

Title: Sticks AND Carrots: Encouraging Open Science at its source

Authors: Sabina Leonelli^{1*}, Daniel Spichtinger², Barbara Prainsack³

Affiliations:

¹Exeter Centre for the Study of the Life Sciences and Department of Sociology, Philosophy and Anthropology, University of Exeter

²European Commission, DG Research and Innovation, Unit A.6. – Science Policy, Foresight and Data.

³Department of Social Science, Health and Medicine, Kings College London

*Correspondence to: s.leonelli@exeter.ac.uk

Abstract: The Open Science (OS) movement has been seen as an important facilitator for public participation in science. This has been underpinned by the assumption that widespread and free access to research outputs leads to (i) better and more efficient science, (ii) economic growth, in particular for small and medium-sized enterprises wishing to capitalise on research findings and (iii) increased transparency of knowledge production and its outcomes. The latter in particular could function as a catalyst for public participation and engagement. Whether OS is likely to help realise these benefits, however, will depend on the emergence of systemic incentives for scientists to utilise OS in a meaningful manner. While some areas, the environmental sciences have a long tradition of open ethos, citizen inclusion and global collaborations, such activities need to be more systematically supported and promoted by funders and learned societies in order to improve scientific research and public participation.

Introduction

OS has been seen as an important factor in facilitating and catalysing public participation in science. Increasing parts of the material that used to be inaccessible but for professional experts are now accessible to wider groups of people. Open access to scientific peer reviewed publications has led this trend, which is now also expanding to original research data. There are, however, still a number of obstacles to OS, which currently hinder the full realisation of its benefits.

At least in principle, OS involves the public dissemination of all elements involved in scientific inquiry, ranging from lab journals and research notes to publications, materials, data, methods/protocols, models, code and software. While not all of these elements may be freely available in all cases, a commitment to facilitate the sharing of these materials underpins the OS movement. This commitment is seen to play a central role in enabling researchers to effectively re-use existing outputs for their own purposes (1), and to foster the intelligibility and reproducibility of research findings across disciplinary boundaries. It also makes it possible for researchers to pick up and continue research that was started, but never completed, by others (2). Finally, it is expected to enhance recognition for the efforts involved in producing research components other than journal publications, which could in turn enhance impact and citations for whoever develops such components (3), and to encourage the use of high standards in research,

e.g. careful data production, well-tested modelling and robust software (4).

At the same time, increasing transparency in research practices can have unintended consequences. Anything that is open to public scrutiny can be used to assess the practices in question, which however may be premature for ongoing projects that need time to yield clear and widely intelligible results. It may also compound researchers' fears of being scooped. It is not hard to imagine that researchers forced to render lab or field notes, protocols, or software freely accessible to others will feel the need to create shadow procedures and infrastructures for those parts of their practice that they do not want, or cannot share (5-7). Finding ways to decide *how* sharing and transparency can be organised to be as fruitful as possible is one of the main challenges at present, while consensus that we need to share seems to have become solid.

Robust funder requirements – for example, as implemented in the US (8) and in the EU's Multiannual Research Framework Programme 'Horizon 2020' (9) – are key to a widespread uptake of OS. So far, however, such mandates pertain mainly to access to *publications*. As part of the changing *modus operandi* of the research system, search engines are replacing discipline-specific journals as the first point of access for most scientists looking for relevant research published by others. It has become anachronistic to address the publication of research findings in scholarly journals as a separate concern from processes through which research is actually conducted: all elements of scientific inquiry, and their different roles in specific phases of research, need to be taken into account. The European Commission, for one, has already started thinking about other means of facilitating OS. It sees OS, sometimes also referred to as 'Science 2.0' or 'Science in transition', as helpful in addressing the Grand Challenges of our times, such as demographic change, climate change, health, food security, clean energy and others mentioned within Horizon 2020 (10). The notion of 'Science 2.0' signifies that every aspect of scientific practice is currently undergoing changes (11). Examples for such changes are the emergence of alternative reputation systems, the growing use of scientific blogs, open annotation, and widening access to data and publications. 'Science 2.0' as a *holistic approach* thus includes much more than Open Access. It represents a paradigm shift in the *modus operandi* of research and science spanning the entire scientific research cycle, from the inception of research to its publication and future use. It also affects the evaluation of the quality and impact of research. For these reasons, the European Commission is conducting a stakeholder consultation on the issue, including an online public consultation (closed September 2014) whose results are currently being analysed.

