

Publisher interest towards a role for journals in data-sharing: The findings of the JoRD Project

Marianne Bamkin^{a1}, Jane Smith^a, Paul Sturges^b, Azhar Hussain^a, Bill Hubbard^a
^a*Centre for Research Communication, University of Nottingham*
^b*Emeritus Professor, Loughborough University*

Extended Abstract

Introduction

The sharing of the data generated by research projects is increasingly being recognised as an academic priority by funders and researchers. For example, out of 110 listed funders on the JULIET² service, 32 have data policies of some form [1]. The topic has been discussed by national and international organisations, for example, ICSU (the International Council for Science), [2] the OECD (Organisation for Economic Co-operation and Development) [3] and the UK's Royal Society [4]. The public statements that emerge from these scientific bodies call for both research transparency and freely available access to research data created with public funding for possible re-use. The rights associated with the sharing of data and the environment in which it can be done is also of interest to publishers. This interest can be attributed to two motivating factors: to support the academic function of data such as the corroboration of research findings and the facilitation of the re-use of data; and to respond to a strategic, commercial development, for instance, an engagement with the rights, process and environment of data sharing. Currently some publishers are introducing contractual policies on the archiving and sharing of data in addition to policies governing the deposit and sharing of research articles through repositories [5]. The issue of policies on sharing set out by academic journals has been raised by scientific organisations, such as the US National Academy of Sciences, which urges journals to make clear statements of their sharing policies [6]. On the other hand, the publishing community whilst broadly supporting the principle of open and accessible research data expresses concerns over the intellectual property implications of archiving shared data.

The JoRD Project was a feasibility study on the possible shape of a central service on journal research data policies, funded by the UK JISC under its Managing Data Research Programme. It was carried out by the Centre for Research Communication at Nottingham University (UK) with contributions from the Research Information Network and Mark Ware Consulting Ltd. The project used a mix of methods to examine the scope and form of a sustainable, international service that would collate

¹ Corresponding Author.

² <http://www.sherpa.ac.uk/juliet/>

and summarise journal policies on research data for the use of researchers, managers of research data and other stakeholders. The purpose of the service would be to provide a ready reference source of easily accessible, standardised, accurate and clear guidance and information, on the journal policy landscape relating to research data. The specific objectives of the study were: to identify the current state of journal data sharing policies; to investigate the views and practices of stakeholders; to develop an overall view of stakeholder requirements and possible service specifications; to explore the market base for a JoRD Policy Bank Service; and to investigate and recommend sustainable business models for the development of a JoRD Policy Bank Service.

1. Consultation of Literature

A review of relevant literature was conducted in order to: define the problem; identify existing data sharing initiatives; discover the current state of data sharing policies; gain an understanding of the issues arising from different data types; and gather information which might inform the design of applicable business models and sustainability options. The review suggested that the subject of journal data sharing policies has not been greatly researched. What does emerge is evidence that scientific institutions are attempting to draw attention to the importance of journal data policies and a sense that the scientific community in general is in favour of the concept of data sharing [7]. At the same time it appears that more needs to be done to alert the publishing world that should it chose to engage with the academic drive towards data sharing, greater consistency in data policy and author guidelines are necessary, with a focus on questions such as when and where authors should deposit data for sharing.

Such studies of journal policies as do exist found that a large percentage of journals do not have a policy on data sharing, and that there are inconsistencies between the traceable journal data sharing policies. They report that whilst some journals offered little guidance to authors, others stipulated specific compliance mechanisms. A valuable distinction is made in some policies between two categories of data: *integral*, which directly supports the arguments and conclusions of the article, and *supplementary*, which enhanced the article, but was not essential to its argument. What we judged the most significant study on journal policies [8], defined journal data sharing policies as “strong”, “weak” or “non-existent”. A strong policy stipulates the deposit of data as a condition of publication, whereas a weak policy merely requests the deposit of data. There is some indication from previous studies that researchers’ data sharing behaviour is as inconsistent as the journal policies [9] [10]. However, there is general assent to the data sharing concept and many researchers would be prepared to submit data for sharing along with the articles they submit to journals.

2. Examination of journal policies

A substantial sample of journal policies was then investigated in order to establish our own picture of the journal policy landscape. A selection of 400 international and national journals were purposefully chosen to represent the top 200 most cited journals

(high impact journals), and the bottom 200 least cited (low impact journals), equally shared between Science and Social Science, based on the Thomson Reuters citation index. The data policy of each journal was broken into different aspects such as: what, when and where to deposit data; accessibility of data; types of data; monitoring data compliance and consequences of noncompliance. These were then systematically entered onto a matrix for comparison. Where no policy was found, this was indicated on the matrix. Policies were categorised as either being “weak”, only requesting that data is shared, or “strong”, stipulating that data must be shared.

