation (CSSI, Evanston, Illinois), Science of Science & Computational Discovery Lab (University of Colorado), NEtwoRks, Data, and Society (NERDS, IT University of Copenhagen), School of Public Policy (Georgia Institute of Technology), etc.

These initiatives allow the Research on Research field to be highly active on multiple and diverse dimensions, which can all benefit Open Science. Let us cite Nanobubbles (how the correction of science works or fails), Pathways (investigating how data is gathered about research careers), Randomization (experiments to test how well lottery-style research funding works), the Harbinger project[14] (research about the place of Early Career Researchers in the current transformation of the academic world), ON-MERRIT (how and if open and responsible research practices could worsen existing inequalities), research on Wikipedia (links between academic literature and Wikipedia contents and users) or Data journeys in the Sciences (understanding disciplinary, geographical and historical differences and similarities in data management).

Some Research on Research initiatives come with a digital service aimed at filling a technical gap in order to improve openness, transparency and reproducibility, such as the Open Science Framework, the Leiden Ranking, COKI (Curtin Open Knowledge Initiative), the Initiative for Open Citations I4OC, Open Journals System, Zotero, Protocols.io, Software Heritage, CRediT (Contributor Roles Taxonomy), GeneRation Of BIbliographic Data (GROBID), SoftCite and Datastet (detection of authors and mentions of publications, datasets and software), the Cochrane Library (a major meta-analysis initiative in health research) and Dataverse (software dedicated to data repositories)[15]. This list indicates that Research on Research is continuously fuelling the academic ecosystem with Open Science tools, dedicated to both research itself and to policy-makers.

Research on Research has already provided many results that are of great interest for Open Science policies. For example, the open access citation advantage[16] and the open data citation advantage[17] provide a major incentive that could be used by policy-makers to help

---

[11] These organizations are described in a more detailed manner in the annex of this report.
[12] National Institute for Informatics, Tokyo
[13] National Institute of Science and Technology Policy, Tokyo
[14] https://elico-recherche.msh-lse.fr/programme/harbinger-research-project
[15] These examples of research programs and Open Science services are detailed in the annex of this report.
[16] Piwowar, H. *et al. The State of OA: A large-scale analysis of the prevalence and impact of Open Access articles*. https://peerj.com/preprints/3119 (2017) doi:10.7287/peerj.preprints.3119v1. McKiernan, E. C. *et al.* How open science helps researchers succeed. *eLife* **5**, e16800 (2016). https://doi.org/10.7554/eLife.16800
[17] Colavizza, G., Hrynaszkiewicz, I., Staden, I., Whitaker, K. & McGillivray, B. The citation advantage of linking publications to research data. *PLOS ONE* **15**, e0230416 (2020).

researchers to take the leap towards Open Science[18]. Other research works have outlined that open access is beneficial not only to researchers themselves, but also to students (at a massive level) and to society outside academia, which, represents approximately 30% of the usage of the open access publication platforms[19]. The latter is dominated by professional activities, outside academia, but personal uses have been evidenced and are meaningful.

Figures are also available on clinical trial results/publications, which are required legally in many countries. On lung cancer, for instance, all existing systematic reviews are consistently incomplete, because they rely on only 45% to 70% of the existing trials, which is the fraction that is actually published[20]. The results are not yet monitored on a regular basis, though this monitoring is necessary to understand if and how research on these public health issues can be advanced. On another theme, recent research on COVID 19 genome sharing has highlighted the complexity of the topic, the interest for associated data to be fully open, and the need to overcome the controlled access by changing the recognition rules and considering the need for reward and recognition for the data providers.

Though some striking results concerning the economic impact of Open Science are available, the Human Genome Project and the opening of the Web by CERN would probably be the best examples. Nevertheless, there is still a lot of research needed to deeply understand and evaluate the economic impact of Open Science. In this spirit, the French Parliament has requested a report on the production and valorization of research software (both open source and proprietary). This report will fuel further knowledge about the impacts of opening research software and potential policy evolutions. The European Commission has also funded PathOS, a project which seeks to "*better understanding and measuring* [of] *Open Science impacts and their causal mechanisms*"[21]

Similarly, several countries have developed indicators concerning their publication open access ratio, but there is yet only one attempt at estimating the ratio of open research data and software at the country level. The reason for that is that the methodological challenge is huge: in short, assessing the openness and fairness of research objects is nearly impossible because many are hidden in USB keys and in the hard disks of personal computers, far away from any unified catalogue. However, the first findings of Research on Research on this topic indicate that, amongst the French publications in 2021 that mention the production of data, 22% are referring to sharing a dataset, while it amounts to 20% for research software.[22].

The complex question of costs of Open Science is also quite a challenge, to help in driving investments and policy choices. Regarding Open Science infrastructures, there is, to our knowledge, only one research publication on the question of research infrastructure failing to

[18] Peroni, S. & Shotton, D. OpenCitations, an infrastructure organisation for open scholarship. Quantitative Science Studies 1, 428–444 (2020). https://doi.org/10.1162/qss_a_00023.
Also see: https://i4oc.org/

[19] Alperin, J. P. The public impact of Latin America's approach to open access. https://purl.stanford.edu/jr256tk1194

[20] Créquit, P., Trinquart, L., Yavchitz, A. & Ravaud, P. Wasted research when systematic reviews fail to provide a complete and up-to-date evidence synthesis: the example of lung cancer. BMC Medicine 14, 8 (2016). DOI: https://doi.org/10.1186/s12916-016-0555-0

[21] https://pathos-project.eu/

[22] French Open Science Monitor (esr.gouv.fr)

meet their initial goals/expectations[23] . Having more of this kind of research could also help learn from mitigated success in attracting research-based users. Many more topics could be added to this list, such as the needed research about the values of researchers and their adherence to the Open Science principles.

All these topics are detailed in the annex of this report, including research assessment, Open Science infrastructures, data sharing and research cultures, code and software, reproducibility of research and trust, academic, economic and societal impact of Open Science, costs of Open Sciences and Open Science monitoring. The annex also focuses on the case study about genomic data sharing during the COVID 19 pandemic[24], thereby showing the inherent complexity of the topic, which has been shown not to be addressed properly by mere public declarations of willingness to open all COVID 19 research data.

Research on Research Recommendations

There is a need to deepen research on research on Open Science at an international level, as well as increase coordination and knowledge-sharing. Existing research has paved the way; however, there are often needs of confirmation and diversification on a disciplinary level, as most of the existing research is discipline-specific. For less explored questions, there is a need to open new research questions. For example, the questions of the economic and societal impacts of Open Science need more research, including on the costs of Open Science, and its benefits. Moreover, the question of compatibility of values and cultures when confronted with Open Science policies, recommendations, tools, and processes could be addressed. If we want Open Science to become the default in academics' daily life, existing obstacles could be carefully studied, and the facilitating approaches could be tested against a variety of situations and disciplines. The list of questions to be addressed is thus very large; however, one could attempt to group them within 5 main categories (keeping in mind that the list is not exhaustive). The scope of the questions varies based on the level of maturity of the topic.

1. Research Assessment and Incentives
    - What are the lessons learned of the existing experimentations concerning narrative CVs?
    - How can we achieve the goal of fully open and reusable bibliographic databases and other research outputs databases to plan research assessment only on FAIR data?[25]

2. Skill Gaps
    - Among the numerous initiatives dedicated to training and knowledge-sharing of Open Science research data management, software management, and publishing,

[23] Dombrowski, Q. What Ever Happened to Project Bamboo? *Lit Linguist Computing* fqu026 (2014) doi:10.1093/llc/fqu026.
[24] *Intelligent open science: viral genomic data sharing during the COVID-19 pandemic*. 69 https://www.gov.uk/government/publications/intelligent-open-science-viral-genomic-data-sharing-during-the-covid-19-pandemic (2022).
[25] On this topic, see the projects I4OC, OpenAlex, I4OA.

how do we evaluate (qualitatively and quantitatively) success or failure of the initiatives?

- Open Science and Research Assessment are closely linked: how do we address the issue of skills needed to assess other forms of CVs, such as narrative ones?

3. Open Science Monitoring

- How the values of the research community evolve towards Open Science? This study could be initiated under the umbrella of an annual or biannual research, willing to extend the annual "state of open data" published by Digital Science.
- How can a worldwide Open Science Observatory, covering diverse dimensions of Open Science and not only publications, be built by using exclusively open data? How can we assess such an Observatory to ensure it is meaningful and mitigates any unintended biases?

4. Open Science Infrastructures

- Do Open Science badges increase Open Science practices where they have been tested?[26]
- How can we measure the success (or failure) of an Open Science infrastructure, beyond measuring the number of downloaded files or terabytes? How can we know to what extent an existing infrastructure meets the goal it has been designed for?
- How could we evaluate and monitor in the long run the usages, and the benefits, of Open Science infrastructures by and for the academic community and society as a whole?

5. Open Science Impacts

- Open Science is said to benefit the economy. How can it be measured and detailed?
- Open Science has societal impacts: it is said to enhance trust in science though the preprint process is hard to understand. How could we measure this and provide evidenced based solutions?
- Research itself is said to benefit from its opening, in many fields such as progress speed, reliability, cumulative nature, replicability, integrity… Can all this be evaluated, monitored, and described in different fields and situations?

[26] Rowhani-Farid, A., Aldcroft, A. & Barnett, A. G. Did awarding badges increase data sharing in BMJ Open? A randomized controlled trial. *Royal Society Open Science* **7**, 191818.
Kidwell, M. C. et al. Badges to Acknowledge Open Practices: A Simple, Low-Cost, Effective Method for Increasing Transparency. PLOS Biology 14, e1002456 (2016). Doi: 10.1371/journal.pbio.1002456
Rowhani-Farid, A., Aldcroft, A. & Barnett, A. G. Did awarding badges increase data sharing in BMJ Open? A randomized controlled trial. Royal Society Open Science 7, 191818. DOI: 10.1098/rsos.191818

Research on research could inspire a framework of Open Science monitoring. This topic is useful for national purposes and is also currently discussed by UNESCO to implement the UNESCO recommendation on Open Science (2021)[27]. It could address the different dimensions of Open Science and the different aspects of the research life cycle. Some parts of Open Science monitoring could be quantitative. Thanks to Unpaywall, we know how many publications are open access. And some parts could be qualitative: when we deal with culture, habits and resistance, the recommended approach could combine quantitative and qualitative approaches, and deal as much as possible with disciplinary and geographical variations. To get an Open Science monitor dealing with open/FAIR data with as much precision as with publications, we have far more work to do. We should avoid confusing research data available online in the 3000+ research data repositories in the world, which roughly represents 20% of the existing research data that has been used for publishing new knowledge[28], with the hidden iceberg of unknown, untraceable, unFAIR and uncurated research data that populate researchers' personal computers and USB keys.

There is also a need to provide a state of knowledge on Research on Research, to help decision and policy-makers know the latest findings of Research on Research. The International Panel of Climate Change (IPCC) undertakes a similar activity. We need the same kind of effort, but on the smaller scale of Research on Research. Strong meta-analysis published in one or several reports, translated in different languages, would be of great help, obviously published in open access, with open data as rich as possible to help the appropriation and reuse of the results.

All these questions, monitoring, and meta-analysis could be aggregated into something like an Open Science Observatory. The recommendation is to articulate and coordinate existing Research on Research on Open Science. There is also, however, a need to stimulate new research that would fit into the big picture of Open Science evidenced based policy making. To reach good results in a reasonable time and with Open Science policy in mind, international coordination would be really helpful. The already-existing research on research congresses, moreover when they focus on Open Science, along with the Research Data Alliance plenaries, and events of this kind, could be used as opportunities to help in building stronger and larger communities and stimulating transdisciplinary and trans-professional scholarly discussions in the field. There is also a need for academic coordination to draft collaboratively and then publish meta-analyses on the state of the art in research on research. The same coordination is needed to propose accurate indicators based on both qualitative and quantitative approaches.

This coordination could take the form of an international research **initiative** for **Research on Research and Open Science**. As the Open Science agenda is a global topic, such an initiative should not be restricted to G7 countries and be open to any other country willing to contribute to such an effort.

[27] UNESCO Recommendation on Open Science - UNESCO Digital Library
[28] French Open Science Monitor https://frenchopensciencemonitor.esr.gouv.fr/

Annex

1. Mandate of the Research on Research Sub-Working Group

The G7 Open Science Working Group (OSWG) was established to *share open science policies, explore supportive incentive structures, and identify good practices for promoting increasing access to the results of publicly funded research, including scientific data and publications*[29]*,* [...] To inform the deliberations of the G7 OSWG, three sub-working groups were created to explore three key topics and formulate policy recommendations:

- Interoperability and Sustainability of Infrastructures
- Research Assessment and Incentives
- Research on Research

This report pertains to the topic of Research on Research on Open Science. The mandate of the G7 OSWG Research on Research Sub-Working Group consisted of:

- Examining questions at the intersection of Research on Research and Open Science, which are relevant to the adoption of Open Science practices.
- Helping to mobilize existing Research on Research knowledge to build more evidence-based Open Science policies.
- Identifying knowledge gaps in Research on Research that should be filled, drafting a first attempt of Research on Research agenda, to define more efficient Open Science policies.
- Proposing actions based on lessons learned, to inform decision-making and accelerate the Open Science journey.

The report was informed by two workshops held by the G7 OSWG Research on Research Sub-Working Group, which convened Open Science experts from across the world to discuss topics such as data-sharing, reproducibility and research evaluation (October 2021), as well as academic, societal and economic impacts of Open Science (June 2022). It was also informed through Research on Research itself, as well as discussions and knowledge-sharing between all three G7 OSWG Sub-Working Groups (Research on Research, Interoperability and Sustainability of Infrastructures, and Research Assessment and Incentives). The report was also informed by exchanges with individual researchers (see the list of all contributors at the end of the report). It consists of an overview of the Research on Research landscape on Open Science, as well as research recommendations on behalf of the G7 OSWG Research on Research Sub-Working Group.

After more than 10 years of Open Science policies, 20 years of Open Science debates[30] and 30 years after the first digital Open Science initiative (ArXiv), there is still much work to be done to achieve a fully Open Science minded academic landscape. Some studies have demonstrated that close to 50% of contemporary research publications are open access

---

[29] Communiqué of the G7 Ministers of Science and Technology 2016 (utoronto.ca)

[30] In 1997, the report *Bits of Power: Issues in Global Access to Scientific Data* of the National Research Council already stated the importance of open research data.1. National Research Council. *Bits of Power: Issues in Global Access to Scientific Data*. (National Academies Press, 1997). doi:10.17226/5504.

globally[31]; however, that proportion fluctuates widely depending on the field of research and countries. With regards to Open Data or FAIR data[32] (findable, accessible, interoperable, and reusable), there are no conclusive studies providing an estimate of the proportion of data globally that is open and/or FAIR. However, the proportion is likely close to 20%, which corresponds more or less to the so-called "big sciences"[33]. In addition, results are still low on issues such as reproducible research[34] and publication bias[35], and research assessment reform is only in its first stages[36] to facilitate the Open Science transformation. This clearly demonstrates that the path to Open Science has just begun, despite three decades of individual, collective and political efforts. If the Open Science paradigm is difficult to achieve due to the radical nature of the transformation it implies, the slowness of the associated cultural changes must be understood as resulting from the significant challenge in addressing the cultural changes with adequate, fine-tuned and contextualized Open Science policies. In other words, if the big picture is clearly drafted, the steps needed to achieve it in a reasonable time are not all clearly identified. Additional efforts to understand obstacles and efficient solutions are needed. In this respect, evidence-based Open Science policies will be of great help, and research on research could play a key role[37] to address the Open Science challenge.

2. State of Knowledge and Research Agenda

2.1 Research Assessment and Incentives

By enabling broader access to research results and methodologies, such as data, protocols and publications, Open Science can increase the transparency, reproducibility, and impact of public funding of research and societal engagement[38].

---

[31] Piwowar, H. *et al. The State of OA: A large-scale analysis of the prevalence and impact of Open Access articles*. https://peerj.com/preprints/3119 (2017) doi:10.7287/peerj.preprints.3119v1.

[32] We use this definition of scientific data: "The recorded factual material commonly accepted in the scientific community as of sufficient quality to validate and replicate research findings, regardless of whether the data are used to support scholarly publications. Scientific data do not include laboratory notebooks, preliminary analyses, completed case report forms, drafts of scientific papers, plans for future research, peer reviews, communications with colleagues, or physical objects, such as laboratory specimens." https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html

[33] Borgman, C. L. *Big Data, Little Data, No Data: Scholarship in the Networked World*. (The MIT Press, 2015).

[34] Schweinsberg, M. et al. Same data, different conclusions: Radical dispersion in empirical results when independent analysts operationalize and test the same hypothesis. Organizational Behavior and Human Decision Processes 165, 228–249 (2021). DOI: https://doi.org/10.1016/j.obhdp.2021.02.003
Reproducibility of scientific results in the EU: scoping report. (2020). https://data.europa.eu/doi/10.2777/341654

[35] Fanelli, D. Negative results are disappearing from most disciplines and countries. *Scientometrics* 90, 891–904 (2012). https://doi.org/10.1007/s11192-011-0494-7
Vinkers, C. H. *et al.* The methodological quality of 176,620 randomized controlled trials published between 1966 and 2018 reveals a positive trend but also an urgent need for improvement. *PLOS Biology* 19, e3001162 (2021). DOI: https://doi.org/10.1371/journal.pbio.3001162

[36] Towards a reform of the research assessment system: scoping report. (Publications Office of the European Union, 2021). https://data.europa.eu/doi/10.2777/707440

[37] Ioannidis, J. P. A. Meta-research: Why research on research matters. PLOS Biology 16, e2005468 (2018). https://doi.org/10.1371/journal.pbio.2005468

[38] Susi, T., Heigtz M., Hnatkova, E., Koch, W., Leptin, M., Andler, M., Masia M., Garfinkel, M. (2022). Centrality of researchers in reforming research assessment- Route to improve research by aligning rewards with Open Science practices. Initiative for Science in Europe. Pg. 1-32 2022-03-16_ise_report_online_final.pdf (initiative-se.eu)

Although some aspects are being implemented by researchers in different organizations (e.g., Government, Universities, Institutions), Open Science overall is still far from being fully embraced by the research community. At the same instant, the ways in which research assessment is accomplished at present have become an important issue both for researchers and organizations. For example, research and researcher assessment for career progression are often based on the number of journal based and publication based metrics as an over-reliance on Journal Impact factor (JIF) and H-index[39]. Also, many academic hiring and promotion committees rely in part on journal reputation and prestige, when evaluating individual scholars and/or outputs. Thus, journals are often used as proxies of research quality. On the other hand, too few academic institutions value Open Science practices. Some institutions do have Open Access mandates, but those generally do not go beyond Open Access to publications and do not include the other aspects of Open Science such as promoting or mandating practices in utilizing FAIR data and community outreach for mutual learning and coordination. Few funders and institutions assess whether publications are Open Access and, in many cases, the mandates they put forward work poorly[40]. Regarding judgment of quality exercised by journals and peer reviewers, it is difficult to be confident in how journals evaluate the quality of research because most of the peer-review process is closed. In addition, systemic bias might result in "elite" researchers appearing more laudable, where the disproportionate scientific productivity can be explained by their substantial labour advantage rather than inherent differences in research aptitude[41]. Most research evaluation that occurs as part of the peer review process is unstructured or minimally structured and is not shared beyond the people directly involved with the peer review of the manuscript (i.e., authors, reviewers, editor). Therefore, the validity of journal-based peer review as a measure of quality is limited. In summary, intrinsic quality, integrity in the conduct of research, and contributions to the research community and to society are not part of researcher assessment in most of the research performing organizations (RPOs), globally.