At the same time, a more widespread uptake of OS requires the support and understanding of researchers on the ground: if OS is perceived by researchers primarily as another piece of bureaucracy imposed by funders, compliance will be a best half-hearted. In addition to the "stick" of compulsory mandates, "carrots" are therefore also needed. This can only be achieved by changes in the scientific culture at large. In particular the following systemic shifts are needed:

- **Recognition of sharing practices in credit structures.** Meaningful sharing takes time; data that are just dumped into a repository without sufficient metadata, annotations, or other relevant information may meet sharing requirement imposed by some funders or universities, but the chances that the data will be discoverable and usable by others are

very low. Thus, effort and time that researchers invest in the meaningful sharing of data, protocols, notes, and results need to be considered in career progression decisions (e.g. promotion, tenure, etc.) and research assessments. Metrics to aid this process need to be developed and implemented (12). OS is, by definition, not a solipsistic activity but a community effort. If assessment metrics for scientific researchers took into consideration the contribution that they make to facilitate the free flow of information and ideas within the scientific community as well as within society as a whole, this would be a strong incentive for people who would like to support OS but cannot afford to (because they, for example, need to focus on activities that will get them tenure instead).

- **Creation of more meaningful incentives for researchers to engage with OS.** A current obstacle to a wider uptake of OS is that many researchers know very little about the variety of formats of OS and their consequences. For example, research institutions should provide systematic training to scientific researchers on practices such as self-archiving, on different formats of data sharing and its advantages and potential risks (in the medical domain in particular, where an important risk is the possible re-identification of individuals), or on how to make information intelligible for specific intended user groups. Moreover, as an increasing number of institutions runs research or teaching initiatives around ‘big data’, it is important that these are not narrowly focused on technical skills such as predictive analytics or data cleaning, but they deal with big data comprehensively, including its societal, ethical, philosophical, and regulatory aspects. This will lead to higher levels of awareness of the potential benefits and drawbacks of OS among scientific researchers as well as among wider publics, which in turn facilitates more meaningful and targeted support of OS. Also within the more limited area of open access to published findings, we have not reached the end of the road. Better incentives to engage with open access publishing of research findings will need to be created, in order to demonstrate to researchers that OS can be a way to reach audiences and users more effectively and that will be beneficial to researchers themselves, for instance as concerns their citation rate (13). The use of social media is important, but by far not the only aspect of this. As incentives are likely to vary depending on the specific features of each research area, both discipline-specific research funders and learned societies have an important role to play in promoting meaningful engagement with OS within their disciplines. We already mentioned the need for medical professionals and others working with personal information to be trained in a kind of ‘social impact assessment’ of sharing research data. Given the discipline-specific nature of the nature and uses of data themselves, learned societies may be better placed to train researchers on modalities of data sharing than universities or funders, as they would be able to develop and provide specific guidelines addressing the concerns and context characteristic of each field.
- **Recognition of the role of alternative metrics (“altmetrics”) and changing publication cultures.** The emergence of data journals and citable repositories, which incentivise the acknowledgement of data production as a research outcome in itself, is one step in this direction, although it is unclear which role such tools will play in future research assessment exercises. Initiatives promoting the publication of models and protocols have lagged behind in comparison to data sharing tools, and the situation is even worse when it comes to materials such as specimens or cell cultures, whose

standardisation and dissemination have mostly been achieved through the open ethos and efforts of specific individuals and communities (when they were achieved at all; 14, 15). Altmetrics provide a potential solution to this issue, even if currently available altmetrics focus more on the resonance of research in social media than on the extent to which authors make the materials and data relating to their publications freely available. Given the considerable effort involved in disseminating research components, universities and funding agencies should ensure that researchers receive targeted support for conducting these activities, for instance in the form of dedicated funding and additional personnel devoted specifically to managing information sharing. Additionally, an important aspect is to ensure the citability of data, in order to ensure that data creators are properly acknowledged (16)

Such strategies should be at the centre of the next wave of policy making on OS. Policy makers have already started to realise some of these suggestions. For example, in a 2012 Recommendation the European Commission clearly outlined the need for systemic change (17), suggesting that Member States should "adjust the recruitment and career evaluation system for researchers and the evaluation system for awarding research grants to researchers so that those who participate in the culture of sharing results of their research are rewarded". The Recommendation also states that if researchers make their findings available through open access, this should be taken into account in relevant assessment procedures. The Commission also encourages the use of new and alternative models of assessing and measuring careers and research activities more generally, especially those that encompass not only the publication of research findings but also data and other types of output.

