PASTEUR4OA Briefing Paper: Disciplinary differences in opening research data

Author: Danny Lämmerhirt (OKF)

Reviewers: Alma Swan (EOS) and Mafalda Picarra (Jisc)

April 2016

Introduction

The management and widespread sharing of publicly funded research data has gained significant momentum[1] among governments, funders, institutions, journals and data service providers around the world, including the European Commission (EC)[2], the United States[3], the OECD[4] and the G8[5]. Central is the idea that publicly funded open research data should be openly published, accessible and reusable with the least technical and legal restrictions possible, while protecting privacy, safety, and commercial interests. Contrary to published research outputs, research data are commonly understood as the raw material leading to scientific insights. The EC states in its "Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020" that open access to research data allows to improve the quality of research results, helps to avoid duplication of effort, speeds up innovation and increases the transparency of the scientific process.[6]

However, there is no 'one-size-fits-all' approach to open research data across academic disciplines. Different disciplines produce different types of data and have various procedures for analysing, archiving and publishing it. Some have established data management procedures, norms or policies, making their research data open by default, while others do not.[7] Manifold factors influence which data are opened to what degree in a specific discipline.

---

1 See also: Comprehensive Brief on Research Data Management Policies:
http://www.science.gc.ca/default.asp?lang=En&n=1E116DB8-1
2 EC Commission Recommendation on Access to and Preservation of Scientific Information
http://ec.europa.eu/research/science-society/document_library/pdf_06/recommendation-access-and-preservationscientific-information_en.pdf
3 Increasing Access to the Results of Federally Funded Scientific Research, Office of Science and Technology Policy
http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf
4 OECD Ministers of Science and Technology: Access to Research Data From Public Funding: **http://bit.ly/1lGubJQ**
5 G8 Science Ministers Statement London UK, 12 June 2013: **https://www.gov.uk/government/news/g8-science-ministers-statement**
6 Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020:
https://www.openaire.eu/guidelines-on-open-access-to-scientific-publications-and-research-data-in-horizon-2020
7 The conundrum of sharing research data: **http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1869155**

In a research series examining data sharing practices in selected disciplines, the Research Information Network (RIN) states that

'if the policies and strategies of research funders, universities and service providers are to be effective in optimising the use and exchange of scholarly information, they must be sensitive to the practices and cultures of different research communities.'[8]

This briefing paper presents the current state of open research data across academic disciplines. It describes disciplinary characteristics inhibiting a larger take-up of open research data mandates. Additionally, it presents the current strategies and policies established by funders, institutions, journals and data service providers alongside general data policies.

Discipline-specific data policies: An overview

The key policy instrument to foster good data practice, publication and reuse is the data management plan (DMP), which is part of the data lifecycle. Increasingly, funders and institutions have established data policies mandating DMPs. For instance, the EC, the Research Councils of the United Kingdom (RCUK), Oxford University and the University of Heidelberg. These policies commonly state that researchers should plan *which data* should be *generated, collected*, *prepared*, *managed*, *preserved*, *made accessible* and *re-used, to what extent,* for *which time-span and with which data service provider*.[9]

Often, it is the researcher's individual responsibility to establish a DMP in accordance with the 'community best practice' and aligned with legal, ethical and discipline-specific standards and procedures. Besides these general notes, data policies sometimes are unclear about what such disciplinary standards and processes entail (with problematic implications for disciplines lacking such standards, see also section "Disciplinary data management practices, standards and norms of data sharing").