- *Modern Science Values and Open Science*

A key factor for overcoming these systemic challenges is to improve the evaluation process and reward systems to include Open Science practices. To begin with, one must look at the core Open Science values of what constitutes the ethos of modern science. Robert Merton, an American sociologist, introduced the four best practices of good scientific research: Communalism, Universalism, Disinterestedness, and Organised scepticism[42]. According to Merton, Communalism addresses common ownership of scientific discoveries and the need

---

[39] Wang, J., Veugelers, R., Stephan, P. 2017. Bias against novelty in science: A cautionary tale for users of bibliometric indicators. Research Policy. 46: 1416-1436 https://doi.org/10.1016/j.respol.2017.06.006

[40] Larivière, V., & Sugimoto, C. R. (2019). The Journal Impact Factor: A Brief History, Critique, and Discussion of Adverse Effects. In W. Glänzel, H. F. Moed, U. Schmoch, & M. Thelwall (Eds.), *Springer Handbook of Science and Technology Indicators* (pp. 3–24). Springer International Publishing. https://doi.org/10.1007/978-3-030-02511-3_1

[41] Zhang, S., Wapman, K.H., Larremore, D.B., Clauset, A. 2022. Labour advantages lead to the greater productivity of faculty at elite universities. Science Advances 8: 1-9 https://www.science.org/doi/10.1126/sciadv.abq7056

[42] Merton RK. 1942. The Ethos of Science, J. Legal and Political Sociology. 1: 115-126. Reprinted In: Merton RK, Sztomka P., editor., editors. Social structure and science, Chicago: University of Chicago Press, 1996.

for scientists to publicly share their discoveries. This best practice could be seen as a precursor to modern initiatives such as Open Science. Universalism is the idea that everyone can do science, regardless of race, nationality, gender, or any other differences, and that everyone's scientific claims should be scrutinized equally. These values and practices should be utilized; however, the research climate falls short of this ideal. Research can sometimes be appraised and published based on the authority and status of its authors[43]. The culture of 'publish or perish' and the increased dependence on grants for success can sometimes obfuscate the value of scientific research.

Communism, or "norm communism," speaks to the communal character of scientific knowledge. Communication allowing research findings to be scrutinized by those that either like or dislike the research, which is good for scientific integrity[44]. As stated earlier, a barrier that can hinder the value of scientific research includes closed peer reviews. As an example, peer reviews are often solicited with the assurance that referee confidentiality will be preserved, the use of these reviews as data in studying the effects of secrecy or delays are impeded. Therefore, the researcher's ability to understand a principal mechanism of quality control in science is curbed by the functioning of the system itself[45].

Upholding norms of Open Science is challenging when personal interests and motivations of numerous groups in team efforts, and those of their institutions, collide and contrast. This is exactly where Mertonian values fall short[46]. For instance, a more prudent perspective is provided by those who note that openness "is not only a technical problem to be solved but is also a social, cultural, and moral issue" based on uneven social relationships and therefore, is not in itself positive or negative[47]. In many ways, commitment to scientific reforms and Open Science practices are defining and enlightening activities, discriminating between science as it is, science as it should be, and science as it was. Understanding this shift and its implications requires empirical work and will help shape the borders of various operationalizations of good science under banners such as 'research ethics,' 'research integrity,' 'research quality,' 'responsible research' or 'rigour[48].' Future work could investigate the extent to which early

[43] Latour, B., and S. Woolgar. 1986. Laboratory life: The construction of scientific facts. Princeton University Press.Merton, R. K. The Matthew Effect in Science. *Science* **159**, 56–63 (1968). https://www.jstor.org/stable/1723414
Klebel, T., Fava, I. & Ross-Hellauer, T. Matthew Effects in Open Science and RRI. Preprint at https://doi.org/10.5281/zenodo.4155449 (2020).
[44] Chubin, Daryl E. 'Open Science and Closed Science: Tradeoffs in a Democracy.' Science, Technology, & Human Values, vol. 10, no. 2, 1985, pp. 73–81.
[45] ibid
[46] Hosseini, Mohammad, et al. Messing with Merton: The Intersection between Open Science Practices and Mertonian Values. SocArXiv, 14 Sept. 2022. OSF Preprints, https://doi.org/10.31235/osf.io/9nyh8.
[47] Levin, N., Leonelli, S. (2017). How Does One "Open" Science? Questions of Value in Biological Research. Science, Technology, & Human Values, 42:2. https://doi.org/10.1177/0162243916672071
[48] Hosseini, Mohammad, et al. Messing with Merton: The Intersection between Open Science Practices and Mertonian Values. SocArXiv, 14 Sept. 2022. OSF Preprints, https://doi.org/10.31235/osf.io/9nyh8.

career researchers are able to display Mertonian and Open aspirations, and whether they could play a key role in that change[49].

- *From Novel Findings to Sound Science*

Although scientific progress rests on the repetition and reproduction of data and/or results, many studies cannot be competently analyzed or replicated. Open Science can try to combat the replication crisis, by providing incentives and rewards for research. That is, changing what we measure as a success in research, shifting from a culture that emphasizes novel findings to one that also rewards the many other aspects of practising sound science. To do this, scientists should try to incorporate modern values such as Scientific Integrity, Equity and Communication, and Diversity and Inclusion[50] when planning and conducting research. One way is to incorporate bottom-up approaches to tackling EDI issues within Open Science. Senior management within government and academia should take a more facilitating role rather than a controlling role, supporting initiatives at local and grass-roots levels rather than introducing top-down strategies that may not connect very well with local needs[51]. Adopting a reward system should be spent in a way that reduces the equity gap and enables more people to participate in open research practices. The best possible scenario is to reward those who have expressed interest in undertaking or adapting Open Science practices, by enabling researchers to acquire extra funds and/or security to be able to do so.

- *Inequities and Inclusiveness*

An emerging topic is raising concerns about possible new inequities introduced or reinforced by Open Science. These inequities could be observed in the G7 countries, but also in other countries. There are many different inequities observed. The first one is the inequity inside the academic publishing ecosystem. For example, the ON-MERRIT is a research project funded by the European Commission, which aims at eventually suggesting a set of evidence-based recommendations for science policies, indicators, and incentives, which could address and mitigate cumulative disadvantages, so-called Matthew effects. Many observers have noted that the generalization of article processing charges could provide stronger inequities in the capacity to publish than ever[52] (see also later in the report, the part dedicated to "Open Science costs").

The second major inequity is related to research data. However, this question of inequities and inclusiveness is less discussed concerning research data. The COVID 19 pandemic has proven strong inequities in access to genomic data in controlled repositories[53] (see later in the report, "The case of genomics during COVID 19").

---

[49] Nicholas, David, et al. 'Does the Scholarly Communication System Satisfy the Beliefs and Aspirations of New Researchers? Summarizing the Harbingers Research.' Learned Publishing, vol. 33, no. 2, Apr. 2020, pp. 132–41. DOI.org (Crossref), https://doi.org/10.1002/leap.1284.
[50] 1.4 Core Values | Open Science: An Introduction for Biology (ubco-biology.github.io)
[51] Equity, diversity and inclusivity in open research - BMJ Open Science
[52] Ross-Hellauer, T. Open science, done wrong, will compound inequities. *Nature* **603**, 363–363 (2022).
[53] Wadman, M. Critics decry access, transparency issues with key trove of coronavirus sequences. (2021) Doi: 10.1126/science.abi4496

"Helicopter research" describes predatory and unethical research practices from rich countries undertaking research in poor countries[54]. This issue is frequently associated with data collecting in the countries, without clear benefit and feedback for the researchers, people and countries that become raw material providers, when rich countries' researchers make the best use of it and receive the merit.

- *Skills Gaps*

Opening science means taking advantage of the digital transformations of the 21st century to initiate a deep change toward openness in the way research results and processes are disseminated. This implies evolutions in the research process itself, which should now include the dissemination policy, and thus also in the way research is evaluated, incentivized, and funded[55].

Reinvesting the dissemination part of the research workflow can mean, for instance, the design of new ways of openly communicating scientific writings, along with their business model which should incentivize research ethics and scientific integrity[56]. Furthermore, the very idea of disseminating research data and software is, still nowadays, completely new to many disciplines: the technical skill gap to gain adequate data and software literacy can be high for some communities. Indeed, openly, and efficiently disseminating paper, data and software requires technical skills that are not included in the 20th century research skill set. A paper by Koltay et al.[57] allows for instance to discover the wide range of skills needed to acquire appropriate "data literacy," among which "selecting, synthesizing and combining data," as well as "identify, collect, organize, analyze, summarize, and prioritize" it. Researchers should become data literate for them to efficiently share their data, but the question of data literacy also holds for all the staff[58] within the research process[59]. When it comes to software, the gap

Van Noorden, R. Scientists call for fully open sharing of coronavirus genome data. *Nature* **590**, 195–196 (2021). *doi: https://doi.org/10.1038/d41586-021-00305-7*

[54] Rochmyaningsih, D. Did a study of Indonesian people who spend most of their days under water violate ethical rules?
O'Grady, C. 'Helicopter research' comes under fire at Cape Town conference. (2022). doi:10.1126/science.add3544.

[55] McKiernan EC, Bourne PE, Brown CT, et al. How open science helps researchers succeed. Elife. 2016 Jul;5:e16800. DOI: 10.7554/elife.16800. PMID: 27387362; PMCID: PMC4973366.

[56] See for instance the project 101 innovations by Bianca Kramer and Jeroen Bosman, "Innovations in Scholarly Communication - Changing Research Workflows", Innovations in Scholarly Communication, accessed 9 February 2023, https://101innovations.wordpress.com/.

[57] Koltay, T. (2017). Data literacy for researchers and data librarians. Journal of Librarianship and Information Science, 49(1), 3–14. https://doi.org/10.1177/0961000615616450

[58] An attempt at a precise definition of the skills and jobs that are associated to data and data stewardship has for instance been released in the Netherlands :
https://doi.org/10.5281/zenodo.4623713

[59] This whole process of skill identification and training to access data literacy for all the involved communities is described in details in the framework of the European Open Science Cloud:
https://op.europa.eu/en/publication-detail/-/publication/af7f7807-6ce1-11eb-aeb5-01aa75ed71a1/language-en/format-PDF/source-190694287

is even higher because of the fragmented and network nature of software along its fragility[60]. Both for data and software literacy, research is needed to better demonstrate a way to bridge the gap, moreover for the latter, for which literature is very scarce when concerning the research community.

The skills gap that was addressed in the above lines concerns the skills needed to open research outcomes. For the sake of clarity, let us call them technical skills. These technical skills will take a piece of research to be clearly identified, as well as the way to efficiently bridge the gap. Nonetheless, initiatives attempting at bridging this gap are already appearing in various places around the world, coming both from the research and the libraries community. It is not the case for a second and distinct set of skills, for which the gap is probably higher, and more difficult to bridge. This second set of skills concerns the evaluation of research.

Another sub-working group of the G7 Open Science working group is working on evaluation. Its conclusions show that an evolution of the research evaluation practices is needed to allow for change to happen. We will not detail this as it is done by the other subgroup. However, should this change in the evaluation process and criteria occur, their implementation will require, on the one hand, researchers to change the way they are producing CVs to file for hiring, promotion or research projects and, on the other hand, peer reviewers to change the way they are evaluating CVs and files. Bridging the gap for this second set of skills is all the more difficult because the research community of peers must do it by itself. DORA, through its website[61], is among the very few initiatives providing tools that aim at bridging this gap. The newly born coalition CoARA[62] will certainly help in this direction equally. Fortunately, research on the new skills that are needed is already underway, as summarized in a report for the European University Association by Bregt Saenen and Lidia Borrell-Damián[63]. The width of the gap being now approximately known, more insight is needed on how to bridge it, moreover because of the smallness of the fraction of the research community that is aware of the necessary changes, a community whose skills are needed to carry out the peer evaluation process.

2.2 Open Science Infrastructures

Several recent studies have indicated that an increasing number of researchers now accept the value of openly sharing research data[64]. But there is a major difference between seeing

---

[60] Roberto Di Cosmo. Building the software pillar of Open Science. In *Open Science European Conference (OSEC 2022)*, pages 183--193. OpenEdition Press, 2022. DOI: https://doi.org/10.4000/books.oep.15829
[61] https://sfdora.org/
[62] https://coara.eu
[63]https://eua.eu/resources/publications/825:reflections-on-university-research-assessment-key-concepts,-issues-and-actors.htmlhttps://eua.eu/resources/publications/825:reflections-on-university-research-assessment-key-concepts,-issues-and-actors.html
[64] Gregory Goodey, Mark Hahnel, Yuanchun Zhou, Lulu Jiang, Ishwar Chandramouliswaran, Amy Hafez, Taunton Paine, Susan Gregurick, Samuel Simango, Juan Miguel Palma Peña, Holly Murray, Matt Cannon, Rebecca Grant, Kate McKellar, Laura Day, *The State of Open Data 2022*. (2022) doi:10.6084/m9.figshare.21276984.v5.
Béchec, Mariannig Le, et al. State of open science practices in france (SOSP-FR). Comité pour la science ouverte, 2022, p. 112 p. hal-lara.archives-ouvertes.fr, https://doi.org/10.52949/5.

the value of open research data provided by others, and the willingness of sharing the data you have produced with great efforts, budgets, time, and creativity. Yet, the actual practice of sharing data—especially in forms that comply with FAIR principles—remains a challenge for many researchers to integrate into their workflows and prioritize among the demands on their time[65]. The barriers to open data sharing are numerous. However, sustained funding from federal agencies in the United States including the NSF and NIH (e.g., the Data Management and Sharing Policy)[66] and important initiatives in other countries such as Canada's Tri-Agency Research Data Management Policy and the European Union's European Open Science Cloud (EOSC), and Japan's National Institute for Informatics Research Data Cloud (NII RDC)[67] are creating a growing infrastructure for open sharing of research data[68].As defined in the UNESCO Recommendation on Open Science[69], infrastructure refers to shared research infrastructures that are needed to support Open Science and serve the needs of different research communities. Examples of Open Science infrastructures include major scientific equipment or sets of instruments, and knowledge-based resources such as collections, journals and open access publication platforms, repositories, archives, and research data[70]. Open Science infrastructures are often the result of bottom-up or community-centred efforts, which are crucial for their long-term sustainability and therefore should guarantee permanent and unrestricted access and reuse rights to the public to the greatest extent possible. One critical component of Open Science Research Infrastructures is the identification of research papers by unique persistent identifiers such as digital object identifiers (DOIs), ORCID IDs, Software Heritage ID (SWID) and research activity identifiers (RaiD) for research projects[71]. They provide essential open and standardized services to manage and provide access, portability, analysis of data, scientific literature, thematic science priorities or community engagement. These include Open Science platforms and repositories for publications, research data and source codes, digital research services and open laboratories.

However, challenges of data sharing within collaborative research projects can occur due to the constraints of sharing and utilizing data which is identified as highly sensitive and carries extra levels of protection. Specifically, these challenges include establishing a regular framework for data protection across the EU, and elsewhere, a lack of consensus on the choice of interoperability standards was reported, and suboptimal data science literacy in the

[65] See chapter 8 of Borgman, C. L. *Big Data, Little Data, No Data: Scholarship in the Networked World*. (The MIT Press, 2015). https://mitpress.mit.edu/9780262529914/big-data-little-data-no-data/ and the French translation : BORGMAN, Christine L. *Qu'est-ce que le travail scientifique des données? Big data, little data, no data.* Nouvelle édition [en ligne]. Marseille : OpenEdition Press, 2020. Disponible sur Internet : <http://books.openedition.org/oep/14692>. ISBN: 9791036565410. DOI: https://doi.org/10.4000/books.oep.14692.

[66] Please see: https://sharing.nih.gov/data-management-and-sharing-policy/about-data-management-and-sharing-policies/data-management-and-sharing-policy-overview for an overview of the NIH Data Management Sharing Policy

[67] NII RDC is a core infrastructure for research data management in Japan, which collaborate with EOSC and other initiatives.

[68] Ruediger, D., MacDougall, R., Cooper, D.M., Carlson, J., Herndon, J., Johnston, L. 2022. Leveraging Data Communities to Advance Open Science- Findings from an Incubation Workshop Series. Ithaka S+R: 1-38.

[69] UNESCO 2022. https://doi.org/10.54677//QZPQ1991

[70] ibid

[71] ibid

health sector was noted[72]. For example, data collection and harmonization across different cohorts may be the solution to reach adequate statistical power to test several hypotheses and control for relevant biases[73]. Proper harmonization allows data to be comparable and, as a final output, will produce reliable and valid data for integrated research examination. Initiatives such as Maelstrom Research[74] have provided data harmonization guidelines that ensure data quality, reproducibility, and transparency of the process. Unfortunately, the process of harmonizing and integrating data from existing projects is often poorly documented and, furthermore, implemented. Human resources and economic efforts should be made available to support the development and implementation of international standards and protocols for harmonization. Future approaches for large-scale harmonization of data are needed to address the current trials caused by tedious manual work.

Recently, initiatives such as to address the inherent regulatory complexities and privacy concerns, use a federated data analysis and machine learning approach[75]. With these approaches, where individual-level data never leave the institution but are analyzed locally. This kind of infrastructure enables conducting participant-level analysis without revealing/exchanging the participant-level data, as only the computed parameters describing the entire datasets are communicated to the analyst.