While the European Commission will continue to support these actions – for instance through a call in the 2014-2015 "Science with and for Society" Work Programme, the bulk of the changes need to be implemented the level of EU Member States, or even at the sub-member states level (depending on the respective research system of the Member State in question). Here, cooperation in the framework of the European Research Area (ERA), which contains an item on "improving knowledge circulation", is potentially very valuable, since it involves both member states and stakeholder groups, such as various associations of European Research Organisations (LERU, Science Europe, EUA, NORDFORSK, CESAR and EARTO). Currently existing approaches that go in the right direction, such as the mandate of the University of Liege for publications, should be explored further and could potentially be supported as "best practice" examples. At the same time, adequate training and support needs to be provided to researchers, so that they are (i) aware of what OS entails and what the potential benefits and concerns are and (ii) they can support OS without having to deal with additional administrative burdens. While some EU supports exists, such as the OpenAIRE and FOSTER projects, research-performing organisations will have a major role to play in providing this support. This could take the form of dedicated information managers and could also involve research libraries. Learned societies also need to take responsibility for helping researchers to identify appropriate infrastructures, publication opportunities and relevant tools that may inform both the planning and the dissemination of their work. In parallel to this development, calls for patients or research participants to be given access to the data that are held about them in clinical and research repositories or within medical devices are becoming more frequent (18, 19).

Within the environmental and life sciences, research communities such as those involved in long-term longitudinal environmental studies and model organism research have long cultivated an ethos of data sharing and open communication, including efforts to disseminate results on a global scale and to include amateur scientists into research initiatives and publications (e.g. 14, 20). These practices emerged in a relatively hostile environment thanks to the perseverance and vision of specific groups of individuals and funding initiatives, and were developed in response to the specific challenges of the research areas in which they emerged, as well as the social and political context in which they were situated. In this sense, they constitute a role model for future research in other areas, and need to be systematically supported and widely publicized by relevant learned societies and funding bodies. This is particularly important in relation to geography, where the modalities and extent of dissemination of results varies enormously across projects and between the physical and human realms, and constructive debate across such research contexts would be highly beneficial to the development of the field as a whole. Physical geography has a long history of data-heavy research and effective data sharing, particularly in areas such as cartography and oceanography (21, 22). Nevertheless, the field as a whole has yet to exploit the vast opportunities offered by recent technological and institutional shifts (23) – a situation partly due to the disciplinary politics underlying large-scale data collection efforts during and after the Cold War (when geology, climatology and environmental science commanded most data initiatives of potential relevance to geographers), and the extent to which commercial and military institutions have taken ownership of such efforts (e.g. 24, 25). The same can be said about human geography, where, moreover, data are typically more sensitive than those collected within physical geography, and researchers need to exercise particular care both in data collection and in the evaluating of which data can be openly disseminated, and how (26). Establishing networks to discuss modes of data sharing and reflect on their implications could be an productive way to bring physical and human geography into closer dialogue with each other, and enable researchers in these areas to learn and profit from each others' skillset (such as experience in handling large datasets in the case of physical geography, and in protecting the confidentiality of data on human subjects in the case of human geography).

The IT revolution – which started several decades ago – continues to reverberate in the scientific system. In order to reap the benefits of OS it needs to be implemented in an institutional context that creates incentives for researchers to share and re-use data, addresses transparency concerns and provides adequate support structures. Merely telling researchers to engage and to learn how to use social media is not in the spirit of a kind of OS that will help public engagement.

References:

1. Royal Society 2012 *Science as an Open Enterprise*. Available at: <http://royalsociety.org/policy/projects/science-public-enterprise/report/> Accessed on 10 November 2014
2. Levin N, Weckowska D, Castle D, Dupre J & Leonelli S Under Review. How do scientists understand openness? Assessing the impact of UK Open Science Policies on biological research.
3. Piwowar HA, Day RS & Fridsma DB 2007 Sharing detailed research data is associated with increased citation rate. *PLoS ONE* 2(3): e308.

