

Open Data in Global Environmental Research: Findings from the Community

Birgit SCHMIDT^{a,1}, Birgit GEMEINHOLZER^b and Andrew TRELOAR^c

^a*University of Göttingen, State and University Library*

^b*University of Giessen*

^c*Australian National Data Service*

Abstract. This paper presents selected findings of the Belmont Forum's survey on open data which targeted the global environmental research and data infrastructure community. It highlights users' perceptions of the term "open data", expectations of infrastructure functionalities, and barriers and enablers for the sharing of data. Respondents also pointed out a wide range of good practice examples and a desire for enhancement and consolidation.

Keywords. Open data, sharing, e-infrastructures, global environmental change

1. Introduction

The Belmont Forum [1], a group of high-level representatives from major funding agencies across the globe, coordinates funding for collaborative research to address the challenges and opportunities of global environmental change. In the course of this, the Belmont Forum E-Infrastructures and Data Management Collaborative Research Action [2] was initiated in 2013, to survey the state of the art and establish recommendations on how the Belmont Forum can leverage existing resources and investments to better foster a more coordinated, holistic, and sustainable approach to the funding of global environmental change research.

Experts from more than 14 countries collectively assessed existing international e-infrastructure capabilities, gaps and overlaps. They prioritized challenges, and provided recommendations for developing and sustaining human and technical international data infrastructures.

In the context of the working group on open data (one of six working groups), a survey invited researchers from various science communities, interested laypersons, government employees, and others who are providing and/or using open data in the scope of environmental change, or are planning/interested in doing so in the future, to share their views and experiences on data publishing, access and (re)use.

The main aim of the survey was to learn more about key open data initiatives of relevance for global environmental change from a data user/provider/manager perspective; areas where users' desire to share could be enhanced by new/other developments; and to detect barriers to "open data sharing" from a user perspective.

¹ Corresponding Author. E-mail: bschmidt@sub.uni-goettingen.de.

2. Methods and Results

From September to November 2014, the survey collected 1,330 responses through a web form [3]. Of these, 1,253 qualified as valid responses. The survey was distributed to about 20 disciplinary and professional mailing lists, and to all the authors of a well-renowned open access publisher, central to the research area (Copernicus Publications). A potential bias in the responses should be taken into account as the participants of the survey might not be representative of the community, and might also be more positive towards the topic of “open data” than the average researcher. Note that “open data” was not defined in the survey – respondents were free to respond on the basis of their own understanding.

Table 1 indicates the regional distribution of the collected data.

Table 1. Population of the survey

	Frequency (N=1248)	Percentage
Germany	205	16.4
United States	184	14.7
Italy	117	9.4
United Kingdom	88	7.1
France	68	5.5
Australia	45	3.6
Spain	43	3.4
China	39	3.1
Other countries (76)	459	36.8

As expected, the majority of respondents belonged (multiple answers were allowed) to earth and environmental sciences (67.5%, 846 answers) as well as climate and atmospheric sciences (30.8%, 386 answers). In addition, there were 50 or more answers from the biological sciences (20.6%, 258 answers), physical sciences (12.9%, 162 answers), engineering (7%, 88 answers), computer sciences (6.8%, 85 answers), social sciences (5.3%, 66 answers), agricultural and veterinary sciences (4.2%, 53 answers) and chemical sciences (4%, 50 answers).

The survey results highlight the users’ perspective about the term “open data”, such as the importance of information that enables users to assess the quality of data (82% very important), to select data based on metadata (78%), and to easily access (76%) and reuse the data (70%). The provision of unrestricted data was considered as very important by 2 out of 3 respondents. Open data already seems to be of substantial relevance for the global environmental change research community as more than 4 out of 5 respondents considered open data as important for advancing research. Half of the respondents saw open data as important for supporting applications to societal problems.

Moreover, motivators and barriers to publish data as open data were analyzed. The desire to publish data as open data was mainly linked to research-intrinsic motives ranging from general considerations, i.e. the acceleration of scientific research and applications, to personal motivations, i.e. dissemination and recognition of research results, personal commitment to open data and requests from data users (see Fig. 1).

From all the policies, funder policies seem to be the most important motivator, supporting the conclusion that stronger mandates will likely further strengthen the case for data sharing.

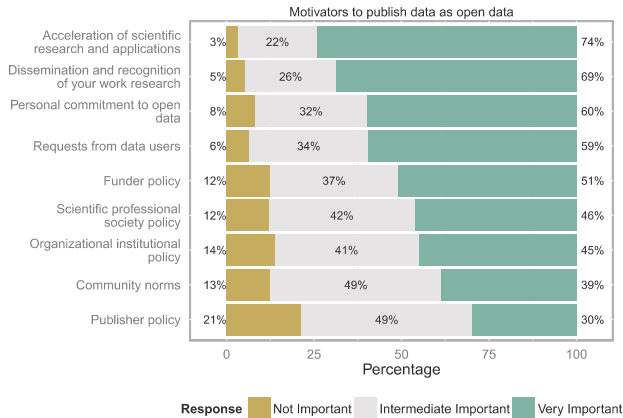


Figure 1. Motivators to publish data as open data.