To date, research funding tools have adhered to the classical format in terms of calls for standard procedures and calls for grant competition by different competitive groups. However, the pandemic has shown limitations of this format, which can be slow to generate evidence and knowledge. In times of dynamic change, there is a need for producing evidence “on the fly” to enable policy-makers to make scientifically informed decisions. This includes establishing research calls including incentives for transdisciplinary groups and data owners; geographic origin could be interpreted more broadly by encouraging and facilitating participation of different European regions instead of a single partner institution could be set aside from Member States; and dedicated funding could be used for coordination and harmonization, of multicentric studies[76]. These experimentations should then be evaluated by research on research to know more about their impacts, difficulties and lessons learned. For example, Project Bamboo, a humanities cyberinfrastructure initiative funded by the Andrew W. Mellon Foundation, aimed to augment art and humanities research through the development of shared technology services. The planning phase brought researchers, librarians, and IT staff together to build a scholarly community of future developers and users of Bamboo’s services. However, Bamboo struggled to be fully implemented because the project struggled to define itself clearly. The early emphasis on a service-oriented design approach intrigued librarians and IT technologists, while many researchers felt that their needs lay elsewhere entirely. The technical team and the research team had very different

[72] Tacconelli, E. *et al.* Challenges of data sharing in European Covid-19 projects: A learning opportunity for advancing pandemic preparedness and response. *The Lancet Regional Health - Europe* **21**, 100467 (2022).
[73] ibid
[74] Fortier, I. *et al.* Maelstrom Research guidelines for rigorous retrospective data harmonization. *Int. J. Epidemiol.* dyw075 (2016) doi:10.1093/ije/dyw075.
[75] Tacconelli, E., et al. 2022. Challenges of data sharing in European Covid-19 projects: A learning opportunity for advancing pandemic preparedness and response. The Lancet Regional Health - Europe 21: 1-9 https://doi.org/10.1016/j.lanepe.2022.100467
[76] ibid

perspectives on what was needed, which led to frustration and communication failure from both sides[77].

2.3 Data Sharing and Research Cultures

- *The Complexity of the Data Sharing Ecosystem*

Research communities determine what is useful research data for their discipline and/or region. Research cultures vary across these research environments, defined as they are by their respective research discipline and stakeholders. Policy deliberations need to represent the diversity of data cultures.

The complexity of the data sharing ecosystem is complicated by overlapping roles among researchers, funders, publishers, platforms, and Research Performing Organizations (RPOs). This complexity underpins the growing appreciation, across research communities and society, of open and transparent science/research practices. Despite increased awareness about the value of open and transparent practices, inconsistent data sharing persists. Perceived barriers may explain this discrepancy. Knowledge gaps can include processes for the analysis, management, stewardship, and sharing of data. Related uncertainty can lead to subjective value judgments about data. Moreover, perceptions of value are also related to different data practices, to data as standalone output, and to the reuse potential of data. Many researchers are reluctant to share their data because they are afraid of lack of academic recognition and there are many researchers trying to bargain data sharing for co-authorship, even when it is inappropriate. Reuse concerns also apply to sensitive content (safety, security) and inappropriate use (misuse, misinterpretation) of research data. Understanding sharing rights is complicated by collaboration, funding, affiliation, and publication. Concerns about accessibility, interoperability, and sustainability may deter researchers from expending the time, effort, and cost required to store data for reuse.

- *Disincentives*

A lack of resources, recognition, and rewards disincentivize[78] open and transparent science/research practices. Data curation needs to be recognized as an integral part of the research process[79]—and resourced, accordingly, with sufficient time and funding.

- *Systemic Challenges*

Research institutions are either slow to implement open data policies (analysis paralysis, insufficient funding, not discipline-specific), or fail to consistently monitor and enforce compliance. Publishers are also accountable as many publications include partial data or no data, without penalty for including data that is not openly shared. While technology (repositories) can support open and transparent research, it can also impede discoverability,

---

[77] Dombrowski, Q. What Ever Happened to Project Bamboo? *Lit Linguist Computing* fqu026 (2014) doi:10.1093/llc/fqu026.

[78] D. G. E. Gomes et al. (2022) Why don't we share data and code? Perceived barriers and benefits to public archiving practices. *Proc. R. Soc. B.* **289:** 20221113. https://doi.org/10.1098/rspb.2022.1113

[79] Leonelli, S., Spichtinger, D. and Prainsack, B. (2015) Sticks AND Carrots: Incentives for a Meaningful Implementation of Open Science Guidelines. Geo, 2: 12–16 doi: 10.1002/geo2.2

tools and repositories can be far from being FAIR friendly, and there are also strong limitations of data search engines.

These findings suggest areas for further research, as outlined below, which would also provide additional evidence of the benefits of data sharing to reuse, alongside best practices[80] across different research environments[81].

- *Research Agenda*

Research is needed to improve data management, discoverability, and reuse throughout the Open Science life cycle. Solutions will include measures to prevent the loss of either standard data (preservation strategies) or data in a format less readily shared—and thereby temper the bias in the data ecosystem.[82] Research will aim to address the complexity around common identifiers (global / PIDs) and vocabularies to improve circulation and to facilitate reuse (FAIR Data Principles).[83] [84]

An impact assessment of data sharing could begin with the practices of existing professional groups (4TU.ResearchData, Research Data Alliance (RDA), The Committee on Data of the International Science Council (CODATA), Global Biodata Commission (GBC)[85], Center for Open Science (COS[86]). Data-dependent organizations (food, biotech, pharmaceutical industries) would need to be engaged to ensure the depth and breadth of insight. In both cases, data sharing is recognized as a means to an end, which necessarily centres reuse in any impact assessment. However, to date, data reusing studies are rare and fragmented. Research is needed to establish a knowledge base around data sharing, reuse, and citation practices across disciplines and career stages. Qualitative aspects refer to the support researchers need to fulfil their role relative to data collection such that sensitivity and privacy issues, plus the potential for reuse, are addressed.

Related recommendations could be validated and refined through a feedback loop across research communities. This approach would see researchers using Research on Research methodology to evaluate proposed solutions based on their implementation: whether (how and why) solutions were adopted, adapted, circumvented, or even rejected. Possible policy recommendations, more overarching than research on research-specific, are outlined below.

---

80 Neylon, Cameron. (2017). Knowledge Exchange Approach towards Open Scholarship. https://doi.org/10.5281/zenodo.826643

81 Leonelli, S. (2017) Global Data Quality Assessment and the Situated Nature of "Best" Research Practices in Biology. Data Science Journal 16(32): 1-11. DOI: 10.5334/dsj-2017-032

82 *Ibid.*

83 Pasquetto, I.V., Randles, B.M. and Borgman, C.L., 2017. On the Reuse of Scientific Data. Data Science Journal, 16, p.8. DOI: http://doi.org/10.5334/dsj-2017-008

84 Pasquetto, I.V., Borgman, C.L., & Wofford, M.F. (2019). Uses and Reuses of Scientific Data: The Data Creators' Advantage. Harvard Data Science Review, 1(2). https://doi.org/10.1162/99608f92.fc14bf2d

85 Global Biodata Coalition: https://globalbiodata.org/

86 The Centre for Open Science published the Transparency and Openness Promotion / TOP Guidelines (2015).

2.4 Code and Software

Whereas the question of code and software has not been addressed during the workshops of the Research on Research Sub-Working Group, comprehensive Open Science can only be achieved when codes and software[87] associated with research are properly opened under open licence (FOSS: Free and Open Source Software).

There is a growing appreciation for the value of code - be it programming scripts or other research software - to process research data, and to automate, share, reproduce, and reuse research methods.[88] Research software has also become a key output of research projects.[89] Likewise, extending citation to software supports reproducibility and reliability, in addition to recognition. Funders can play a critical role in addressing the urgent need to sustain research software[90], and to invest in the specialists (research software engineers) who develop and maintain it. For example, the vision of the Research Software Alliance (ReSA)[91] is to recognize those who develop and maintain research software. Research software is a key driver of innovation and the economy. ReSA focuses on software that is developed to serve a research purpose, rather than the standard software that some researchers also use for research purposes [92]. Related issues include contributor community development, governance, diversity, and inclusion.[93] The Software Sustainability Institute has aided in the development and implementation of cultivating a more sustainable research software hub to enable world-class research. The Institute employs experts with a breadth of experience in software development, training, program management, and community engagement, and is forefront in driving research software policy and instigating projects researching research software.[94] There is also a need to consider not only comprehensive software, but also small pieces of code and scripts that are used in the research data pipeline and have a role in the fine-tuning of data which is needed before extracting knowledge from data.

Open source software was created in the 1980s[95], far before the Open Science debates even started. This may be the reason Open Source Software was paradoxically mainly forgotten during the first years of Open Science policies. While focusing on open access and FAIR data, Open Science policies were nearly open source software blind, and the "*unsung heroes of*

---

[87] Scholarly Infrastructures for Research Software: Report from the EOSC Executive Board Working Group (WG) Architecture Task Force (TF) SIRS. Publications Office of the European Union, 2020. Publications Office of the European Union, https://data.europa.eu/doi/10.2777/28598.
[88] Gomes Dylan G. E. et al. 2022 Why don't we share data and code? Perceived barriers and benefits to public archiving practices. Proc. R. Soc. B. 289 20221113 https://doi.org/10.1098/rspb.2022.1113
[89] Netherlands eScience Centre: Practical Guide to Software Management Plans https://www.escienceCentre.nl/national-guidelines-for-software-management-plans/
[90] Joris van Eijnatten, et al. (2022). Amsterdam Declaration on Funding Research Software Sustainability (0.2). Zenodo. https://doi.org/10.5281/zenodo.7330542
[91] https://www.researchsoft.org/about-resa/
[92] ibid
[93] Strasser C, et al. (2022) Ten simple rules for funding scientific open source software. PLoS Comput Biol 18(11): e1010627. https://doi.org/10.1371/journal.pcbi.1010627
[94] https://www.software.ac.uk/
[95] Tozzi, C. For Fun and Profit: A History of the Free and Open Source Software Revolution. (The MIT Press, 2017). doi:10.2307/j.ctt1t88w3c.

science software" were left behind[96]. This explains the unsophisticated way we address this issue, with less research, less funding mechanism, and less mature services. Some major initiatives exist that try to overcome this problem, by creating open research software communities[97], by archiving the whole of the open source software patrimony of the earth[98], or by publishing dedicated conferences[99] and journals[100] … But most of the time, software is confused with research papers or research data, though research software is of a different nature. Consequently, famous research mistakes in important publications reveal lack of knowledge in basic statistics and, moreover, lack of capacities in the data pipeline. A famous example in economy has been detected, with false conclusions driven by mistakes in Microsoft Excel, but this is only the visible part of the iceberg[101].

Software and code sharing, along with dedicated processes for worldwide collaboration on its elaboration, should become a specific part of Open Science policies. Software should not be confused with publication or with data, as they are different by nature. Indeed, software should be considered as a first-class research object as it represents a game changer in research. This is the reason France has created Open Science awards dedicated to research software[102].

An impact assessment of research software and code sharing could get inspired by the decades' long-standing communities that have been built around free software such as, for instance, the Linux Foundation[103] which coordinates hundreds of Open Source projects with hundreds of thousands of developers worldwide. These high figures are the symbol of a dynamic worldwide open source software ecosystem, which may hide its fragility, especially when it comes down to specialized ecosystems such as that of research software. Indeed, the culture change toward free and open source software has yet to happen, in the research

[96] Singh Chawla, D. The unsung heroes of scientific software. Nature 529, 115–116 (2016). Doi: 10.1038/529115a

[97] https://www.researchsoft.org/

[98] About https://www.softwareheritage.org/, see: Cosmo, R. D. & Zacchiroli, S. Software Heritage: Why and How to Preserve Software Source Code. in iPRES 2017 - 14th International Conference on Digital Preservation, Sep 2017 (2017). https://hal.science/hal-01590958/
Pietri, A., Spinellis, D. & Zacchiroli, S. The Software Heritage Graph Dataset: Large-scale Analysis of Public Software Development History. in Proceedings of the 17th International Conference on Mining Software Repositories 1–5 (Association for Computing Machinery, 2020). doi:10.1145/3379597.3387510.

[99] Many computer sciences conferences, like the International Conference on Software Engineering, have Artifact Evaluation Committees that evaluate the software associated to research publication. For example : https://2021.icse-conferences.org/committee/icse-2023/icse-2023-artifact-evaluation-artifact-evaluation

[100] IPOL is a research journal of image processing and image analysis which emphasizes the role of mathematics as a source for algorithm design and the reproducibility of the research. Each article contains a text on an algorithm and its source code, with an online demonstration facility and an archive of experiments. Text and source code are peer-reviewed, and the demonstration is controlled. IPOL is an Open Science and Reproducible Research journal. https://www.ipol.im/
Journal of Open Source Software. https://joss.theoj.org/

[101] Peng, R. The reproducibility crisis in science: A statistical counterattack. Significance 12, 30–32 (2015). Doi: 10.1111/j.1740-9713.2015.00827.x

[102] https://www.ouvrirlascience.fr/open-science-free-software-award-ceremony/

[103] https://www.linuxfoundation.org/

community as elsewhere, with the notable exception for some specific disciplines[104]. This results in open source projects often relying on fragmented small fragile teams. Research software here mimics what is happening for the generic use of Open Source software. Though Open Source software is now decades old, the research community has yet to take the full advantage of it.

Aside from this crucial efficiency issue is another one, which concerns the very process of science itself, any software used in a scientific process should be opened to allow understanding, traceability, and replicability[105], of the research process itself[106]. As the data is the raw result and the scientific paper is its analysis, software is on the path from the former to the latter. Replicability and confidence on the research process is at stake here, research software should be precisely preserved, opened and made citable as other research products[107] [108].

Research on the research process and protocols itself is thus needed to understand what is hindering this potential worldwide collaboration and required piece of scientific evidence. It seems particularly important because the tools that technically enable them to have already been around for years[109]. A second important matter that has yet to be clarified is the way to provide a secure economical model for research Open Source software, as it has been found for others of its uses[110], to remove the present fragility of the system.
An example, the French parliament has asked for a study on the French production and valorization of research software, either open source or not, to gain the knowledge that is missing to adjust the French policies in the field[111].

2.5 Reproducibility of Research and Trust

The report on reproducibility of scientific results in the EU[112] defines reproducibility as the possibility for scientists to obtain the same results as the originators of some specific scientific

---

104 Fortunato Laura and Galassi Mark, 2021, The case for free and open source software in research and scholarship, Phil. Trans. R. Soc. A.3792020007920200079, http://doi.org/10.1098/rsta.2020.0079
105 See for instance the replicability stamps initiative in Computer Graphics: Graphics Replicability Stamp Initiative
106 Clément-Fontaine, M., Cosmo, R. D., Guerry, B., Moreau, P. & Pellegrini, F. Encouraging a wider usage of software derived from research. 6 p. (Comité pour la science ouverte, 2019). doi:10.52949/4.
107 Roberto Di Cosmo and Stefano Zacchiroli. 2017. Software Heritage: Why and How to Preserve Software Source Code. In Proceedings of the 14th International Conference on Digital Preservation, iPRES 2017.
108 Software Preservation: A Stepping Stone for Software Citation. https://doi.org/10.25815/0ZBH-2W14
109 As examples: the Free and Open Source git and gitlab enable worldwide traceability and collaboration while Software Heritage allows preservation and citeability. Open Archives with dedicated document types can also play a role in dissemination: the French Open Archive HAL is for instance interconnected with Sofware Heritage.
110 See for instance the above mentioned gitlab society, which is based on the Open Source git and that aims for Nasdaq: Boorstin, Julia; Fortt, Jon (October 14, 2021). "GitLab goes public on Nasdaq a $10 billion IPO". *CNBC TechCheck*. Could also be mentioned Canonical, the well-known maker of the Ubuntu linux distribution.
111 Article 163 of the French Law n° 2022-217 of February 21, 2022
112 Assessing the reproducibility of research results in EU Framework Programmes for Research - Publications Office of the EU (europa.eu)

findings. The US National Academies of Sciences, Engineering and Medicine report on reproducibility and replicability in science[113] define reproducibility as obtaining the same results having used the same (original) data, code, methods, and analysis. As computing and data play an increasingly important role across all of science and engineering, ensuring the reproducibility of software and data-enabled research is important; it is a prerequisite to replicability to improve trustworthiness. For example, the research community is seeing a growing number of "fake papers," often enabled by artificial intelligence-based tools, which manage to get past flawed review processes[114]. Moreover, review and publication processes are evolving to include a trend toward greater unreviewed preprint release, and lack of publication of null results, make it difficult to evaluate the statistical significance of positive published results and the potential lack of replicability. In fact, transparency and trust are at stake, where many areas of scientific research are hard to reproduce and there are limited incentives to run replication studies.

In general, policies should seek to improve reproducibility and normalize full reporting of all results, including data and code sharing. However, policy actions on reproducibility depend on mutual cooperation of a range of stakeholders, specifically research funders, publication companies, and Research Performing Organizations (RPOs). Protocol sharing will help, in addition to Data Management Plan. For example, Protocols.io[115] is a platform where researchers could organize and collaborate with others to improve their research methodology, facilitate teaching and accelerate progress in data reproducibility across most research disciplines.

As a collaboration tool, The Open Science Framework enables connections to many research tools, in order to streamline processes and effectively share methodology steps within a research project to eliminate data silos and information gaps.

Publishers could complement the work of reviewers with basic reproducibility checks and data/code inspection by adapting to publication formats that support reproducibility. In addition, publishers could offer registered reports where the protocol is peer reviewed and provisionally accepted for publication before the data are collected and analyzed[116]. As a certification tool for code and data in Economy and Management, the Cascad reproducibility certification[117] attests that the numerical results (tables and figures) reported in a scientific publication can be reproduced using an identifiable set of numerical resources (code and/or data) made available by the authors of this publication. This certification involves a rigorous evaluation process conducted by a referee under the supervision of an editor, at the end of

[113] Reproducibility and Replicability in Science | National Academies
[114] Parashar, M., Heroux, M.A., Stodden, V. 2010. Research Reproducibility. Computer. 55: 16-18 doi: 10.1109/MC.2022.3176988.
[115] Teytelman, L., Stoliartchouk, A., Kindler, L. & Hurwitz, B. L. Protocols.io: Virtual Communities for Protocol Development and Discussion. *PLOS Biology* **14**, e1002538 (2016).protocols.io is an example of a secure platform for developing and sharing reproducible methods: https://www.protocols.io/
[116] Chambers, C. D. & Tzavella, L. The past, present and future of Registered Reports. Nat Hum Behav 6, 29–42 (2022). https://doi.org/10.1038/s41562-021-01193-7
[117] https://www.cascad.tech/

which a certification rating is delivered. This is not only important with open data, but also with confidential data[118].

Open peer review could also be part of the solution. For example, BMC Medicine has an open access, transparent peer-reviewed system[119], where, if the article is published, the reviewer reports are published online alongside the researcher's article. The benefit of open reviewers' reports is that it can serve an educational purpose in helping facilitate training and research into peer review. But open peer review is a complex approach, with identified issues, and with very diverse configurations[120] (22 different approaches have been identified by research on research)[121]. Open peer review is probably an item in the Open Science toolbox, but it cannot be generalized widely without precaution and variations. There is still room for experimentation and lessons to be learned.

RPOs could train research staff on reproducible research and fund training as part of early career development. It has been highlighted that work reproducibility should put early career researchers in a better career position. For example, reproducibility has been raised as a major concern in software development, and in industrial biomedical and pharmaceutical research. Early adoption of open and reproducible methods which can be evidenced, will open opportunities for collaborations in consortia or research networks and connect with others to build a local Open Science community[122].