4. *Nature Special* 2013 Challenges in Irreproducible Research. Available at: <http://www.nature.com/nature/focus/reproducibility/index.html>. Accessed 10 November 2014
5. Tenopir C et al. 2012 Data sharing by scientists: practices and perceptions. *PloS ONE* 6(6): e21101.
6. Poline JB et al. 2012 Data sharing in neuroimaging research. *Frontiers in Neuroinformatics* 6, 9.
7. Schäfer A et al. 2011 *Baseline Report on Drivers and Barriers in Data Sharing. Opportunities for Data Exchange*. Available at: http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2011/11/ODE-WP3-DEL-0002-1_0_public_final.pdf Accessed 10 November 2014
8. Office of Science and Technology Policy 2013 *Increasing Access to the Results of Federally Funded Scientific Research*. Available at: http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf Accessed 10 November 2014
9. European Commission 2013 *Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020*. Available at: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf Accessed 20 November 2014
10. European Commission 2013 Societal challenges of Horizon 2020. Available at <http://ec.europa.eu/programmes/horizon2020/en/h2020-section/societal-challenges> Accessed 23 November 2014
11. European Commission 2014 *Background document: Public Consultation 'Science 2.0.': Science in Transition*. Directorates-General for Research and Innovation (RTD) and Communications Networks, content and Technology (CONNECT). Available at: <http://ec.europa.eu/research/consultations/science-2.0/background.pdf> Accessed on 23 November 2014
12. Bourne B 2014 Ten Weeks as ADDS. Pebourne Blog. Available at: <http://pebourne.wordpress.com/2014/05/17/ten-weeks-as-adds/>
13. Caball M, Louca L, Pochet R, Spichtinger D & Prainsack B 2013 The best Open Access policies put researchers in charge. *Impact of Social Science* blog. Available at: <http://blogs.lse.ac.uk/impactofsocialsciences/2013/12/18/good-open-access-policies-put-researchers-in-charge/> Accessed on 10 November 2014
14. Leonelli S & Ankeny RA 2012 Re-thinking organisms: The epistemic impact of databases on model organism biology. *Studies in the History and Philosophy of the Biological and Biomedical Sciences* 43, 1: 29-36.
15. Leonelli S 2013 Why the current insistence on Open Access to scientific data? Big Data, knowledge production and the political economy of contemporary biology. *Bulletin of Science, Technology and Society* 33: 1/2, 6-11.
16. Kotarski R, Reilly S, Schrimpf S, Smit E, Walshe K 2012 *Report on Best Practices for Citability of Data and on evolving Roles in Scholarly Communication*. Executive summary. ODE Project. Available at <http://www.alliancepermanentaccess.org/wp->

<content/uploads/downloads/2012/08/ODE-ExecutiveSummary-ReportBestPracticesCitabilityDataEvolvingRolesScholarlyCommunication.pdf> Accessed 23 November 2014

17. European Commission 2012 *Recommendation on access to and preservation of scientific information*. Available at: http://ec.europa.eu/research/science-society/document_library/pdf_06/recommendation-access-and-preservation-scientific-information_en.pdf Accessed on 12 November 2014
18. Lunshof JE, Church GM & Prainsack B 2014 Raw Personal Data: Providing Access. *Science* 343, 6169: 373-374.
19. Dockser Marcus A & Weaver C 2012 Heart Gadgets Test Privacy Law Limits. *The Wall Street Journal online*. Available at: <http://online.wsj.com/news/articles/SB10001424052970203937004578078820874744076?mg=reno64-wsj&url=http%3A%2F%2Fonline.wsj.com%2Farticle%2FSB10001424052970203937004578078820874744076.html>. Accessed 10 November 2014
20. The Long Term Ecological Research Network <http://www.lternet.edu> Accessed 17 November 2014
21. OceanDataPortal <http://www.oceandataportal.org> Accessed 19 November 2014
22. Martin D, Kitchin R and Perkins C 2009 *Re-Thinking Maps*. New York: Routledge.
23. Kitchin R 2013 Big Data and Human Geography: Opportunities, Challenges and Risks. *Dialogues in Human Geography* November 3, 3: 262-267
24. Aronova E, Baker K & Oreskes N 2010 Big Science and Big Data in Biology: From the International Geophysical Year through the International Biological Program to the Long Term Ecological Research (LTER) Network, 1957–Present. *Historical Studies in the Natural Sciences* 40, 2: 183–224.
25. Aronova E 2017 (forthcoming) Geophysical Datascape: The World Data Centers and the Cold War Political Economy of Data Exchange, 1950s – 1960s. *Osiris*
26. Sarah Elwood, Michael F. Goodchild & Daniel Z. Sui (2012) Researching Volunteered Geographic Information: Spatial Data, Geographic Research, and New Social Practice. *Annals of the Association of American Geographers* Vol 102, Issue 3, p.571-590.

Acknowledgments: All views expressed herein are entirely of the authors, do not reflect the position of the European Institutions or bodies and do not, in any way, engage any of them. The authors are grateful to the participants of the COST Strategic Initiative on Open Access (http://www.cost.eu/media/newsroom/open_access) and the Workshop “The Value of Open Science” (<http://www.datastudies.eu/events/15-the-value-of-open-science-workshop>) for inspiring discussions, particularly Gail Davies. SL was funded by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement n° 335925; the UK Economic and

Social Research Council (ESRC) grant number ES/F028180/1; and the Leverhulme Trust grant award RPG-2013-153.