Funders like the UK's Economic and Social Research Council (ESRC) and the Biotechnology and Biological Sciences Research Council (BBSRC) or the German Research Foundation (DFG) embrace disciplinary characteristics more actively. Their data policies and guidelines may include topical data management solutions (DFG), recommend suitable data service providers that can offer guidance and expertise in managing data, or legal guidelines for data publication and licensing.[10] The BBSRC also prescribes particular data types that grantees have to publish. Increasingly, institutional and discipline-specific repositories as well as journals run their own data management services and provide open licences. Prominent repositories for research data are Dryad (for life and medical sciences), Open Context (for archaeology), and Pangea (for earth sciences).[11] Exemplary journals include the Public Library of Science (PLOS), Ubiquity Press Journals, Molecular Ecology or the American Economic Review. The latter actively enforces the publication of data and states that it will publish papers only if

---

8 Reinventing Research? Information Practices in the humanities: **http://www.rin.ac.uk/our-work/using-and-accessing-information-resources/information-use-case-studies-humanities**

9 A detailed list of data management policies may be found here: Comprehensive Brief on Research Data Management Policies: **http://www.science.gc.ca/default.asp?lang=En&n=1E116DB8-1**

10 BBSRC Data Sharing Policy: **http://www.bbsrc.ac.uk/about/policies/policy-foi/policy/data-sharing-policy/**; ESRC Research Data Policy: **http://www.esrc.ac.uk/funding/guidance-for-grant-holders/research-data-policy/** and DFG's guidelines on the "Handling of Research Data": **http://www.dfg.de/en/research_funding/proposal_review_decision/applicants/submitting_proposal/research_data/index.html**

11 The Registry of Research Data Repositories provides a comprehensive overview of repositories with open licenses: **http://www.re3data.org/**

"data used in the analysis are clearly and precisely documented and are readily available to any researcher for purposes of replication."[12] Thereby, the amount of published data could be increased.[13]

As part of the EC's strategy for opening science[14], the EC is running an Open Data Pilot (ODP) in Horizon 2020. The ODP aims to maximise access and reuse of research data in selected focus areas. It mandates that Horizon 2020 grantees must provide a DMP and deposit at least all research data underlying final research results, free of charge, in a repository, enabling the use and re-use of scientific research results.[15] The Open Access Infrastructure for Research in Europe (OpenAIRE) provides support to link publications to underlying research data and the EUDAT B2SHARE tool helps researchers to choose an adequate licence for research data such as CC-BY or CC0.[16] With general policies and a supporting infrastructure the ODP aims to encourage good data management and may be regarded as "as open as possible, as closed as necessary".

There is considerable heterogeneity amongst research data policies, some of them addressing disciplinary differences rather generically, others aligning their policies proactively with discipline-specific requirements. Some are aspirational or normative in scope, while others actively enforce particular data management practices. But while surveys repeatedly show that researchers support these normative efforts, compliance with data policies is patchy at best.

The state of open research data

Whether researchers publish data varies significantly across academic disciplines. A recent survey by the academic publisher Wiley investigated the likelihood of researchers to share research data with others. It stated that 66 percent of all life scientists share their data with other researchers, mostly in dedicated repositories or journals, while only 36 percent of all surveyed social scientists and humanities scholars do, often informally sharing them within their own research team or institution.[17]

Other studies show similar findings and repeatedly demonstrate that individual practices of data publication are seldom compliant with data policies, except from disciplines like genetics, astronomy or crystallography (see "Research on discipline-specific data publication" in the further readings-section). These studies often only regard the sharing of research data in broad terms, including informal sharing practices among trusted researchers. It is fair to say that data publications under open licences like CC-BY or CC 0 are significantly low in number. Reports by the Research Information Network (RIN), the Digital Curation Centre (DCC), the Knowledge Exchange Group (KE) and various individual research efforts commonly agree upon a range of factors that influence the likelihood of research data being published in distinct disciplines (see further readings). These factors include discipline-specific data management practices, disciplinary norms and cultures of data sharing, economic factors, career-related factors, infrastructural factors, and legal, commercial and ethical questions of public access.