Dedicated research programs are needed to evaluate the reproducibility of research, and to identify key findings that support reproducible research and non-reproducible research. Reproducible research should be encouraged, and research communities should promote incentives for all researchers to publicly discuss their methods, but also the ways in which they learn from unexpected and incongruent findings[123].

2.6 Academic, Economic and Societal Impacts

- *Academic Impacts*

Most of the texts published about Open Science focus on the quantification of the openness of the production, which is already something difficult to assess. Nevertheless, the "open access citation advantage" is identified in over one hundred publications and the majority conclude that there is a positive impact on citation when publications are open[124]. However,

[118] Pérignon, C., Gadouche, K., Hurlin, C., Silberman, R. & Debonnel, E. Certify reproducibility with confidential data. Science 365, 127–128 (2019).
[119] https://bmcmedicine.biomedcentral.com/submission-guidelines/peer-review-policy
[120] For example, post publication open peer review on books: Dandieu, C. & Consortium, H. Report on Post-Publication Open Peer Review Experiment. (2019) doi:10.5281/zenodo.3275651.
[121] Ross-Hellauer, T. What is open peer review? A systematic review. Preprint at https://doi.org/10.12688/f1000research.11369.2 (2017).
[122] Allen, C., Mehler, D.M.A. 2019. Open science challenges, benefits, and tips in early career and beyond. PLOS Biology 17 (12):1-14 https://doi.org/10.1371/journal.pbio.3000587
[123] Leonelli, S. (2018). Re-Thinking Reproducibility as a Criterion for Research Quality. Research in the History of Economic Thought and Methodology. 36: 129-146 DOI: 10.1108/S0743-41542018000036B009
[124] Piwowar, H. *et al. The State of OA: A large-scale analysis of the prevalence and impact of Open Access articles*. https://peerj.com/preprints/3119 (2017) doi:10.7287/peerj.preprints.3119v1.

the scholarly communication environment is dynamic and already the intensive usage of pirated papers via Sci-Hub[125] suggests fades in impact. Although there are fewer studies about the impact of having open research data attached to a research paper, all identify a clear open research data citation advantage[126]. In biomedical research, which has the most extensive study based on more than 500,000 papers, an increase of 25% of citations is reported when an article is enriched by open data[127]. There is a clear need for more research like this to understand open data citation advantage, both longitudinal within a discipline and within other disciplines. Moves towards data (and software) citation should make the impact of data, and the crediting of data creators, curators, and repositories more tractable[128]. However, although the technologies are readily available, widespread changes in citation practice have been slow to take hold[129]. Meanwhile, data providers are forced to use indirect measures to build impact measures that justify their continued sustainability[130]. Understanding why adoption of seemingly straightforward changes in scholarship practice is so slow is essential.

- *Societal Impacts*

In a world of closed content, users must provide identification and credentials, so it is straightforward to track usage. However, in a world of open access to publications, data, and code, most of the users have access without authentication, so the knowledge of the users is very limited. The limited research suggests that there is probably a high societal impact of Open Science. Juan Pablo Alparin's study of the two major open access platforms in Latin America stands out. Redalyc and Scielo host more than two thousand open access journals and gather millions of visits each month. "*By way of summary, there can be said to be three main types of readers: Students, Academics, and the Public (Table 4.1). This composite summary of the article readership is a key result, challenging the assumption that academics write for academics (and assumed by academics as much as anyone). It challenges assumptions about students reading original research (rather than textbooks). It challenges*

McKiernan, E. C. *et al.* How open science helps researchers succeed. *eLife* **5**, e16800 (2016). https://doi.org/10.7554/eLife.16800

[125] Maddi, A. & Sapinho, D. On the culture of open access: the Sci-hub paradox. Preprint at https://doi.org/DOI: 10.21203/rs.3.rs-2357492/v1 (2022).

[126] Piwowar, H. A. & Vision, T. J. Data reuse and the open data citation advantage. *PeerJ* **1**, e175 (2013).

Jon Sears. Data Sharing Effect on Article Citation Rate in Paleoceanography. (17:50:19 UTC).

Zhang, L. & Ma, L. Does open data boost journal impact: evidence from Chinese economics. *Scientometrics* (2021) doi:10.1007/s11192-021-03897-z.

Henneken, E. A. & Accomazzi, A. Linking to Data - Effect on Citation Rates in Astronomy. *arXiv:1111.3618 [astro-ph]* (2011).

Piwowar, H. A., Day, R. S. & Fridsma, D. B. Sharing Detailed Research Data Is Associated with Increased Citation Rate. *PLOS ONE* **2**, e308 (2007).

[127] Colavizza, G., Hrynaszkiewicz, I., Staden, I., Whitaker, K. & McGillivray, B. The citation advantage of linking publications to research data. *PLOS ONE* **15**, e0230416 (2020).

[128] Cousijn, H., Kenall, A., Ganley, E. *et al.* A data citation roadmap for scientific publishers. *Sci Data* **5**, 180259 (2018). https://doi.org/10.1038/sdata.2018.259

[129] Buneman et al. u, Why data citation isn't working, and what to do about it, *Database*, Volume 2020, 2020, baaa022, https://doi.org/10.1093/databa/baaa022

[130] Aurélie Névéol, W. John Wilbur, Zhiyong Lu, Extraction of data deposition statements from the literature: a method for automatically tracking research results, *Bioinformatics*, Volume 27, Issue 23, 1 December 2011, Pages 3306–3312, https://doi.org/10.1093/bioinformatics/btr573

the assumption that the public has no interest or capacity for research (although this study does not investigate the nature of this public use)."[131]

Table 4.1: Summary of Users Reached

Type of Reach	Approximate Proportion	
	SciELO	RedALyC
Students	50%	55%
Staff (including faculty)	25%	22%
Professional	20%	17%
Personal	9%	6%

Note: This table represents a composite of the results presented in this chapter, derived from the author's interpretations of the data. As such, the numbers should be treated only as approximations, and the percentages should not be expected to add to 100%.

In France, early results of the ongoing research into the uses of OpenEdition show that this French platform for humanities and social sciences receives more than one hundred million visits a year[132]. This number is remarkable when one considers the "niche" that many OpenEdition journals cover. Moreover, beyond the quantification of users, the qualification of users provides findings that could be counter-intuitive for a highly specialized platform dedicated to history, geography, philosophy and other humanities and social science disciplines. The portal is used by companies in the aeronautics sector, banking, insurance, automotive, energy and by the media. In the public sector, beyond university campuses, there are also identified users in central and territorial administrations[133]. Research in the Netherlands considering publications in Science, Technology and Mathematics (STM) shows that *"approximately 40% of readers surveyed for this analysis on Springer Nature websites were classified as non-academic audiences, including 15% "Halo" users (likely to be reading research for professional purposes but not conducting or publishing research themselves) and 28% "General" users (likely to be reading out of personal or professional interest but outside of a role where conducting, publishing or citing research is typical)."* There is clear evidence that this kind of readership is excluded by paying walls[134]. Other research gathered evidence of the impact of consensus reports of the National Academies of Science, Engineering, and Medicine (NASEM)[135]. Before downloading, the users were asked to report why they downloaded the reports. By analyzing the 1,6 million downloads left by these National Academies reports, Diana Hicks *et al* found that *"half of reported use to be academic—research, teaching, or studying. The other half reveals adults across the country seeking the*

---

[131] Alperin, J. P. The public impact of Latin America's approach to open access. https://purl.stanford.edu/jr256tk1194
[132] OpenEdition reports provide many figures: https://www.openedition.org/25480?lang=en
[133] For early results, see Dacos, M. Des nains sur les épaules de géants : ouvrir la science en France. *Revue Politique et Parlementaire* (2019). https://hal.science/hal-02366604
[134] Wirsching, H. *et al.* Open for all: exploring the reach of open access content to non-academic audiences. (2020) doi:10.5281/zenodo.4143313.
[135] Hicks, D., Zullo, M., Doshi, A. & Asensio, O. I. Widespread use of National Academies consensus reports by the American public. *Proceedings of the National Academy of Sciences* **119**, e2107760119 (2022). https://www.pnas.org/doi/full/10.1073/pnas.2107760119

highest-quality information to improve how they do their job, to help family members, to satisfy their curiosity, and to learn." These results "*establish the existence of demand for high-quality information by the public and that such knowledge is widely deployed to improve provision of services. Knowing the importance of such information, policy-makers can be encouraged to protect it.*" As a conclusion, "*The picture contrasts starkly with the dominant narrative of a misinformed and manipulated public targeted by social media.*" Some figures talk even about "open access best sellers," which seems counter-intuitive because open access books are not sold, but it makes sense when a newly born open access university press reaches one million downloads after only 3 years of existence[136]. The societal impact of open data is easy to understand and has been advocated[137] in many opinion and commentary pieces, yet we still lack systematic and evidential study.

- Economic Impacts

An important example of academic, economic, and societal impact associated to opening research and data is the Human Genome Project, a large Open Science project that has created a tremendous amount of economic value, providing, according to some estimates, around 1 trillion U.S. dollars to the U.S. economy in 2013[138].

This amount may be underestimated since the associated report only focuses on human applications. If one extrapolates from that example, especially in the context of a project that ceased operations, we can only imagine how the creation of the World Wide Web – an open infrastructure that greatly support Open Science, has incalculable positive economic repercussions - and this, notably in terms of productivity growth. It is a creation that has been brought to the whole society and economy by a decision of CERN in 1990. Originally, the Web was invented to meet the demand for automated information-sharing between institutes around the world and scientists in universities[139]. Some research has also provided some estimations of the impact of Open Science infrastructures dedicated to research data, such as EMBL-EBI (*European Molecular Biology Laboratory - European Bioinformatics Institute*), a European research infrastructure being part of the ESFRI roadmap. The most direct measure of the value is the time researchers spend using EMBL-EBI data resources. This added up to more than 140 million hours during 2020, equivalent to an estimated £5.5 billion. Another key value of open data lies in the deduplication of research effort. An impressive 58% of survey respondents stated that they could not have collected the last dataset they used themselves, nor obtained it elsewhere. If the time saved by users from not having to create or, more precisely, recreate the data enabled more research to be done, it could be worth almost £6

---

[136] McKenzie, L. Open-Access Best Sellers. *Inside Higher Ed* https://www.insidehighered.com/news/2018/06/01/most-popular-open-access-monographs (2018).
[137] Geoffrey Boulton, Michael Rawlins, Patrick Vallance, Mark Walport Science as a public enterprise: the case for open data, The Lancet 2011 DOI:https://doi.org/10.1016/S0140-6736(11)60647-8

[138] Wadman, M. Economic return from Human Genome Project grows. *Nature* (2013) doi:10.1038/nature.2013.13187.
[139] "A short history of the web," CERN, https://home.cern/science/computing/birth-web/short-history-web (accessed, 1 February 2023).

billion annually[140]. This kind of research should and could be applied to other fields, to identify differences and specificities.

The European Commission has funded research[141] about the cost-benefit of not having FAIR research data. They make an estimation based on the cost related to research data collection and processing during the whole research process (collection of data, pre-processing and data cleaning, integration of data, analysis of data, registration and publication, peer review). During all these stages, the lack of FAIR data cost at the European level $10.2 billion/year. They acknowledge the fact that changing data into FAIR data is expensive, but their estimation is that the cost is lower than the benefits. This very high-level research is a first step but needs to be confirmed and refined by more research. The European Commission has also funded PathOS, a project which seeks to "*better understanding and measuring* [of] *Open Science impacts and their causal mechanisms*"[142]

Formally, there are a lot of data sharing statements attached to research publications that have a clear commitment to data sharing, specifically when the data cannot be open to everybody, but can be shared under a controlled environment to trusted and skilled people. These precautions are dedicated to assuring the confidentiality of sensitive data. Some recent research[143] shows that there is a strong difference between the willingness to share that is stated in the data sharing statements and the facts. Researchers tested a sample and showed that the statements are not reliable: 93% of the researchers either did not respond to the data request or declined the request. This shows another side of the interest of research on research: being able to compare the official discourse and real daily practices is necessary to achieve the whole Open Science journey. Otherwise, wrong statements could be taken literally, and the Open Science policies efforts could be stopped too early or be mis-calibrated.

- *Open Science Policies Impacts*

Research on Research indicates that Open Science policies are needed, but their impact may vary a lot, depending on numerous factors. For example, open access mandates do not always produce the same results[144]. Having a clearly stated mandate to make all the grant-funded publications open is, indeed, not enough to have a mandate compliance close to 100%. There are numerous obstacles that the policy must avoid. For example, a mandate without a

---

[140] Institute, E. B. Open data on the rise: the value of EMBL-EBI data resources. https://ebi.ac.uk/about/news/announcements/value-and-impact-emblebi-2021/.
[141] *Cost-benefit analysis for FAIR research data - Publications Office of the EU*. https://op.europa.eu/en/publication-detail/-/publication/d375368c-1a0a-11e9-8d04-01aa75ed71a1/language-en.
[142] https://pathos-project.eu/
[143] Gabelica, M., Bojčić, R. & Puljak, L. Many researchers were not compliant with their published data sharing statement: mixed-methods study. *Journal of Clinical Epidemiology* **0**, (2022).
[144] Vincent-Lamarre, P., Boivin, J., Gargouri, Y., Lariviere, V. & Harnad, S. Estimating Open Access Mandate Effectiveness: The MELIBEA Score. Preprint at https://doi.org/10.48550/arXiv.1410.2926 (2016).
Simard M-A, Ghiasi G, Mongeon P, Larivière V (2022) National differences in dissemination and use of open access literature. PLoS ONE 17(8): e0272730. https://doi.org/l10.1371/journal.pone.0272730
Huang, Chun-Kai (Karl), et al. 'Evaluating the Impact of Open Access Policies on Research Institutions.' ELife, edited by Julia Deathridge et al., vol. 9, Sept. 2020, p. e57067. eLife, https://doi.org/10.7554/eLife.57067.

clear and simple process for the grant leader to be compliant has been proven to be less effective. Furthermore, a funder-imposed data publication requirement seems to be no sufficient to inspire data sharing[145]. First, the different disciplinary cultures lead to vastly different reactions, understandings, and capacities of application of the data sharing mandate. Secondly, the success is also connected to the dedicated resources the funder invests in enforcing compliance with data requirements, providing data-sharing tools and technical support to awardees. Thirdly, finally, some mandates come without strict consequences for those who ignore data sharing preconditions. A deeper knowledge of the success stories (and failure) and of the most efficient actions to transform a policy into reality is needed in the coming years. This would accelerate the Open Science journey.

2.7 The Case of Research on Research of Genomics during COVID 19

Research commissioned by the UK Government's Department for Business, Energy & Industrial Strategy (BEIS) has addressed the question of viral genomic data sharing during the COVID-19 pandemic[146]. Despite many official declarations stating that the countries would share all their data, publications and results in a timely and open manner, the COVID-19 pandemic exposed fundamental weaknesses in Open Science maturity and in pandemic preparedness across global health systems. On the topic of SARSCOV-2 genome sharing, this can be attributed to a lack of sustained investment across the sequencing pipeline. The inconsistent access to the elements outlined above has led to variability in sequencing capability, availability, and quality. In turn, we have been left with gaps in the global knowledge base around how COVID-19 mutated and spread. The research brings together an evidence base of 295 sources, the views of twenty-four interviewees, and insights from eighteen international peer reviewers. It finds five key lessons:

- The long-term investment made in developing international standards and infrastructure for data sharing in genomics has been repaid many times over when this data became central to the pandemic response.
- Effectively tackling global crises like the COVID-19 pandemic requires representative data from all parts of the world. Not all countries and regions have sufficient data-generating capacity or trained human resources to collect, disseminate, and analyze these data.
- If rapid and open data sharing is to be encouraged, the contributions of data generators need to be recognized and rewarded.
- Established norms around the timing and extent of data sharing were in many cases set aside in the COVID-19 crisis, with multiple actors recognizing that the immediate availability of data to a broad set of users was paramount. Yet the pandemic also provides an opportunity to reassess these established norms, whose deficiencies were in some cases sharply exposed. Ongoing efforts to reform academic incentives must

[145] Couture, J. L., Blake, R. E., McDonald, G. & Ward, C. L. A funder-imposed data publication requirement seldom inspired data sharing. *PLOS ONE* **13**, e0199789 (2018). DOI: 10.1371/journal.pone.0199789

[146] *Intelligent open science: viral genomic data sharing during the COVID-19 pandemic*. 69 https://www.gov.uk/government/publications/intelligent-open-science-viral-genomic-data-sharing-during-the-covid-19-pandemic (2022).

be accompanied by corresponding work to incentivize sharing by public health actors, with strengthened expectations for data-sharing by all parties in an emergency context.

- This study has exposed divergent perspectives within and between the research and public health communities on the merits of open and controlled models of access to genomic viral data. Fully open-access infrastructures for data sharing offer demonstrably greater benefits than controlled access repositories in terms of data re-use and integration, but these benefits cannot be realized in practice unless these infrastructures are accompanied by a transparent and globalized approach to funding, governance and benefits sharing. Proponents of open-access infrastructures must also give greater consideration to mechanisms for incentivizing and crediting data deposits and enabling the creation of high-quality metadata. An intelligent approach to Open Science means moving beyond our existing data-sharing paradigms to be better prepared for future emergencies.

Extending the subject to other types of data, several studies suggest that the pandemic has not significantly transformed sharing practices, apart from the partial development of pre-publications[147]. This research does not, however, allow for definitive conclusions: more in-depth studies are needed. However, it does allow us to identify a major problem: a global pandemic does not seem sufficient to transform data and code-sharing practices[148]. This is a particularly strong confirmation that the transition to Open Science is not just a matter of political will, or even collective will, and that it requires complex systemic changes in all the layers of the academic ecosystem, which is a sophisticated arrangement of culture and beliefs, rights and regulations, technologies and standards, infrastructures, and “know-how.”