Approximately half the journals examined had no data sharing policy. Nearly three quarters of the policies found were assessed as weak and only just under one quarter were deemed to be strong (76%: 24%). The high impact journals were found to have the strongest policies, whereas not only did fewer low impact journals include a data sharing policy, those policies were less likely to stipulate data sharing, merely suggested that it may be done. The policies generally give little guidance on which stage of the publishing process that data is expected to be shared.

3. Stakeholder consultations

Throughout the duration of the project, representatives from publishing and other stakeholders were consulted in different ways. Representatives of publishing were selected from a cross section of different types of publishing house; the researchers which were consulted were self-selected through open invitations by way of the JoRD Blog. Nine of them attend a focus group and 70 answered an online survey. They were drawn from every academic discipline and ranged over a total of 36 different subject areas. During the later phases of the study, a selection of representatives of stakeholder organisations was asked to explore the potential of the proposed JoRD service and to comment on possible business models. These included publishers, librarians, representatives of data centres or repositories, and other interested individuals. This aspect of the investigation included a workshop session with representatives of leading journal publishers in order to discuss the potential for funding a JoRD Policy Bank service. Subsequently an analysis of comparator services and organisations was performed, using interviews and desk research.

4. Conclusion

Our conclusion from the various aspects of the investigation was that although the idea of making scientific data openly accessible for sharing is widely accepted in the scientific community, the practice confronts serious obstacles. The most immediate of these obstacles is the lack of a consolidated infrastructure for the easy sharing of data. In consequence, some researchers quite simply do not know how to share the data that supports published journal articles. Data sharing practices and infrastructure are not reliant on publishers, because funders' and institutional policies and developing support systems also contribute to the sharing environment. However, in such a dynamic landscape, the researcher needs clear policies and guidance and where publishers choose to engage with the academic need for sharing, the case for clear policies to be articulated by journals seems unchallengeable. At the present juncture, when publishers

policies are either not available, or provide inadequate guidance, researchers acknowledge a need for the kind of information that a policy bank would supply. The market base for a JoRD policy bank service would be the research community, and researchers indicated that such a service would be well used. At the same time, because publishers are not convinced that the infrastructure and financial models for sharing are sufficiently robust to guarantee a high quality sharing environment they expressed some willingness to provide financial support for a service that would make the sharing process transparent and consequently simplify their own engagement with sharing.

Three levels of possible business models for a JoRD service were identified and finally these were put to a range of stakeholders. These stakeholders found it hard to identify a clear cut option of service level that would be self sustaining. The funding models of similar services and organisations were also investigated. In consequence, an exploratory two phase implementation of a service is proposed. The first phase would be the development of a database of data sharing policies, engagement with stakeholders and API development with the intention to build use to the level at which a second phase, a self sustaining model, would be possible.

References

- [1] JULIET: <http://www.sherpa.ac.uk/juliet.>, [Accessed 20 January 2013]
- [2] International Council for Science. ICSU Report of the CSPR Assessment Panel on Scientific Data and Information. 2004:1-43.
- [3] OECD. OECD Principles and Guidelines for Access to Research Data from Public Funding 2007:1-22.
- [4] Royal Society. Science as an Open Enterprise. 2013; Available at: http://royalsociety.org/uploadedFiles/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf. [Accessed 01 February 2013]
- [5] Cicerone, Ralph. Ensuring Integrity in Science. Science, 2010. Available at: <http://www.nasonline.org/about-nas/leadership/president/cicerone-editorial-science.pdf>. [Accessed 12 March 2013]
- [6] National Academy of Sciences. Sharing Publication-Related Data and Materials: Responsibilities of Authorship in the Life Sciences. 2003; Available at <http://www.nap.edu/catalog/10613.html>. [Accessed 12 March 2013]
- [7] Hrynaszkiewicz, Ian. The need and drive for open data in biomedical publishing. 2011; Serials 24, 31-37
- [8] Piwowar, Heather. and Chapman, Wendy. A Review of Journal Policies for Sharing Research Data. Open Scholarship: Authority, Community and Sustainability in the Age of Web2.0 - Proceedings of the 12th international Conference on Electronic Publishing (ELPUB); June 25-27; Toronto: ELPUB; 2008.
- [9] Spencer, Hilary. Thoughts on the sharing of data and research materials and the role of journal policies. 2010. Available at http://www.stanford.edu/~vcs/Nov21/hilary_spencer_rdcscsJan2010.pdf. [Accessed 12 March 2013]
- [10] Kuipers, Tom and van der Hoeven, Jeffrey. PARSE: Insight into issues of permanent access to the records of science in Europe. Survey report. 2009; Brussels: European Commission.