---

12 American Economic Review - Data Availability Policy: **https://www.aeaweb.org/aer/data.php**
13 Open Data in Economics: The Basis of Reproducible Research:
http://blogs.lse.ac.uk/impactofsocialsciences/2013/10/23/open-data-in-economics-the-basis-of-reproducible-research/
and Researcher data sharing insights: **http://exchanges.wiley.com/blog/2014/11/03/how-and-why-researchers-share-data-and-why-they-dont/**
14 PASTEUR4OA Discussion Paper: Researchers and Open Science:
http://www.pasteur4oa.eu/sites/pasteur4oa/files/resource/Discussion%20Paper_Researchers%20and%20Open%20Science.pdf
15 Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020:
https://www.openaire.eu/guidelines-on-open-access-to-scientific-publications-and-research-data-in-horizon-2020
16 **https://b2share.eudat.eu/?ln=en**
17 Researcher data sharing insights: **http://exchanges.wiley.com/blog/2014/11/03/how-and-why-researchers-share-data-and-why-they-dont/**

Disciplinary data management practices, standards and norms for data sharing

In some disciplines standardised data is tightly interlinked with norms for data sharing. Sometimes community-specific policies play an important role. The Bermuda Principles and the Fort Lauderdale Principles established worldwide standards for gene sequencing and regulate the period within which data has to be published. The climate models underlying the reports of the International Panel of Climate Change (IPCC) also rely on the archival of highly standardised (meta-)data enabling comparability and compatibility across different modelling rounds.[18]

Sometimes the nature of a discipline makes it scientifically necessary to share the data underlying research publications. Disciplines like palaeogenetics or forensics have a strong peer-pressure to publish data in order to verify data quality and the knowledge claims derived from them. Well-established data standards, publication procedures and specialised repositories are well-integrated and support the publication of research data in these disciplines.[19]

In other cases disciplines may lack a coherent understanding of data and data practices may vary significantly.[20] Evidence shows that social scientists often collect and manage data on an ad-hoc basis, which is adapted to the researcher's immediate needs to publish findings and which is often stored locally on a hard disc.[21] Similarly, arts and humanities scholars who work with heterogeneous data are generally more concerned with linking and finding dispersed data sources than with publishing data and they rarely use digital repositories.[22] Those 'small sciences' may develop local and unique data management practices without agreed-upon standards, and researchers tend to generally work individually rather than in larger teams.[23] This is aggravated by the fact that researchers across disciplines consider time pressure and costs as being the main obstacles for data sharing. Requirements to modify data for long-term storage may be considered an additional administrative burden.[24]

Some funders, institutions and data publishers complement general data policies with specific measures to actively foster data management. In the social sciences, the ESRC-funded UK Data Service and the Inter-University Consortium for Political and Social Research (ICPSR) provide guidance for data preparation, preservation and publication as well as best practices for the use of data.[25] Similarly the web-portal biosharing.org provides data exchange formats for submitters in cases where no community standards exist. The Data Documentation Initiative (DDI) develops a structured machine-readable metadata standard for the social sciences and humanities that is increasingly used by major repositories and recommended by ESRC to grantees.[26]

Other policies acknowledge the importance of librarians and data managers in helping to tackle time constraints and a potential lack of data management expertise prevailing in some disciplines. The DataONE network, funded by the US National Science Foundation (NSF), collaborates with institutional librarians who understand discipline-specific needs. The aim is to foster an increased interaction

---

18 The Case for International Sharing of Scientific Data: A Focus on Developing Countries: Proceedings of a Symposium: **http://www.nap.edu/catalog.php?record_id=17019**
19 When Data Sharing Gets Close to 100%: What Human Paleogenetics Can Teach the Open Science Movement: **http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0121409**
20 The conundrum of sharing research data: **http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1869155**
21 The Problem of Data: **http://www.clir.org/pubs/reports/pub154**
22 Reinventing Research? Information Practices in the humanities: **http://www.rin.ac.uk/our-work/using-and-accessing-information-resources/information-use-case-studies-humanities**
23 Data sharing, small science and institutional repositories: **http://rsta.royalsocietypublishing.org/content/368/1926/4023**
24 Data Sharing by Scientists: Practices and Perceptions: **http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0021101**
25 The Guide to Social Science Data Preparation and Archiving: **https://www.icpsr.umich.edu/icpsrweb/content/deposit/guide/**
26 For Attribution -- Developing Data Attribution and Citation Practices and Standards: Summary of an International Workshop: **http://www.nap.edu/catalog/13564/for-attribution-developing-data-attribution-and-citation-practices-and-standards**