2.8 Costs of Open Science

- *Cost of Wasted Research: Costs of Closed Science*

It is impossible to discuss the economic impact of Open Science without discussing the topic of wasted research, mostly due to closed approaches. The basic assumption is that research badly shared, for example not published because of negative results, or published with too few details and data associated, lead to an expansive waste of research, meaning a waste of knowledge advances, leading to duplication of research, slowing down the pace of findings, and wasting research budget because of the issue of lack of publicity of results. There are already strong findings about that issue: “*However, protocols, full study reports, and participant-level datasets are rarely available, and journal reports are available for only half of all studies and are plagued by selective reporting of methods and results. Furthermore,*

[147] Waltman, L. *et al. Scholarly communication in times of crisis: The response of the scholarly communication system to the COVID-19 pandemic*. https://rori.figshare.com/articles/report/Scholarly_communication_in_times_of_crisis_The_response_of_the_scholarly_communication_system_to_the_COVID-19_pandemic/17125394/1 (2021) doi:10.6084/m9.figshare.17125394.v1.
Fraser, N. et al. *The evolving role of preprints in the dissemination of COVID-19 research and their impact on the science communication landscape*. PLOS Biology 19, e3000959 (2021).
[148] Larivière, V., Lebaron, F., Vincent-Lamarre, P. & Larregue, J. COVID-19: Where is the data? *Impact of Social Sciences* https://blogs.lse.ac.uk/impactofsocialsciences/2020/11/30/covid-19-where-is-the-data/ (2020).

information provided in study protocols and reports varies in quality and is often incomplete. When full information about studies is inaccessible, billions of dollars in investment are wasted, bias is introduced, and research and care of patients are detrimentally affected."[149] Research in specific fields, such as lung cancer, systematic reviews fail to provide a complete and up-to-date evidence synthesis[150] : "*from 2009 to 2015, the evidence covered by existing systematic reviews was consistently incomplete: 45 % to 70 % of trials; 30 % to 58 % of patients; 40 % to 66 % of treatments; and 38 % to 71 % of comparisons were missing. In the cumulative networks of randomized evidence, 10 % to 17 % of treatment comparisons were partially covered by systematic reviews and 55 % to 85 % were partially or not covered.*" These dramatic results "*illustrate how systematic reviews of a given condition provide a fragmented, out-of-date panorama of the evidence for all treatments.*" This not only has consequences on the efficiency of public budgets and spendings, but also on public health, therefore on our capacity to cure people. Finally, there is a need to estimate and follow-up on a regular basis the evolution of the amount of money wasted and to estimate the impact on public health.

- *Costs of Open Science: Issues and Choices to Make*

While some may have thought in the early 2000s that Open Science would be cheaper than existing science, it is now clear that Open Science does not aim or claim to be cheaper. On the contrary, taking care of the data, code, and protocols, publishing negative results and making the whole set of methods more explicit to achieve the goal of greater reproducibility and trust takes time. Political and academic choices must be made in light of these costs, to find the good balance and consider the systemic implications of the upraise of the Open Science paradigm.

For example, the open access to publication was intended to democratize access to the worldwide research results[151]. However, the business model whereby authors pay to publish to cover article processing charges may have made articles more widely accessible but has made publishing, at least in Gold Open Access journals, the prerogative of well-funded research organizations[152]. The academic community is expected to publish in esteemed journals that charge high fees yet it is unclear what added value these journals offer[153] when the total cost for the research organizations is high (30M€ for France in 2020, tripled in ten

[149] Chan, A.-W. et al. Increasing value and reducing waste: addressing inaccessible research. The Lancet 383, 257–266 (2014). DOI: https://doi.org/10.1016/S0140-6736(13)62296-5
[150] Créquit, P., Trinquart, L., Yavchitz, A. & Ravaud, P. Wasted research when systematic reviews fail to provide a complete and up-to-date evidence synthesis: the example of lung cancer. BMC Medicine 14, 8 (2016). DOI: https://doi.org/10.1186/s12916-016-0555-0
[151]Huang, C.-K. (Karl) et al. Evaluating the impact of open access policies on research institutions. eLife 9, e57067 (2020). https://elifesciences.org/articles/57067
[152] Khoo, S. Y.-S. Article Processing Charge Hyperinflation and Price Insensitivity: An Open Access Sequel to the Serials Crisis. LIBER Quarterly: The Journal of the Association of European Research Libraries 29, 1–18 (2019).2. https://liberquarterly.eu/article/view/10729 DOI: 10.18352/lq.10280
Segado-Boj, F., Prieto-Gutiérrez, J.-J. & Martín-Quevedo, J. Attitudes, willingness, and resources to cover article publishing charges: The influence of age, position, income level country, discipline and open access habits. Learned Publishing, DOI: 10.1002/leap.1455
[153] S. Du, J. Opinion: Is Open Access Worth the Cost? *The Scientist Magazine.* https://www.the-scientist.com/critic-at-large/opinion-is-open-access-worth-the-cost-70049

years, expected to double in the next decade, without reducing the other subscription costs[154]). Managing open data requires significant investments in technical and human infrastructure and new business models for its sustainability[155], requiring shared local, national, and global community action[156]. As Borgman and Bourne state[157]: "*Data sharing is not a simple matter of individual practice, but one of infrastructure, institutions, and economics. Governments, funding agencies, and international science organizations all will need to invest in common approaches for data sharing to develop into a sustainable international ecosystem*." The same applies to software, which has the added need for vigilant and continuous attention to maintain it and access it[158].

Understanding the costs of Open Science and identifying models of fair and sustainable distribution of the burden, is necessary.

2.9 Open Science Monitoring

Open Science monitoring is not a straightforward task. Opening science means opening publications, methodology, data, software, and other research outputs. The indicators that could serve as a basement for such a monitoring should therefore provide the ratio of the research outputs that have been opened versus all the research outputs, opened or not. Obtaining a reliable measure for both is a challenge, especially considering the fact that, as recommended by UNESCO[159] and the European Council[160], "*data and bibliographic databases used for research assessment should, in principle, be openly accessible and that tools and technical systems should enable transparency*," implying data sources should be restricted to openly available ones.

Research on Research will help in addressing this challenge. Several initiatives are already proving that it can be within reach, provided the necessary efforts are made. Indeed, drawing a global openness view from openly available data has already been successfully attempted in several domains. For example, OpenAlex[161] from OurResearch is a free and open catalogue of the world's scholarly papers, researchers, journals, and institutions. In the same spirit, the Curtin Open Knowledge Initiative (COKI) Open Access dashboard[162] provides insights into the open access status of publications for nations around the world using only data from public

154 Blanchard, A., Thierry, D. & Graaf, M. van der. Retrospective and prospective study of the evolution of APC costs and electronic subscriptions for French institutions. (Comité pour la science ouverte, 2022). doi:10.52949/26.

155 Bourne, P., Lorsch, J. & Green, E. Perspective: Sustaining the big-data ecosystem. *Nature* 527, S16–S17 (2015). https://doi.org/10.1038/527S16a

156 Anderson, W. A global coalition to sustain core data. *Nature* **543**, 179 (2017). https://doi.org/10.1038/543179a

157 Borgman, C. L., & Bourne, P. E. (2022). Why It Takes a Village to Manage and Share Data. *Harvard Data Science Review*, *4*(3). https://doi.org/10.1162/99608f92.42eec111

158 European Commission, Directorate-General for Research and Innovation, *Scholarly infrastructures for research software : report from the EOSC Executive Board Working Group (WG) Architecture Task Force (TF) SIRS*, Publications Office, 2020, https://data.europa.eu/doi/10.2777/28598

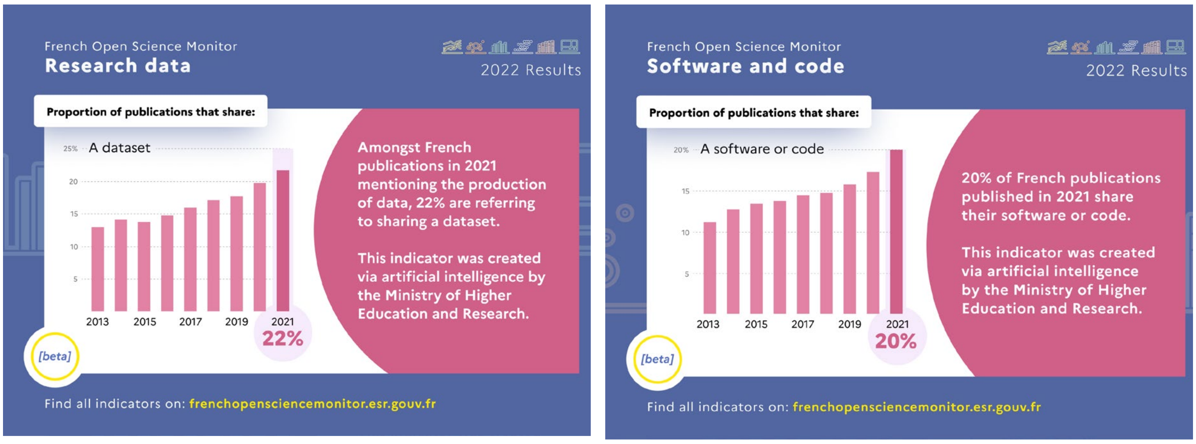
159 https://en.unesco.org/science-sustainable-future/open-science/recommendation

160 https://www.consilium.europa.eu/media/56958/st10126-en22.pdf

161 https://openalex.org

162 https://open.coki.ac

sources. At the national level, the French Open Science Monitor[163] [164] has been developed using a text mining[165] and machine learning approach[166]. It allows to evaluate the openness of French publications[167] and will soon be enriched with openness indicators about data and research software, along with other French research productions. These studies have been conducted by using only openly available data and show that amongst the French publications in 2021 that mention the production of data, 22% are referring to sharing a dataset, while it amounts to 20% for research software. A similar approach had already been previously used at the journal level by PLOS for their Open Science Indicators[168].



Some first beta results from the French Open Science Monitor about Open Research Data and Open Source Research Software

There are many more indicators to build, such as the ratio of clinical trials that publish their results, even if they are negative, moreover because the publication is mandatory in many countries. A large amount of Research on Research[169] has been made about this issue, showing a strong need of monitoring of this critical public health issue.

---

[163] https://frenchopensciencemonitor.esr.gouv.fr/

[164] Eric Jeangirard. Monitoring Open Access at a national level: French case study. *ELPUB 2019 23rd edition of the International Conference on Electronic Publishing*, Jun 2019, Marseille, France. 10.4000/proceedings.elpub.2019.20. hal-02141819

[165] L'Hôte, A. and Jeangirard, E., "Using Elasticsearch for entity recognition in affiliation disambiguation," *arXiv e-prints*, 2021. doi:10.48550/arXiv.2110.01958.

[166] Lopez, Patrice et al. "Mining Software Entities in Scientific Literature: Document-level NER for an Extremely Imbalance and Large-scale Task." *Proceedings of the 30th ACM International Conference on Information & Knowledge Management* (2021): n. pag.

[167] Lauranne Chaignon, Daniel Egret; Identifying scientific publications countrywide and measuring their open access: The case of the French Open Science Barometer (BSO). *Quantitative Science Studies* 2022; 3 (1): 18–36. doi: https://doi.org/10.1162/qss_a_00179

[168] Explore the first Open Science Indicators dataset—and share your thoughts. *The Official PLOS Blog* https://theplosblog.plos.org/2022/12/open-science-indicators-first-dataset/ (2022). https://theplosblog.plos.org/2022/12/open-science-indicators-first-dataset/

[169] DeVito, N. J., Bacon, S. & Goldacre, B. Compliance with legal requirement to report clinical trial results on ClinicalTrials.gov: a cohort study. The Lancet 395, 361–369 (2020).
Harriman, S. L. & Patel, J. When are clinical trials registered? An analysis of prospective versus retrospective registration. Trials 17, 187 (2016).
Reveiz, L., Villanueva, E., Iko, C. & Simera, I. Compliance with clinical trial registration and reporting guidelines by Latin American and Caribbean journals. Cad Saude Publica 29, 1095–1100 (2013).

3. Existing Research on Research Initiatives

Research on Research can be used to study, mobilize, improve, and accelerate the Open Science journey. In recent years, the importance of Research on Research on Open Science has been demonstrated by the number of global initiatives underway. A 2021 report commissioned by the French Open Science Committee has identified different approaches and initiatives in the field. "*Some are theoretically and methodologically rooted in traditional academic fields such as sociology, economics, political science, philosophy, or information science (bibliometrics and scientometrics). Others with more data-intensive approaches come from computational social sciences or biomedical fields and have evolved in a favourable context for open science in terms of public policies. While the former currents are based on well-established pre-existing academic fields and methods, such as STS [science and technology studies] or scientometrics the latter have appeared more recently, and have adopted a prescriptive, change-oriented focus as well as a normative commitment to foster better and more open science.*"[170] The report also refers to the evolving nature of Research on Research, notably being fuelled by contemporary debates, "*such as reproducibility, evidence-based practices, integrity and inclusivity in research, and some community-issued warnings about not "reinventing the wheel"*". According to the authors, new alliances "*are forming between research centres and laboratories, funding institutions, policy-makers and data providers in order to support public policy-makers with evaluation tools and research protocols to guide decision-making and action.*" There are thus basically two kinds of researchers that are specialized in Research on Research: both Humanities and Social Sciences (HSS) and Science, Technology, Medicine (STM) researchers that specialized in Research on Research itself; and researchers not specialized in Research on Research, but who invest their time and research effort in their own discipline, in order to improve the efficiency, transparency and reproducibility of their own research.

Depending on its history, its methodologies and problematic, Research on Research encompasses different connected research fields: sociology of science[171], science and technology studies (STS)[172], meta-research[173], metascience[174] and science of science[175].

Nguyen, T.-A.-H., Dechartres, A., Belgherbi, S. & Ravaud, P. Public availability of results of trials assessing cancer drugs in the United States. J. Clin. Oncol. 31, 2998–3003 (2013).

[170] Gruson-Daniel, C. & Anderson-González, M. Étude exploratoire sur la « recherche sur la recherche » : acteurs et approches. 66 p. (Comité pour la science ouverte, 2021). doi:10.52949/24.

[171] Merton, R. K. The sociology of science. Theoretical and Empirical Investigations. (The University of Chicago Pres, 1973).

[172] Felt, U., Fouche, R., Miller, C. A. & Smith-Doerr, L. The Handbook of Science and Technology Studies. https://mitpress.mit.edu/9780262035682/the-handbook-of-science-and-technology-studies/

[173] Ioannidis, J.P.A. (2018). Meta-research: Why research on research matters. PLoS Biol 16(3): e2005468. https://doi.org/10.1371/journal.pbio.2005468Ioannidis, J. P. A., Fanelli, D., Dunne, D. D. & Goodman, S. N. Meta-research: Evaluation and Improvement of Research Methods and Practices. PLOS Biology 13, e1002264 (2015).
https://doi.org/10.1371/journal.pbio.1002264

[174] Schooler, J. W. Metascience could rescue the 'replication crisis.' *Nature* 515, 9–9 (2014). https://doi.org/10.1038/515009a

[175] Fortunato, S. et al. Science of science. Science 359, eaao0185 (2018). DOI: 10.1126/science.aao0185

This research community publishes in dedicated Research on Research journals, such as *Journal of Infometrics*[176], *Scientometrics*[177], and *Quantitative Science Studies*[178]. However, the community also publishes in journals dedicated to a specific research field (for example PLOS journals) or in journals where Research on Research plays an important role without being the focus of the publication (for example, *Journal of the Association for Information Science and Technology, Revue d'anthropologie des connaissances*[179]) and

Research on Research has been structured around the creation of research centres or specialized initiatives dedicated to creating a community of researchers:

- For example, the **Association for Interdisciplinary Meta-Research and Open Science** (**AIMOS**), *seeks to advance the interdisciplinary field of meta-research by bringing together and supporting researchers in that field*[180]. To do so, AIMOS hosts an annual conference that *brings together researchers from across a number of disciplines* [with a particular focus on Open Science] *to talk about how research is done, and how it can be improved*[181]. The last AIMOS conference took place in November 2022. Another interesting example is the Research on Research Institute (RoRi) whose objective is to *translate ideas and evidence into practical, real world solutions to improve research culture and systems*. They do so *by bringing together people and organizations that care about research, gathering information and developing tools to inform and improve how research is funded, practised, communicated, and evaluated*[182]. RoRi projects examine topics such as research assessment, FAIR data principles and peer review.
- The **Center for Open Science (COS)** mission is to *increase openness, integrity, and reproducibility of research*[183]. As part of the work they undertake for their mission, COS organizes the Metascience Conference - *a global gathering for knowledge sharing, community building, and opportunities to define a roadmap of research and intervention priorities to accelerate science*[184]. The program committee includes representatives from AIMOS, RoRi and various universities.
- The **Centre for Science and Technology Studies (CWTS)** at Leiden University *studies scientific research and its connections to technology, innovation, and society*[185]. Knowledge is shared along four thematic research hubs, which include an Open Science hub. The Open Science thematic hub *aims to synthesize CWTS research on the current policy drive towards "open science" and translate our research results as well as theoretical, empirical, and technical expertise to applicable ideas,*

---

[176] Journal of Informetrics | ScienceDirect.com by Elsevier
[177] Scientometrics | Home (springer.com)
[178] Quantitative Science Studies | MIT Press
[179] https://journals.openedition.org/rac
[180] AIMOS
[181] AIMOS 2022 | November 28, 2022 - November 30, 2022 (eventcreate.com)
[182] About (researchonresearch.org)
[183] Our Mission (cos.io)
[184] Metascience
[185] CWTS - Centre for Science and Technology Studies - Leiden University

advice, and technical solutions[186]. It explores the topics of infrastructures, policies, data, indicators, and research assessment.

- The **CSI - centre de sociologie de l'innovation** is, in France, a centre with dedicated teams to research on research. "*Building on the achievements of the STS (Science and Technology Studies) and of pragmatist-inspired approaches, the CSI develops research on these public concerns and on the investigations carried out on them by a plurality of actors. This research aims to examine the knowledge and the devices produced by the actors to qualify the problems at stake and tell how they affect them. The approaches implemented aim to renew the old slogan of the actor-network theory, "follow the actors themselves," by seeking cooperation with people and collectives who strive to make these unprecedented realities count and who design new ways of reporting on them: publics, users, public authorities, companies, scientists, activists, associations."*[187]
- **The Meta-Research Innovation Center at Stanford (METRICS)** is a research-to-action Centre that is focused on transforming research practices to improve the quality of scientific studies in biomedicine and beyond. METRICS fosters multidisciplinary research collaborations to help produce solutions that increase the effectiveness and value of scientific investigation[188].
- The **Canada Research Chair on the Transformations of Scholarly Communication** (Université de Montréal, Canada). *This research program,* led by Vincent Larivière, "*aims to better understand the evolution in research practices and publication in the digital era, as well as their effects on the structure of the scientific community. This knowledge will help inform the development of public policy related to scientific research and technology"*[189].
- Moving forward, the Ministry of Higher Education and Research (France) plans to create a **centre for Research on Research on Open Science (LabSo)**. This will be part of the work undertaken in response to the country's second National Plan for Open Science.

There are, however, many more other growing initiatives such as the QUEST Center for Responsible Research (Berlin), ELICO (Lyon), Institut Francilien Recherche Innovation Société (IFRIS, Paris), Center for Science of Science and Innovation (CSSI, Evanston, Illinois)[190], Science of Science & Computational Discovery Lab (University of Colorado) [191], NEtwoRks, Data, Society (NERDS, IT University of Copenhagen), and School of Public Policy (Georgia Institute of Technology)[192], etc.