The most important barriers (see Fig. 2) were the desire to publish results before releasing data (54% very important), legal constraints (47%), loss of credit and recognition (41%) and possible misinterpretation or misuse (37%). The ranking of these perceived barriers varied across fields, e.g. legal constraints were the most important barrier in economics, computer sciences and engineering. As expected, the desire to publish results before releasing data was most prevalent at early stages of a research career, i.e. the age from early to mid 30s, and was perceived as a major barrier by 69% of all respondents.

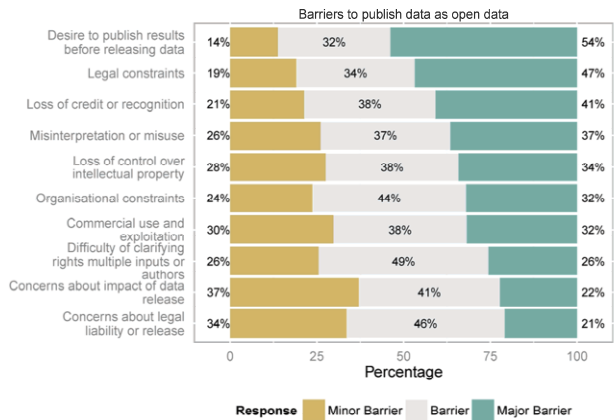


Figure 2. Barriers to publish data as open data.

When it comes to expectations towards infrastructure, the most important functionalities are that authorship and attribution are highlighted (75% most important, 23% intermediate important), data are citable via persistent identifiers (73% and 25% resp.), links to publications are provided (63% and 35% resp.) and restrictions, conditions and/or licensing information is communicated (61% and 36% resp.). In addition, a wide range of good practice examples was pointed out by respondents,

Our findings confirm that accelerating research and applications and scientific merits are the main motivators for publishing data as open data. The high number of examples provided and the wishes expressed by respondents also seem to indicate that global environmental research highly depends on data sharing.

Among all policies, funder policies were ranked first as a motivator, and 88% of all respondents acknowledged their importance. Therefore it seems that acceptance of open data could be further enhanced by making open data archiving mandatory which is currently not the case for funders in several countries. However, this will be one recommendation of the Belmont Forum E-Infrastructures and Data Management Collaborative Research group to the Belmont Forum to support global environmental change research. Although ranked lowest among policies as a motivator, references in journals were the top route for the discovery of data (followed by search engines and data repositories and other discovery routes), and were also mentioned several times in the free text comments. This however should not lead to an encouragement of publishers to establish commercial databases for data storage, as paying for data access was not well perceived by the respondents.

Based on the findings of the survey, we have made the following recommendations to the Belmont Forum:

- that funders should make open data archiving mandatory, to take into account the main motivators revealed by the survey,
- to strengthen support and training activities,
- to further facilitate interoperability between data infrastructures, and
- to support the long-term sustainability of archives and data infrastructures.

Acknowledgements. The authors would like to thank Jonathan Hodge and André Santachè for their contribution to the design of the survey, Kim Oakley for collecting the data, and all survey respondents for their input.

References

- [1] Belmont Forum, available at: <https://igfagcr.org/> [Accessed 14 May 2015].
- [2] Belmont Forum E-Infrastructures and Data Management Collaborative Research Action, available at: <http://www.bfe-inf.org/> [Accessed 14 May 2015].
- [3] B. Schmidt, B. Gemeinholzer, A. Treloar, J. Hodge, A. Santanchè, K. Oakley (2015). Belmont Forum Open Data Survey 2014. Zenodo. doi:10.5281/zenodo.16384.
- [4] S. Sinclair, G. Rockwell, Cirrus, *Voyant*, 2015. Available at: <http://voyant-tools.org/tool/Cirrus/> [Accessed 14 May 2015].
- [5] H.A. Piwowar, R.S. Day, D.B. Fridsma. Sharing detailed research data is associated with increased citation rate. *PLoS ONE* 2(3) (2007), e308 (2007). doi:10.1371/journal.pone.0000308
- [6] C. Tenopir, S. Allard, K. Douglass, A.U. Aydinoglu, L. Wu et al., Data sharing by scientists: practices and perceptions, *PLoS ONE* 6(6) (2011), e21101. doi:10.1371/journal.pone.0021101.
- [7] N. Enke, A. Thessen, K. Bach, J. Bendix, B. Seeger, B. Gemeinholzer (2012), The user's view on biodiversity data sharing – Investigating facts of acceptance and requirements to realize a sustainable use of research data, *Ecological Informatics* 11 (2012), 25–33. doi:10.1016/j.ecoinf.2012.03.004
- [8] B. Fechner, S. Friesike, M. Hebing. What drives academic data sharing? *PLoS ONE* 10(2) (2015), e0118053. doi:10.1371/journal.pone.0118053.
- [9] V. Van den Eynden, V., L. Bishop, Incentives and motivations for sharing research data, a researcher's perspective, 2014. A Knowledge Exchange Report, available at: <http://knowledge-exchange.info/Default.aspx?ID=733> [Accessed 14 May 2015].