between researchers and librarians and to improve data management.[27] In a similar vein, several British funders plan to encourage and fund career paths for data managers and data scientists within institutions, aiming to increase institutional research data management capacities.[28]

Data policies may stimulate the use of specific data service providers or provide funding for subject repositories. Active funding of repositories may reinforce existing sharing practices, as demonstrated by repositories like ChemSpider, funded by the Royal Society of Chemistry.[29]

Career-related incentives and individual perceptions

Studies suggest that researchers publish data mainly for three reasons: due to community pressure, to make scientific advances in their disciplines, or to advance in their career.[30] Research publications are still the main channel to build reputation and secure promotion across disciplines. In competitive environments where publication pressure prevails, researchers are more protective of their data, sometimes considering them as intellectual capital, accumulated through considerable effort and time. Researchers may want to preserve exclusive rights to publish research findings first and often they consider themselves as 'data owners'. The fear of being 'scooped' by competitors is common and in some disciplines researchers fear that their data could be misused or misinterpreted. Researchers across disciplines state that they are willing to share data as long as they can keep control over it.[31] There is a difficult-to-reconcile conflict of interest here.

Policies therefore provide individual researchers with the possibility to restrict the use of data (embargoed publication, safeguarded access to data).[32] Journals and data service providers such as the American Economic Review, the UK Data Service or Dryad have specific terms of access and use, ranging from restricted access models to open licences like CC-BY or CC0 and to waiving copyright, attribution or other conditions such as "share-alike".[33] An alternative approach to the common publication system is to incentivise the publication of data by developing merit systems. These systems foster a formal data citation, including Digital Object Identifiers (DOIs) and ORCIDs as well as data-level metrics. Some data service providers allow complex datasets to be cited at different levels of detail, acknowledging discipline-specific requirements.[34] The "Making Data Count" project"[35] and Thomson Reuters' Data Citation Index[36] develop data-level metrics to track and measure data use across repositories. Sufficient acknowledgement of researchers can thereby be ensured through attribution,

---

27 DataOne for Librarians: **https://www.dataone.org/for-librarians**

28 Establishing incentives and changing cultures to support data access: Response from the EAGDA funders: **http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/EAGDA/WTP056496.htm**

29 **http://www.chemspider.com/**

30 Sowing the seed: incentives and motivations for sharing research data: researcher's perspectives: **http://knowledge-exchange.info/event/sowing-the-seed**

31 What Drives Academic Data Sharing: **http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0118053**

32 The Digital Curation Centre provides an overview over data sharing policies, also including demands for timeliness of data publication at: **http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies**

33 More open repositories are Dryad for Life Sciences and Health Sciences and Pangea for Earth Sciences, while the UK Data Service provides the majority of its data under a 'safeguarded access': **https://www.ukdataservice.ac.uk/get-data/data-access-policy**. The use of CC-SL (share-alike) licenses may cause legal conflicts in cases where derivative work wants to avoid using an open license and restricts the re-usability of research data to cases where open licensing is applicable. For further information, see also: Safe to be open Study on the protection of research data and recommendations for access and usage: **https://www.openaire.eu/search/publication?articleId=od_____323::d052becf1674cfad70faca45f07fa5fa**

34 For Attribution - Developing Data Attribution and Citation Practices and Standards: Summary of an International Workshop: **http://www.nap.edu/catalog.php?record_id=13564**