---

[186] CWTS: Open Science
[187] https://www.csi.minesparis.psl.eu/recherches/
[188] https://metrics.stanford.edu/about-us
[189] Program | Canada Research Chair on the Transformations of Scholarly Communication (umontreal.ca)
[190] https://www.kellogg.northwestern.edu/research/science-of-science.aspx
[191] https://scienceofscience.org/
[192] https://spp.gatech.edu/

The Research on Research field is highly active on multiple dimensions. One can cite some of them to show the diversity and pertinence for Open Science of such approaches:

- **Nanobubbles**[193] : The project focuses on how, when, and why science fails to correct itself. To understand how the correction of science works or fails, the NanoBubbles project combines approaches from the natural sciences, engineering (natural language processing) and humanities and social sciences (linguistics, sociology, philosophy, and history of science). They have notably proven the "prevalence of nonsensical algorithmically generated papers in the scientific literature"[194].
- **Pathways**[195] : The project is investigating how data is gathered about research careers, so that we can work out how best to gather this information in the future. They are looking at research career pathways in six countries: Austria, Canada, Denmark, Germany, the UK, and the USA.
- **Randomization**[196] : The project is running a series of experiments to test how well lottery-style research funding works so that we can understand the benefits and drawbacks of these approaches and how best to use them.
- **Harbinger project**[197] : The project is the first international research project to examine the place of Early Career Researchers in the current transformation of the academic world. Funded by the Publishing Research Consortium (PRC) and led by Professor David Nicholas, the project focuses on the way this younger generation adapts to the Open Science challenge, seeking to evaluate whether they rather lead the change or resist to it[198].
- **ON-MERRIT**[199] : ON-MERRIT is a 30-month project funded by the European Commission to investigate how and if open and responsible research practices could worsen existing inequalities.
- **Research on Wikipedia**[200] : a set of diverse research teams has conducted research on Wikipedia, notably on the question of the appropriation of the collaborative

---

193 https://nanobubbles.hypotheses.org/
194 Cabanac, G. & Labbé, C. Prevalence of nonsensical algorithmically generated papers in the scientific literature. *Journal of the Association for Information Science and Technology* **72**, 1461–1476 (2021).
195 https://researchonresearch.org/projects#!/tab/273951116-4
196 https://researchonresearch.org/projects#!/tab/273951116-6
197 https://elico-recherche.msh-lse.fr/programme/harbinger-research-project
198 Nicholas, David; Jamali, Hamid R.; Herman, Eti; Xu, Jie; Boukacem-Zeghmouri, Chérifa; Watkinson, Anthony; Rodríguez-Bravo, Blanca; Abrizah, Abdullah; Świgoń, Marzena; Polezhaeva, Tatiana (2020). "How is open access publishing going down with early career researchers? an international, multi-disciplinary study." Profesional de la información, v. 29, n. 6, e290614. DOI: 10.3145/epi.2020.nov.14
199 https://on-merrit.eu/
200 Teplitskiy, M., Lu, G. & Duede, E. Amplifying the impact of open access: Wikipedia and the diffusion of science. *Journal of the Association for Information Science and Technology* **68**, 2116–2127 (2017).
Kousha, K. & Thelwall, M. Are wikipedia citations important evidence of the impact of scholarly articles and books? *Journal of the Association for Information Science and Technology* **68**, 762–779 (2017).

encyclopaedia by the public and the question of the links between academic literature and Wikipedia contents. This is one of the ways we could have a proxy showing a direct link between research literature and society.

- **Data journeys in the Sciences**[201] : After five years of research, this project has highlighted the process where "data move from the sites in which they are originally produced to sites where they can be integrated with other data, analyzed and re-used for a variety of purposes." This allows us to understand "disciplinary, geographical and historical differences and similarities in data management, processing and interpretation, thus identifying the key conditions of possibility for the widespread data sharing associated with Big and Open Data." This is, of course, complementary to the research of Christine L. Borgman[202] where she argues that "to manage and exploit data over the long term, requires massive investment in knowledge infrastructures; at stake is the future of scholarship."

Some Research on Research initiatives come with a digital service aimed at filling a technical gap, to improve openness, transparency, and reproducibility:

- The Center for Open Science (Charlottesville, USA) has created the **Open Science Framework**, a free and open source project management tool that supports researchers throughout their entire project lifecycle[203].
- The Centre for Science and Technology (CWTS) (Leiden, NL) has created the **Leiden Ranking**[204].
- The **Initiative for Open Citations I4OC**[205] (Bologna, Italy) is a collaboration between scholarly publishers, researchers, and other interested parties to promote the unrestricted availability of scholarly citation data[206]. They publish many research articles as they provide an Open Science infrastructure dedicated to citations.
- Public Knowledge Project (Vancouver, Canada) has created **Open Journals System** (OJS) and other open source publishing software, to give academic publishers the power to publish open access journals within their own infrastructure.

Piccardi, T., Redi, M., Colavizza, G. & West, R. Quantifying Engagement with Citations on Wikipedia. in *Proceedings of The Web Conference 2020* 2365–2376 (Association for Computing Machinery, 2020). doi:10.1145/3366423.3380300.
Maggio, L. A., Steinberg, R. M., Piccardi, T. & Willinsky, J. M. Reader engagement with medical content on Wikipedia. *eLife* **9**, e52426 (2020).5.
Singh, H., West, R. & Colavizza, G. Wikipedia citations: A comprehensive data set of citations with identifiers extracted from English Wikipedia. *Quantitative Science Studies* **2**, 1–19 (2021).
[201] Leonelli, S. & Tempini, N. *Data Journeys in the Sciences*. (Springer Nature, 2020). doi:10.1007/978-3-030-37177-7.
[202] Borgman, C. L. *Big Data, Little Data, No Data: Scholarship in the Networked World*. (The MIT Press, 2015).
[203] https://www.cos.io/products/osf
[204] https://www.leidenranking.com/
[205] https://i4oc.org/
[206] Peroni, S. & Shotton, D. OpenCitations, an infrastructure organisation for open scholarship. Quantitative Science Studies 1, 428–444 (2020). https://doi.org/10.1162/qss_a_00023

- The Roy Rosenzweig Center for History and New Media (Washington, USA) has created the successful, open source reference manager called **Zotero**[207].
- **Protocols.io**[208] has been created by Lenny Teytelman after a frustrating postdoc experience at MIT. It is a secure platform to organize up-to-date/versionable methods with history and concurrent editing. In addition, this platform allows users to create and discover reproducible, experimental, and computational methods.
- **Software Heritage** has been created by INRIA and UNESCO. It is a global infrastructure that aims to propose a global solution to Software preservation and citability[209].
- **CRediT[210]** has been created has a proposed solution to the so-called Matthew Effect that has been identified by Research on Research decades ago[211] and to the complexity of the cycles of credits[212]. It is a new standard dedicated to the description of the specific role of each co-authors in a publication. Major publishers have implemented it. It is a vocabulary including fourteen different contribution roles, including roles related to research data, data curation, software or visualization, and not only writing, drafting, reviewing and editing.
- **GROBID, SoftCite and Datastet** has been created by Patrice Lopez[213] to use the so-called artificial intelligence to identify authors (and their metadata) in PDF publications (GROBID), and mention of data (Datastet) and software (SoftCite) in academic papers. GROBID, SoftCite and Datastet are open source software relying on high quality learning databases.
- The **Cochrane Library[214]** is a collective, worldwide effort, to publish meta-analyses in health research, building stronger evidence for informed decisions. The meta-analyses are translated into numerous languages, to be easily accessible and easy to understand by physicians.

---

207 https://www.zotero.org/about/
208 How to make your protocol more reproducible, discoverable, and user-friendly (protocols.io)
209 Roberto Di Cosmo and Stefano Zacchiroli. 2017. Software Heritage: Why and How to Preserve Software Source Code. In Proceedings of the 14th International Conference on Digital Preservation, iPRES 2017.
210 Brand, A., Allen, L., Altman, M., Hlava, M. & Scott, J. Beyond authorship: attribution, contribution, collaboration, and credit. *Learned Publishing* **28**, 151–155 (2015).
211 Larivière, V., Pontille, D. & Sugimoto, C. R. Investigating the division of scientific labor using the Contributor Roles Taxonomy (CRediT). *Quantitative Science Studies* **2**, 111–128 (2021).
212 Latour, B., and S. Woolgar. 1986. Laboratory life: The construction of scientific facts. Princeton University Press.
213 Lopez, P. GROBID: Combining Automatic Bibliographic Data Recognition and Term Extraction for Scholarship Publications. in *Research and Advanced Technology for Digital Libraries* (eds. Agosti, M., Borbinha, J., Kapidakis, S., Papatheodorou, C. & Tsakonas, G.) 473–474 (Springer, 2009). doi:10.1007/978-3-642-04346-8_62.
Lopez, P., Du, C., Cohoon, J., Ram, K. & Howison, J. Mining Software Entities in Scientific Literature: Document-level NER for an Extremely Imbalance and Large-scale Task. in *Proceedings of the 30th ACM International Conference on Information & Knowledge Management* 3986–3995 (Association for Computing Machinery, 2021). https://doi.org/10.1145/3459637.3481936
214 https://www.cochranelibrary.com/

- The **Dataverse Project**[215] is an open source web application to share, preserve, cite, explore, and analyze research data. It facilitates making data available to others and allows you to replicate others' work more easily. Researchers, journals, data authors, publishers, data distributors, and affiliated institutions all receive academic credit and web visibility. It is run by a consortium led by the Institute for Quantitative Social Science (IQSS), the Harvard University Library and Harvard University Information Technology organisation.

This list indicates that Research on Research is continuously fuelling the academic ecosystem with Open Science tools dedicated both to the research itself and to policy-makers.

4. An International Research on Research Initiative

There is a need to deepen research on research on Open Science at an international level, as well as increase coordination and knowledge-sharing. Existing research has paved the way; however, there are often needs of confirmation and diversification on a disciplinary level, as most of the existing research is discipline specific. For less explored questions, there is a need to open new research questions. For example, the questions of economic and societal impact of science opening need more research, such as the questions of the costs of Open Science and its benefits. Moreover, the question of the compatibility of values and cultures when confronted with Open Science policies, recommendations, tools, and processes could be addressed. If we want Open Science to become the default academics' daily life, the existing obstacles could be carefully studied, and the facilitating approaches could be tested against a variety of situations and disciplines. The list of questions to be addressed is thus large, but one could attempt to group them within 5 main categories. The scope of the questions varies based on the level of maturity of the topic.

1. **Research Assessment and Incentives**
    - What are the lessons learned of the existing experimentations concerning narrative CVs?
    - How can we achieve the goal of fully open and reusable bibliographic databases and other research outputs databases to plan research assessment only on FAIR data?[216]
2. **Skill Gaps**
    - Among the numerous initiatives dedicated to training and knowledge-sharing of Open Science research data management, software management, and publishing, how do we evaluate (qualitatively and quantitatively) success or failure of the initiatives?
    - Open Science and Research Assessment are closely linked: how do we address the issue of the skills needed to assess other forms of CVs, such as narrative ones?
3. **Open Science Monitoring**

[215] https://dataverse.org/

[216] On this topic, see the project I4OC, OpenAlex, I4OA.

- How do the values of the research community evolve towards Open Science? This study could be initiated under the umbrella of an annual or biannual research, willing to extend the annual "state of open data" published by Digital Science.
- How can a worldwide Open Science Observatory, covering diverse dimensions of Open Science and not only publications, be built by using open data exclusively? How can we assess such an Observatory to ensure it is meaningful and mitigates unintended biases?

4. **Open Science Infrastructures**

- Do Open Science badges increase Open Science practices where they have been tested?[217]
- How can we measure the success (or failure) of an Open Science infrastructure, beyond measuring the number of downloaded files or terabytes? How can we know to what extent an existing infrastructure meets the goal it has been designed for?
- How could we evaluate and monitor in the long run the usages, and the benefits, of Open Science infrastructures by and for the academic community and society as a whole?

5. **Open Science Impacts**

- Open Science is said to benefit the economy. How can it be measured and detailed?
- Open Science has societal impacts: it is said to enhance trust in science though the preprint process is hard to understand. How could we measure this and provide evidenced based solutions?
- Research itself is said to benefit from its opening, in many fields such as progress speed, reliability, cumulative nature, replicability, integrity… Can all this be evaluated, monitored, and described in different fields and situations?

Research on research could also inspire a framework of Open Science monitoring. This topic is useful for national purposes and is also currently discussed by UNESCO to implement the UNESCO recommendation on Open Science (2021)[218]. It could address the different dimensions of Open Science and the different aspects of the research life cycle. Some parts of Open Science monitoring could be quantitative. Thanks to Unpaywall, we know how many publications are open access. And some parts could be qualitative when we deal with culture,

---

[217] Rowhani-Farid, A., Aldcroft, A. & Barnett, A. G. Did awarding badges increase data sharing in BMJ Open? A randomized controlled trial. *Royal Society Open Science* **7**, 191818.
Kidwell, M. C. et al. Badges to Acknowledge Open Practices: A Simple, Low-Cost, Effective Method for Increasing Transparency. PLOS Biology 14, e1002456 (2016). Doi: 10.1371/journal.pbio.1002456
Rowhani-Farid, A., Aldcroft, A. & Barnett, A. G. Did awarding badges increase data sharing in BMJ Open? A randomized controlled trial. Royal Society Open Science 7, 191818. DOI: 10.1098/rsos.191818
[218] UNESCO Recommendation on Open Science - UNESCO Digital Library

habits, and resistance. The recommended approach could combine quantitative and qualitative approaches, and deal as much as possible with disciplinary and geographical variations. To get an Open Science monitor dealing with open/FAIR data with as much precision as with publications, we have far more work to do. We should avoid confusing research data available online in the 3000+ research data repositories in the world, which roughly represents 20% of the existing research data that has been used for publishing new knowledge[219], with the hidden iceberg of unknown, untraceable, unFAIR and uncurated research data that populate researchers' personal computers and USB keys.

There is also a need to provide a state of knowledge on Research on Research, to help decision and policy-makers know the latest findings of Research on Research. The International Panel of Climate Change (IPCC) undertakes a similar activity. We need the same kind of effort, but on a smaller scale of research on research. Strong meta-analysis published in one or several reports, translated in different languages, would be of great help, obviously published in open access, with open data as rich as possible to help the appropriation and reuse of the results.

All these questions, monitoring and meta-analyses could be aggregated into something like an Open Science Observatory. The recommendation is to articulate and coordinate existing Research on Research on Open Science. There is also; however, a need to stimulate new research that would fit into the big picture of Open Science policy making and providing strong evidence. To reach good results in a reasonable time and with Open Science policy in mind, international coordination would be really helpful. The already-existing research on research congresses, moreover when they focus on Open Science, but also the Research Data Alliance plenaries and events of this kind, could be used as an opportunity to help building stronger and larger communities and stimulate transdisciplinary and trans-professions scholarly discussions in the field. There is also a need for academic coordination to draft collaboratively and then publish meta-analyses on the state of the art in research on research. The same coordination is needed to propose accurate indicators based on both qualitative and quantitative approaches.

For this coordination to succeed, we could explore the possibility of creating an international research **initiative** for **Research on Research and Open Science**. As the Open Science agenda is a global topic, such an initiative should probably not be restricted to G7 countries and be open to any other country willing to contribute to such an effort.

5. About the Research on Research Sub-Working Group

[219] French Open Science Monitor https://frenchopensciencemonitor.esr.gouv.fr/

The G7 OSWG Research on Research Sub-Working Group consists of experts from Canada (Co-Chair), France (Co-Chair), Japan, the United Kingdom, the United States and the European Union.

The Research on Research Sub-Working Group would like to thank all the subject-matter experts who contributed to this final report either directly or through discussions, which took place during the two Workshops organized by the Sub-Working Group members, namely:

- Stefanie Haustein, University of Ottawa, Canada
- Monica Granados, Creative Commons, Canada
- Vincent Larivière, Université de Montréal, Canada
- David Moher, The Ottawa Hospital Research Institute, Canada
- Richard Gold, McGill University, Canada
- Leigh-Ann Butler, Natural Sciences and Engineering Research Council of Canada, Canada
- Marin Dacos, Ministère de l'enseignement supérieur et de la recherche, France
- Didier Torny, Centre national de la recherche scientifique, France
- Pierre-Carl Langlais, OPSci, France
- Mariannig Le Bechec, Université Claude Bernard Lyon 1, France
- Florian Naudet, Université Rennes 1, France
- Isabelle Boutron Université Paris Cité, France
- Shingo Sogabe, Cabinet Office-Government of Japan, Japan
- Kazuhiro Hayashi, National Institute of Science and Technology Policy, Japan
- Kenkichi Takase, Chuo University, Japan
- Rachel Bruce, UK Research, and Innovation, United Kingdom
- Carole Goble, University of Manchester & Joint Head of Node ELIXIR, United Kingdom
- Sabina Leonelli, University of Exeter, United Kingdom
- Caroline Jay, University of Manchester, United Kingdom
- Sophia Crüwell, University of Cambridge, United Kingdom
- Christopher Allen, Cardiff University, United Kingdom
- Iain Hrynaszkiewicz, Public Library of Science, United Kingdom
- Ian Viney, UK Research and Innovation, United Kingdom
- Despoina Sousoni, ELIXIR-Hub, United Kingdom
- Corinne Martin, ELIXIR-Hub, United Kingdom
- Christine L. Borgman, UCLA, USA
- Merce Crosas, Harvard, USA
- Rebecca Meseroll, National Institutes of Health, USA
- Hashi Wijayatilake, National Institutes of Health, USA
- George Santangelo, National Institutes of Health, USA
- Ameet Doshi, Princeton University, USA
- Diana Hicks, Georgia Institute of Technology, USA
- Michael Arentoft, European Commission, Belgium
- Pantelis Tziveloglou, European Commission, Belgium
- Rene Von Schomberg, European Commission, Belgium
- Kostas Glinos, European Commission, Belgium
- Sybille Hinze, Centre of Open and Responsible Research, Germany

- Isabella Peters, ZBW Leibniz Information Centre for Economics and CAU Kiel University, Germany
- Simine Vazire, University of Melbourne, Australia
- Leonidas Pispiringas, OpenAIRE, Greece
- Ioanna Grypari, Athena Research Centre, Greece
- Marialuisa Lavitrano, University of Milano, Italy
- Francesco Giffoni - Centre for Industrial Studies, Italy

6. Bibliography

140C. "Initiative for Open Citations." https://i4oc.org/. Accessed March 3, 2023.

Allen, Christopher and David M.A. Mehler. "Open science challenges, benefits, and tips in early career and beyond." *PLOS Biology* 17 (12):1-14, 2019. https://doi.org/10.1371/journal.pbio.3000587

Alperin, Juan-Pablo. "The public impact of Latin America's approach to open access." Thesis (Ph.D, Stanford University), 2015, https://purl.stanford.edu/jr256tk1194.

Anderson, W. A global coalition to sustain core data. *Nature* 543, 179, 2017. https://doi.org/10.1038/543179a

Association for Interdisciplinary Meta-Research and Open Science. https://aimos.community/. Accessed March 2, 2023.

Association for Interdisciplinary Meta-Research and Open Science. "Fourth Annual Conference of the Association for Interdisciplinary Meta-research and Open Science (AIMOS 2022) AIMOS 2022 | November 28, 2022 - November 30, 2022 (eventcreate.com). Accessed March 2, 2023.

Baker, Lee, et al., *Reproducibility of scientific results in the EU : scoping report*, European Commission, Directorate-General for Research and Innovation, Lusoli, W. (editor), Publications Office, 2020, https://data.europa.eu/doi/10.2777/341654.

Béchec, Mariannig Le, et al. "State of open science practices in France (SOSP-FR)." *Comité pour la science ouverte*, 2022. https://doi.org/10.52949/5.

BMC Medicine. "Peer Review Policy."https://bmcmedicine.biomedcentral.com/submission-guidelines/peer-review-policy. Accessed February 21, 2023.

Blanchard, Antoine, Thierry, Diane, and Maurits van der Graaf. "Retrospective and prospective study of the evolution of APC costs and electronic subscriptions for French institutions." *Comité pour la science ouverte*, 2022. doi:10.52949/26.

Borgman, Christine L. *Big Data, Little Data, No Data: Scholarship in the Networked World*, The MIT Press, 2015.

Borgman, Christine, and Philip Bourne. "Why It Takes a Village to Manage and Share Data." *Harvard Data Science Review*, *4*(3), 2022. https://doi.org/10.1162/99608f92.42eec111

Boulton, Geoffrey, Rawlins, Michael, Vallance, Patrick, and Mark Walport. "Science as a public enterprise: the case for open data". *The Lancet,* 2011. DOI:https://doi.org/10.1016/S0140-6736(11)60647-8

Bourne, Philip, Lorsch, Jon, and Eric Green. "Perspective: Sustaining the big-data ecosystem." *Nature* 527, S16–S17, 2015. https://doi.org/10.1038/527S16a

Brand, Amy, Allen, Liz, Altman, Micah, Hlava, Marjorie, and Jo Scott. "Beyond authorship: attribution, contribution, collaboration, and credit." *Learned Publishing* 28, 151–155, 2015. https://doi.org/10.1087/20150211.

Buneman, Peter et. al., “Why data citation isn't working, and what to do about it.” *Database*, Volume 2020, 2020, baaa022, https://doi.org/10.1093/databa/baaa022

Cabanac, Guilaume and Cyril Labbé. “Prevalence of nonsensical algorithmically generated papers in the scientific literature.” *Journal of the Association for Information Science and Technology* 72, 1461–1476, 2021. https://doi.org/10.1002/asi.24495.

Cabinet Office, Government of Japan. “ Communique of the G7 Ministers of Science and Technology,” Communiqué of the G7 Ministers of Science and Technology 2016 (utoronto.ca), May 19, 2016, http://www.g7.utoronto.ca/science/2016-tsukuba.html.

Cascad “The first certification agency for scientific code and data.” https://www.cascad.tech/. Accessed March 4, 2023.

Centre de Sociologie de l’innovation. « Les recherches au CSI. » https://www.csi.minesparis.psl.eu/recherches/ . Accessed February 3, 2023.

CERN, “A short history of the web.” https://home.cern/science/computing/birth-web/short-history-web, Accessed February 1, 2023.

Chambers, Christopher, and Loukia Tzavella. “The past, present and future of Registered Reports.” *Nature Human Behaviour*, 6, 29-42. 2022. https://doi.org/10.1038/s41562-021-01193-7

Chan, An-Wen et al. “Increasing value and reducing waste: addressing inaccessible research.” *The Lancet* 383, 257–266, 2014. DOI: https://doi.org/10.1016/S0140-6736(13)62296-5

Chubin, Daryl E. “Open Science and Closed Science: Tradeoffs in a Democracy.” Science, Technology, & Human Values, vol. 10, no. 2, pp. 73–81, 1985.

Clément-Fontaine, Melanie, Cosmo, Roberto, Guerry, Bastien, Moreau, Patrick, and Francois Pellegrini. “Encouraging a wider usage of software derived from research.” 6 p. *In Scholarly Infrastructures for Research Software. Comité pour la science ouverte*, 2019. doi:10.52949/4.

Coalition for Advancing Research Assessment, Agreement of the assessment, accessed February 27, 2023. https://coara.eu

Cochrane Library. https://www.cochranelibrary.com/ . Accessed February 24, 2023.

Colavizza, Giovanni, Hrynaszkiewicz, Iain, Staden, Isla, Whitaker, Kirstie and Barbara McGillivray. “The citation advantage of linking publications to research data.” *PLOS ONE*, 15, e0230416, 2020, https://doi.org/10.1371/journal.pone.0230416.

Cousijn, Helena, et al. “A data citation roadmap for scientific publishers.” *Scientific Data* 5, 180259, 2018. https://doi.org/10.1038/sdata.2018.259.

Couture, Jessica, Blake, Rachel, McDonald, Gavin, and Colette Ward. “A funder-imposed data publication requirement seldom inspired data sharing.” *PLOS ONE* 13, e0199789, 2018. DOI: 10.1371/journal.pone.0199789.

Créquit, Perrine, Trinquart, Ludovic, Yavchitz, Amelie, and Philippe Ravaud. "Wasted research when systematic reviews fail to provide a complete and up-to-date evidence synthesis: the example of lung cancer." *BMC Medicine* 14, 8, 2016, https://doi.org/10.1186/s12916-016-0555-0.

Curtin Open Knowledge Initiative. "Open Access Dashboard" https://open.coki.ac. Accessed on February 24, 2023.

CWTS – Meaningful Metrics. "Centre for Science and Technology Studies." CWTS - Centre for Science and Technology Studies - Leiden University, Accessed February 2, 2023.

CWTS Meaningful Metrics. "CWTS Leiden Ranking 2022." https://www.leidenranking.com/. Accessed March 2, 2023.

CWTS – Meaningful Metrics. "Open Science." https://www.cwts.nl/research/thematic-hubs/open-science, *Leiden University*. Accessed February 2, 2023.

Dacos, Marin. "Des nains sur les épaules de géants : ouvrir la science en France. » *Revue Politique et Parlementaire*, 2019. https://hal.science/hal-02366604

Dandieu, Claire and Hirmeos Consortium. "Report on Post-Publication Open Peer Review Experiment." Zenodo. 2019.

DeVito, Nicholas J. Bacon, Seb, and Ben Goldacre. "Compliance with legal requirement to report clinical trial results on ClinicalTrials.gov: a cohort study." *The Lancet* 395, 361–369,2020. https://doi.org/10.1016/S0140-6736(19)33220-9.

Di Cosmo, Robert. and Stephan Zacchiroli. "Software Heritage: Why and How to Preserve Software Source Code." in *iPRES 2017* - 14th International Conference on Digital Preservation, 2017. https://hal.science/hal-01590958/

Di Cosmo, Robert. Building the software pillar of Open Science. In *Open Science European Conference (OSEC 2022)*, pages 183--193. OpenEdition Press, 2022. https://doi.org/10.4000/books.oep.15829

Dombrowski, Quinn. "What Ever Happened to Project Bamboo?" *Lit Linguist Computing* fqu026, 2014, DOI:10.1093/llc/fqu026.

Du, Jingshan S. "Opinion: Is Open Access Worth the Cost?" *The Scientist Magazine.* 2022. https://www.the-scientist.com/critic-at-large/opinion-is-open-access-worth-the-cost-70049

Equipe de recherche de Lyon en sciences de l'information et de la communication (elico) « Harbinger Research Project, 2016-2020 » https://elico-recherche.msh-lse.fr/programme/harbinger-research-project. Accessed February 23, 2023.

European Commission. "Directorate-General for Research and Innovation, Cost-benefit analysis for FAIR research data: cost of not having FAIR research data." *Publications Office, 2019.* https://data.europa.eu/doi/10.2777/02999

European Commission, Directorate-General for Research and Innovation, *Scholarly infrastructures for research software : report from the EOSC Executive Board Working Group (WG) Architecture Task Force (TF) SIRS*, Publications Office, 2020. https://data.europa.eu/doi/10.2777/28598

European Commission, Directorate-General for Research, and Innovation. Manolan, Natalia et al. *Digital skills for FAIR and Open Science: report from the EOSC Executive Board Skills and Training Working Group*. Publications Office; 2021. doi/10.2777/59065

European Commission, Directorate-General for Research and Innovation, *Towards a reform of the research assessment system : scoping report*, Publications Office, 2021, https://data.europa.eu/doi/10.2777/707440.

European Commission, Directorate-General for Research and Innovation, *Assessing the reproducibility of research results in EU Framework Programmes for Research: final report*, Publications Office of the European Union, 2022. https://data.europa.eu/doi/10.2777/186782

European University Association, Saenen, Bregt and Lidia Borrell-Domian. "Reflections on University Assessment: key concepts, issues, and actors." accessed February 19, 2023. https://eua.eu/resources/publications/825:reflections-on-university-research-assessment-key-concepts,-issues-and-actors.htmlhttps://eua.eu/resources/publications/825:reflections-on-university-research-assessment-key-concepts,-issues-and-actors.html

Fanelli, Daniele. "Negative results are disappearing from most disciplines and countries." *Scientometrics*, 90, 891–904, 2012. https://doi.org/10.1007/s11192-011-0494-7 .

Felt, Ulrike, Fouche, Rayvon, Miller, Clark A. and Laurel Smith-Doerr. "The Handbook of Science and Technology Studies." *The MIT Press*. 2016. https://mitpress.mit.edu/9780262035682/the-handbook-of-science-and-technology-studies/.

Fortier, Ian *et al.* Maelstrom "Research guidelines for rigorous retrospective data harmonization." *Int. J. Epidemiol.* dyw075,2016. doi:10.1093/ije/dyw075.

Fortunato, Laura, and Mark Galassi. "The case for free and open source software in research and scholarship." *Phil. Trans. R. Soc.* A.3792020007920200079, 2021. http://doi.org/10.1098/rsta.2020.0079

Fortunato, Santo. et al. "Science of science." *Science* 359, eaao0185, 2018. DOI: 10.1126/science.aao0185.

Fraser, Nicholas et al. "The evolving role of preprints in the dissemination of COVID-19 research and their impact on the science communication landscape." *PLOS Biology* 19, e3000959, 2021.

French Ministry of Higher Education and Research. "French Open Science Monitor." https://frenchopensciencemonitor.esr.gouv.fr/ . Accessed February 2, 2023.

Gabelica, Mirko, Bojčić, Ruzica, and Livia Puljak. "Many researchers were not compliant with their published data sharing statement: mixed-methods study." *Journal of Clinical Epidemiology* 150, 2022. DOI: 10.1016/j.jclinepi.2022.05.019.

GitLab. "The DevSecsOps Platform." Why GitLab? | GitLab. Accessed March 22, 2023.

Global Biodata Coalition: https://globalbiodata.org/, Accessed February 1, 2023.

Gomes, Dylan G.E. et al. “Why don't we share data and code? Perceived barriers and benefits to public archiving practices.” *Proc. R. Soc. B.* 289: 20221113, 2022. https://doi.org/10.1098/rspb.2022.1113

Graphics Replicability Stamp Initiative. http://www.replicabilitystamp.org/. Accessed February 22, 2023.

Gregory, Goodey, et al. “The State of Open Data 2022.” January 10, 2023. doi:10.6084/m9.figshare.21276984.v5.

Gruson-Daniel, Celya and Maya Anderson-González. *Étude exploratoire sur la « recherche sur la recherche » : acteurs et approches.* 66 p, Comité pour la science ouverte, 2021, doi:10.52949/24.

Gov.UK. “Intelligent open science: viral genomic data sharing during the COVID-19 pandemic,” https://www.gov.uk/government/publications/intelligent-open-science-viral-genomic-data-sharing-during-the-covid-19-pandemic, November 11, 2022.

HAL, “HAL open science: Open and share the knowledge.” https://hal.science/?lang=en. Accessed February 22, 2023.

Hanna, S., Pither, J., Vis-Dunbar, M. *Open Science. An introduction for Biology*, 2021, https://ubco-biology.github.io/OS-Introduction/

Harriman, Stephanie L., and Jigisha Patel. When are clinical trials registered? An analysis of prospective versus retrospective registration. *Trials* 17, 187, 2016. https://doi.org/10.1186/s13063-016-1310-8.

Henneken, Edwin, and Alberto Accomazzi. “Linking to Data - Effect on Citation Rates in Astronomy.” *Computer Science, Digital Libraries*. 2011. https://doi.org/10.48550/arXiv.1111.3618.

Hicks, Diana, Zullo, Matteo, Doshi, Ameet, and Omar I. Asensio. “Widespread use of National Academies consensus reports by the American public.” *Proceedings of the National Academy of Sciences* 119, e2107760119, 2022. https://www.pnas.org/doi/full/10.1073/pnas.2107760119

Hosseini, Mohammad, et al. “Messing with Merton: The Intersection between Open Science Practices and Mertonian Values.” *Accountability in Research: Ethics, Integrity, and Policy.* 2022. DOI: 10.1080/08989621.2022.2141625

Huang, Chun-Kai (Karl), et al. “Evaluating the Impact of Open Access Policies on Research Institutions.” *ELife,* edited by Julia Deathridge et al., vol. 9, Sept. 2020. https://doi.org/10.7554/eLife.57067.

Hypotheses. “Nanobubbles – ERC Synergy Grant Project.” https://nanobubbles.hypotheses.org/. Accessed February 15, 2023.

Ioannidis, John P.A., Fanelli, Danielle, Dunne, Debbie D. and Steven N. Goodman. “Meta-research: Evaluation and Improvement of Research Methods and Practices.” *PLOS Biology* 13, e1002264, 2015. https://doi.org/10.1371/journal.pbio.1002264.

Ioannidis, John P. A. “Meta-research: Why research on research matters.” *PLOS Biology* 16, e2005468, 2018. https://doi.org/10.1371/journal.pbio.2005468.

International Conference on Software Engineering, https://2021.icse-conferences.org/committee/icse-2023/icse-2023-artifact-evaluation-artifact-evaluation. Accessed on February 19, 2023.

IPOL Journal, Image Processing On Line, ISSN: 2105-1232 https://www.ipol.im/ DOI: 10.5201/ipol

Jeangirard, Eric. “Monitoring Open Access at a national level: French case study.” *ELPUB 2019 23rd edition of the International Conference on Electronic Publishing*, 2019. ⟨10.4000/proceedings.elpub.2019.20⟩. ⟨hal-02141819⟩

Jetten, Mijke, et al. “Professionalising data stewardship in the Netherlands. Competences, training, and education. Dutch roadmap towards national implementation of FAIR data stewardship,” zenodo, 2021, https://doi.org/10.5281/zenodo.4623713.

Joris, van Eijnatten, et al. “Amsterdam Declaration on Funding Research Software Sustainability (0.2).” *Zenodo,* 2022. https://doi.org/10.5281/zenodo.7330542

Journal of Informetrics. Journal of Informetrics. Elsevier Press.

Khoo, Shaun Yoo-Seng. “Article Processing Charge Hyperinflation and Price Insensitivity: An Open Access Sequel to the Serials Crisis.” *LIBER Quarterly: The Journal of the Association of European Research Libraries* 29, 1–18, 2019. https://doi.org/10.18352/lq.10280

Kidwell, Mallory C. et al. “Badges to Acknowledge Open Practices: A Simple, Low-Cost, Effective Method for Increasing Transparency.” *PLOS Biology* 14, e1002456, 2016. Doi: 10.1371/journal.pbio.1002456.

Klebel, Thomas, Fava, Ilaria, and Tony Ross-Hellauer. “Matthew Effects in Open Science and RRI.” 2020. Preprint at https://doi.org/10.5281/zenodo.4155449

Klinkhamer, Laura. “Equity, diversity, and inclusivity in open research (Blog).,” *BMJ Open Science*, 2022. Equity, diversity, and inclusivity in open research - BMJ Open Science

Koltay, Tibor “Data literacy for researchers and data librarians.” *Journal of Librarianship and Information Science*, 49(1), 3–14, 2017. https://doi.org/10.1177/0961000615616450

Kousha, Kayvan and Mike Thelwall. “Are Wikipedia citations important evidence of the impact of scholarly articles and books?” *Journal of the Association for Information Science and Technology* 68, 762–779, 2017. DOI:10.1002/asi.23694.

Kramer, Bianca and Jeroen Bosman, “Innovations in Scholarly Communication - Changing Research Workflows,” *Innovations in Scholarly Communication.* accessed 9 February 2023, https://101innovations.wordpress.com/.

Larivière, Vincent and Cassidy R. Sugimoto. “The Journal Impact Factor: A Brief History, Critique, and Discussion of Adverse Effects.” In W. Glänzel, H. F. Moed, U. Schmoch, & M. Thelwall (Eds.), *Springer Handbook of Science and Technology Indicators* (pp. 3–24). Springer International Publishing. 2019. https://doi.org/10.1007/978-3-030-02511-3_1

Larivière, Vincent, Lebaron, Fredric, Vincent-Lamarre, Philippe, and Julien Larregue. "COVID-19: Where is the data?" *Impact of Social Sciences*, 2020. https://blogs.lse.ac.uk/impactofsocialsciences/2020/11/30/covid-19-where-is-the-data/

Larivière, Vincent, Pontille, David and Cassidy R. Sugimoto. "Investigating the division of scientific labor using the Contributor Roles Taxonomy (CRediT)." *Quantitative Science Studies* 2, 111–128, 2021. https://doi.org/10.1162/qss_a_00097.

Latour, Bruno, and Steve Woolgar. *Laboratory life: The construction of scientific facts*. Salk, Jonas, editor. Princeton University Press, 1986.

Lauranne, Chaignon and Daniel Egret. "Identifying scientific publications countrywide and measuring their open access: The case of the French Open Science Barometer (BSO)." *Quantitative Science Studies*; 3 (1): 18–36, 2022.
https://doi.org/10.1162/qss_a_00179.

Leonelli, Sabina, Spichtinger, Daniel and Barabara Prainsack. "Sticks AND Carrots: Encouraging Open Science at its Source." *Geo*, 2: 12–16, 2015. doi: 10.1002/geo2.2

Leonelli, Sabina. "Global Data Quality Assessment and the Situated Nature of "Best" Research Practices in Biology." *Data Science Journal* 16(32): 1-11, 2017. DOI: 10.5334/dsj-2017-032

Leonelli, Sabina. "Re-Thinking Reproducibility as a Criterion for Research Quality." *Research in the History of Economic Thought and Methodology.* 36: 129-146, 2018. https://doi.org/10.1108/S0743-41542018000036B009.

Leonelli, Sabina and Niccolo Tempini. "Data Journeys in the Sciences." *Springer Nature*, 2020. doi:10.1007/978-3-030-37177-7.

Levin, Nadine, and Sabina Leonelli. "How Does One "Open" Science? Questions of Value inBiological Research." *Science, Technology, & Human Values*, 42:2, 2016. https://doi.org/10.1177/0162243916672071

L'Hôte, Anne and Eric Jeangirard. "Using Elasticsearch for entity recognition in affiliation disambiguation," *arXiv e-prints*, 2021. doi:10.48550/arXiv.2110.01958.

Lopez, Patrice "GROBID: Combining Automatic Bibliographic Data Recognition and Term Extraction for Scholarship Publications." in *Research and Advanced Technology for Digital Libraries* (eds. Agosti, M., Borbinha, J., Kapidakis, S., Papatheodorou, C. & Tsakonas, G.) 473–474, Springer, 2009. doi:10.1007/978-3-642-04346-8_62.

Lopez, Patrice et al. "Mining Software Entities in Scientific Literature: Document-level NER for an Extremely Imbalance and Large-scale Task." *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, 2021. https://doi.org/10.1145/3459637.3481936.

Maddi, Abdelghani and David Sapinho. "On the culture of open access: the Sci-hub paradox." 2022. Preprint at https://doi.org/DOI: 10.21203/rs.3.rs-2357492/v1.

Maggio, Lauren, A., Steinberg, Ryan M., Piccardi, Tiziano, and John M. Willinsky. “Reader engagement with medical content on Wikipedia.” *eLife* 9, e52426, 2020. https://doi.org/10.7554/eLife.52426

Marseille : OpenEdition Press, 2020. Disponible sur Internet : <http://books.openedition.org/oep/14692>. ISBN: 9791036565410. DOI: https://doi.org/10.4000/books.oep.14692.

Massachusetts Institute of Technology “MIT News.” Quantitative Science Studies | MIT Press. Accessed February 12, 2023.

McKenzie, Lindsay. “Open-Access Best Sellers.” *Inside Higher Ed.,* 2018. https://www.insidehighered.com/news/2018/06/01/most-popular-open-access-monographs

McKiernan, Erin C. *et al.* “How open science helps researchers succeed.” *eLife* 5, e16800, 2016 https://doi.org/10.7554/eLife.16800.

Merton, Robert K. “The sociology of science: Theoretical and Empirical Investigations.” *The University of Chicago Pres*, 1973.

Merton, Robert K. “The Matthew Effect in Science.” *Science* 159, 56–63, 1986. https://www.jstor.org/stable/1723414

Merton, Robert, K. “The Ethos of Science.” *J. Legal and Political Sociology.* 1: 115-126. Reprinted In: Merton RK, Sztomka P., editor., editors. Social structure and science, Chicago: University of Chicago Press, 1996.

Metascience 2023 Conference. https://metascience.info/?hsCtaTracking=2d3de6c9-7364-40fd-9c94-62da6696f5e0%7C18c17bf0-3fe3-4d0a-98df-c1b6a66b6def Accessed March 13, 2023.

Metrics- Meta-Research Innovation Center at Stanford. “About Metrics.” https://metrics.stanford.edu/about-us. *Stanford University*. Accessed February 3, 2023.

National Academies, Science Engineering and Medicine. “Reproducibility and Replicability in Science.” https://www.nationalacademies.org/our-work/reproducibility-and-replicability-in-science , Accessed March 3, 2023.

National Institutes of Health. *Supplemental Information to the NIH Policy for Data Management and Sharing: Protecting Privacy When Sharing Human Research Participant Data*, USA.gov, 2022. https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html.

National Institutes of Health. Scientific Data Sharing, USA.gov. Data Management and Sharing Policy | Data Sharing (nih.gov), Accessed April 5, 2023.

National Institute of Informatics. National Institute of Informatics (nii.ac.jp)., Accessed April 6, 2023.

National Institute of Science and Technology Policy (NISTEP). National Institute of Science and Technology Policy (NISTEP)., Accessed April 6, 2023.

National Research Council. "*Bits of Power: Issues in Global Access to Scientific Data.*" Washington, DC: The National Academies Press, 1997. https://doi.org/10.17226/5504.

Netherlands eScience Centre. "Software Management Plans: Practical Guide to Software Management Plans." https://www.escienceCentre.nl/national-guidelines-for-software-management-plans/. Accessed February 26, 2023.

Névéol, Aurélie, W. John Wilbur, and Zhiyong Lu. "Extraction of data deposition statements from the literature: a method for automatically tracking research results." *Bioinformatics*, Volume 27, Issue 23, 1 December 2011, Pages 3306–3312, https://doi.org/10.1093/bioinformatics/btr573

Neylon, Cameron. "Knowledge Exchange Approach towards Open Scholarship." *Zenodo,* 2017. https://doi.org/10.5281/zenodo.826643

Nguyen, Thi-Anh-Hoa, Dechartres, Agnes, Belgherbi, Soraya, and Philippe Ravaud. "Public availability of results of trials assessing cancer drugs in the United States." *J. Clin. Oncol.* 31, 2998–3003, 2013. DOI: 10.1200/JCO.2012.46.9577.

Nicholas, David, et al. "Does the Scholarly Communication System Satisfy the Beliefs and Aspirations of New Researchers? Summarizing the Harbingers Research." *Learned Publishing*, vol. 33, no. 2, pp. 132–41. 2020. https://doi.org/10.1002/leap.1284.

Nicholas, David et. al., "How is open access publishing going down with early career researchers? an international, multi-disciplinary study." *Profesional de la información*, v. 29, n. 6, e290614., 2020. DOI: 10.3145/epi.2020.nov.14

Northwestern, The Kellogg Center. "Pioneering the Science of Science: Science of Science and Innovation." https://www.kellogg.northwestern.edu/research/science-of-science.aspx. Accessed February 12, 2023.

O'Grady, Catherine 'Helicopter research' comes under fire at Cape Town conference." *Science*, 2022. doi:10.1126/science.add3544.

On merrit. "Open Science- who is left behind?" https://on-merrit.eu/ . Accessed March 2, 2023.

OpenAlex. https://openalex.org. Accessed March 3, 2023.

OpenEdition, "Reports" https://www.openedition.org/25480?lang=en. Accessed March 3, 2023.

Ouvrir la science! "Open Science Awards for Open Source Research." *Ministre de l'enseignement supérieur et de la recherche*. https://www.ouvrirlascience.fr/open-science-free-software-award-ceremony/. Accessed February 3, 2023.

Parashar, Manesh, Heroux, Michael A., and Victoria Stodden. Research Reproducilibity. in *Computer*, vol. 55, no. 8, pp. 16-18, 2022. doi: 10.1109/MC.2022.3176988.

Pasquetto, Irene, V., Randles, Bernadette M., and Christine L. Borgman. "On the Reuse of Scientific Data." *Data Science Journal*, 16, p.8. 2017. DOI: http://doi.org/10.5334/dsj-2017-008

Pasquetto, Irene V., Borgman, Christine L., and Morgan F. Wofford. "Uses and Reuses of Scientific Data: The Data Creators' Advantage." *Harvard Data Science Review*, 1(2), 2019. https://doi.org/10.1162/99608f92.fc14bf2d

Pathos Open Science Impact Pathways, "Better understanding and measuring Open Science impacts and their casual mechanisms." Accessed February 23, 2023, https://pathos-project.eu/

Peng, R. "The reproducibility crisis in science: A statistical counterattack." *Significance*, 12, 30–32, 2015. https://doi.org/10.1111/j.1740-9713.2015.00827.x

Pérignon, Christophe, Gadouche, Kamel, Hurlin, Christophe, Silberman, Roxane, and Eric Debonnel. "Certify reproducibility with confidential data." *Science* 365, 127–128, 2019. https://dx.doi.org/10.1126/science.aaw2825

Peroni, Silvio and David Shotton. "Open Citations, an infrastructure organization for open scholarship." *Quantitative Science Studies* 1, 428–444, 2020. https://doi.org/10.1162/qss_a_00023.

Picardy, Tiziano, Redi, Miriam, Colavizza, Giovanni, and Robert West. "Quantifying Engagement with Citations on Wikipedia." in *Proceedings of The Web Conference 2020* 2365–2376, *Association for Computing Machinery,* 2020. doi:10.1145/3366423.3380300.

Pietri, Antoine, Spinellis, Diomidis, and Stefano Zacchiroli. "The Software Heritage Graph Dataset: Large-scale Analysis of Public Software Development History." in *Proceedings of the 17th International Conference on Mining Software Repositories 1*–5, *Association for Computing Machinery,* 2020. https://doi.org/10.1145/3379597.3387510

Piwowar, Heather, Day, Roger, and Douglas B. Fridsma. "Sharing Detailed Research Data Is Associated with Increased Citation Rate." *PLOS ONE* 2, e308, 2007. https://doi.org/10.1371/journal.pone.0000308.

Piwowar, Heather and Todd Vision. "Data reuse and the open data citation advantage." *PeerJ* 1, e175. 2013. DOI: 10.7717/peerj.175/supp-1

Piwowar, Heather et al. "The State of OA: A large-scale analysis of the prevalence and impact of Open Access articles." *PeerJ Preprints*, 2017, https://peerj.com/preprints/3119/ .

PLOS Blogs. "Explore the first Open Science Indicators dataset—and share your thoughts." *The Official PLOS Blog, 2022.* https://theplosblog.plos.org/2022/12/open-science-indicators-first-dataset/

Research Center for Open Science and Data Platform., *Overview of the NII Research Data Cloud.,* Overview of Our Projects | Research Center for Open Science and Data Platform (nii.ac.jp). Accessed April 6, 2023.

Research on Research Institute. https://researchonresearch.org/about. Accessed February 24, 2023.

Research on Research Institute "About the Institute." https://researchonresearch.org/projects#!/tab/273951116-4. Accessed February 24, 2023.

Research Software Alliance. ReSA. https://www.researchsoft.org/. Accessed February 02, 2023.

Research Software Alliance. "Global gathering of research software funders sets the agenda for supporting sustainable research software." https://www.researchsoft.org/blog/2022-11/. Accessed February 25, 2023.

Reveiz, Ludovic, Villanueva, Eleana, Chimaaroke, Ika, and Iventa Simera. "Compliance with clinical trial registration and reporting guidelines by Latin American and Caribbean journals." *Cad Saude Publica* 29, 1095–1100, 2013. PMID: 23778541.

Rochmyaningsih, Dyna. "Did a study of Indonesian people who spend most of their days under water violate ethical rules?" *Science*, 2018. doi: 10.1126/science.aau8972

Ross-Hellauer, Tony. What is open peer review? A systematic review. F1000 Research., 2017. doi: 10.12688/f1000research.11369.2

Ross-Hellauer, Tony. Open science, done wrong, will compound inequities. *Nature* 603, 363–363 ,2022. Open science, done wrong, will compound inequities (nature.com), Accessed March 2, 2023.

Rowhani-Farid, Anisa, Aldcroft, Adrian, and Barnett, Adrian, G., "Did awarding badges increase data sharing in BMJ Open? A randomized controlled trial." *Royal Society Open Science* 7, 191818. DOI: 10.1098/rsos.191818.

Ruediger, Dylan, et. al. "Leveraging Data Communities to Advance Open Science- Findings from an Incubation Workshop Series." *Ithaka S+R:* 1-38, 2022. **https://doi.org/10.18665/sr.317145**

San Francisco Declaration on Research Assessment (DORA), "What is DORA?" Home | DORA (sfdora.org) Accessed February 10, 2023.

Scholarly Infrastructures for Research Software: Report from the EOSC Executive Board Working Group (WG) Architecture Task Force (TF) SIRS. Publications Office of the European Union, 2020. Publications Office of the European Union, https://data.europa.eu/doi/10.2777/28598..

School of Public Policy, Ivan Allen College of Liberal Arts. "Developing Policy Leaders for Our Technology-Driven World." *Georgia Tech*. https://spp.gatech.edu/ Accessed February 13, 2023.

Schooler, Jonathan. Metascience could rescue the 'replication crisis.' *Nature* 515, 9–9, 2014. https://doi.org/10.1038/515009a

Schweinsberg, Martin et al. "Same data, different conclusions: Radical dispersion in empirical results when independent analysts operationalize and test the same hypothesis." *Organizational Behavior and Human Decision Processes,* 165, 228–249, 2021, https://doi.org/10.1016/j.obhdp.2021.02.003.

Science of Science and Computational Discovery Lab. “Upcoming events.” https://scienceofscience.org/ . Accessed February 12, 2023.

Scientometrics: An International Journal of all Quantitative Aspects of the Science of Science, Communication in Science and Science Policy. Scientometrics | Home (springer.com). Springer Press.

Sears, John. “Data Sharing Effect on Article Citation Rate in Paleoceanography.” *American Geophysical Union, Fall Meeting,* 2011. Bib code: 2011AGUFMIN53B1628S

Segado-Boj, Francisco, Prieto-Gutiérrez, Juan-Jose, and Juan Martín-Quevedo. “Attitudes, willingness, and resources to cover article publishing charges: The influence of age, position, income level country, discipline and open access habits.” *Learned Publishing*, **https://doi.org/10.1002/leap.1455**

Simard, Marc-Andre, Ghiasi Gita, Mongeon Philippe, and Vincent Larivière. “National differences in dissemination and use of open access literature.” *PLoS ONE* 17(8): e0272730. https://doi.org/l10.1371/journal.pone.0272730

Singh Chawla, Dalmeet. “The unsung heroes of scientific software.” Nature 529, 115–116, 2016. Doi: 10.1038/529115a

Singh, Harshdeep, West, Robert, and Giovanni Colavizza. “Wikipedia citations: A comprehensive data set of citations with identifiers extracted from English Wikipedia.” *Quantitative Science Studies* 2, 1–19, 2021. https://doi.org/10.1162/qss_a_00105.

Software Heritage. “Building a universal software archive.” https://www.softwareheritage.org/ Accessed February 23, 2023.

Software Preservation: “A Stepping Stone for Software Citation.” https://doi.org/10.25815/0ZBH-2W14/. Accessed February 22, 2023.

Software Sustainability Institute., The Software Sustainability Institute | Software Sustainability Institute. Accessed April 6, 2023.

Strasser Carly et al. “Ten simple rules for funding scientific open source software.” *PLoS Computational Biology* 18(11): e1010627, 2022. https://doi.org/10.1371/journal.pcbi.1010627

Stroe, Oana. “Open data on the rise: the value of EMBL-EBI data resources.” *EMBL-EBI news.* https://ebi.ac.uk/about/news/announcements/value-and-impact-emblebi-2021/. Accessed February 1, 2023.

Susi, Toma, Heigtz Monica, Hnatkova, Eva, Koch, Wolfram, Leptin, Maria, Andler, Martin, Masia Marco, and Michele Garfinkel. “Centrality of researchers in reforming research assessment- Route to improve research by aligning rewards with Open Science practices.” *Initiative for Science in Europe*. Pg. 1-32,2022, 2022-03-16_ise_report_online_final.pdf (initiative-se.eu)

Tacconelli, Evelina *et al.* “Challenges of data sharing in European Covid-19 projects: A learning opportunity for advancing pandemic preparedness and response.” *The Lancet Regional Health - Europe* 21, 100467, 2022.

Teplitskiy, Misha, Lu, Grace. And Eamon Duede. “Amplifying the impact of open access: Wikipedia and the diffusion of science.” *Journal of the Association for Information Science and Technology* 68, 2116–2127, 2017. https://doi.org/10.1002/asi.23687.

Teytelman, Leonid, Stoliartchouk, Alexei, Kindler, Lori, and Bonnie Hurwitz. “Protocols.io: Virtual Communities for Protocol Development and Discussion.” *PLOS Biology* 14, e1002538, 2016. https://doi.org/10.1371/journal.pbio.1002538

Teytelman, Lenny, protocols.io. team, and Anita Broellochs. “How to make your protocol more reproducible, discoverable, and user friendly.” *Protocols.io*., 2020. DOI: dx.doi.org/10.17504/protocols.io.bknmcve.

The Centre for Open Science- Science works best in the Open. “The TOP Guidelines were created by journals, funders, and societies to align scientific ideals with practices.” TOP Guidelines (cos.io). Accessed March 2, 2023.

The Dataverse Project. “Open source research data repository software.” https://dataverse.org/ . Accessed February 21, 2023.

The Journal of Open-Source Software. https://joss.theoj.org/. Accessed February 3, 2023.

The Linux Foundation, “Decentralized innovation.” https://www.linuxfoundation.org/. Accessed March 3, 2023.

Tozzi, Christopher. *For Fun and Profit: A History of the Free and Open-Source Software Revolution*. The MIT Press, 2017, doi:10.2307/j.ctt1t88w3c.

UNESCO Digital Library, “UNESCO Recommendation on Open Science,” UNESCO Recommendation on Open Science - UNESCO Digital Library. Accessed March 2, 2023.

Universite de Montreal “Canada Research Chair on the Transformations of Scholarly Communication.” Program | Canada Research Chair on the Transformations of Scholarly Communication (umontreal.ca). Accessed February 3, 2023.

Van Noorden, Richard. “Scientists call for fully open sharing of coronavirus genome data.” *Nature* 590, 195–196 2021, https://doi.org/10.1038/d41586-021-00305-7

Vincent-Lamarre, Philippe, Boivin, Jade, Gargouri, Yassine, Lariviere, Vincent, and Stevan Harnad. “Estimating Open Access Mandate Effectiveness: The MELIBEA Score.” 2016. Preprint at https://doi.org/10.48550/arXiv.1410.2926.

Vinkers, Christiaan H. *et al.* “The methodological quality of 176,620 randomized controlled trials published between 1966 and 2018 reveals a positive trend but also an urgent need for improvement.” *PLOS Biology*, 19 (4): e3001162, 2021. https://doi.org/10.1371/journal.pbio.3001162.

Wadman, Meredith. “Economic return from Human Genome Project grows.” *Nature*,2013. doi:10.1038/nature.2013.13187.

Wadman, Meredith. “Critics decry access, transparency issues with key trove of coronavirus sequences.” *Science*, 2021, doi: 10.1126/science.abi4496

Wang, Jian, Veugelers, Reinhilde, and Paula Stephan. "Bias against novelty in science: A cautionary tale for users of bibliometric indicators." *Research Policy*, 46: 1416-1436, 2017. https://doi.org/10.1016/j.respol.2017.06.006

Waltman, L. *et al.* "Scholarly communication in times of crisis: The response of the scholarly communication system to the COVID-19 pandemic." https://rori.figshare.com/articles/report/Scholarly_communication_in_times_of_crisis_The_response_of_the_scholarly_communication_system_to_the_COVID-19_pandemic/17125394/1 2021. doi:10.6084/m9.figshare.17125394.v1

Wirsching, Harald. *et al.* "Open for all: exploring the reach of open access content to non-academic audiences," *zenodo*, 2020. doi:10.5281/zenodo.4143313.

Zhang, Liwei and Liang Ma. "Does open data boost journal impact: evidence from Chinese economics." *Scientometrics*, 2021. doi:10.1007/s11192-021-03897-z.

Zhang, Sam, Wapman, K. Hunter, Larremore, Daniel, B., and Aaron Clauset. "Labour advantages lead to the greater productivity of faculty at elite universities." *Science Advances* 8: 1-9, 2022. https://www.science.org/doi/10.1126/sciadv.abq7056

Zotero. "About Zotero." https://www.zotero.org/about/. Accessed March 10, 2023.