35 Making Data Count: Developing a Data Metrics Pilot: **http://blogs.plos.org/tech/making-data-count-plos-cdl-and-dataone-join-forces-to-build-incentives-for-data-sharing/**

36 The Data Citation Index: **http://wokinfo.com//products_tools/multidisciplinary/dci/?utm_source=false&utm_medium=false&utm_campaign=false**

citation and co-authorship, depending on the importance of open research data in the findings communicated in a scientific article.[37]

Legal and ethical issues

Legal and ethical issues are primary factors influencing data sharing practices and it is commonly accepted that different data need different scales of security and access. In some disciplines, the disclosure of research data may cause harm, such as the possible destruction of nesting sites for endangered species or archaeologic excavation sites.[38] Project-specific policies such as the International Polar Year 2007-2008 Data Policy mandated that such data have to be exempt from open access.[39] Researchers often state that privacy issues prevent them from publishing data, occurring whenever data are related to human subjects (these issues prevail in disciplines working with person-related data). The publication of such data is only possible under certain conditions such as adequate anonymisation or informed consent. The UK Data Archive provides guidance for sharing confidential and sensitive data including access control, anonymisation and informed consent about data disclosure.[40]

In some disciplines such as life sciences or engineering, research data may have commercial potential or the research infrastructure may be provided by private partners. Institutions and private partners therefore may claim the protection of intellectual property rights and patents and some institutional policies may be an active obstacle to data publication. To obviate this obstacle to sharing, the Montreal Neurological Institute, associated with the McGill University, waived the pursuit of patents and mandated that all data produced as a result of research conducted in the institution must be published under open license.[41] The mandate applies to every researcher affiliated with the institute, including external collaboration partners.

Conclusion

Disciplinary data practices, norms and policies vary significantly and some disciplines publish data more readily than others. Given the disparities across disciplines to open research data, policies must pay attention to data practices, norms, infrastructure as well as to individual incentives to publish research data, particularly due to the obstacles that prevail in some disciplines and within particular research communities. PASTEUR4OA supports current policies and strategies adopted by funders, institutions, journals and data service providers and acknowledges that there is no one-size-fits-all solution to open research data. Some disciplines are currently better-equipped with policies and strategies that are well-integrated and aligned with specific exigencies. PASTEUR4OA therefore suggests that similar efforts are made in disciplines lacking such policies and advocates for more concerted efforts between the actors involved.

---

37 For Attribution - Developing Data Attribution and Citation Practices and Standards: Summary of an International Workshop: **http://www.nap.edu/catalog.php?record_id=13564**
38 Perspectives on Open Science and scientific data sharing: an interdisciplinary workshop: **http://www.ncbi.nlm.nih.gov/pubmed/25020017**
39 International Polar Year 2007-2008 Data Policy: **http://www.api-ipy.gc.ca/pg_IPYAPI_055-eng.html**
40 UK Data Archive: Consent and Ethics: **http://www.data-archive.ac.uk/create-manage/consent-ethics**
41 **http://www.sciencemag.org/news/2016/01/montreal-institute-going-open-accel-erate-science**

Further Information

Governmental Statements on open research data

- European Commission recommendation on access to and preservation of scientific information http://ec.europa.eu/research/science-society/document_library/pdf_06/recommendation-access-and-preservationscientific-information_en.pdf
- Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020: https://www.openaire.eu/guidelines-on-open-access-to-scientific-publications-and-research-data-in-horizon-2020
- G8 Science Ministers Statement London UK, 12 June 2013: https://www.gov.uk/government/news/g8-science-ministers-statement
- Increasing Access to the Results of Federally Funded Scientific Research, Office of Science and Technology Policy: http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf
- OECD Ministers of Science and Technology: Access to Research Data From Public Funding: http://bit.ly/1lGubJQ

Discipline-specific research data policies of funders and institutions

- Comprehensive Brief on Research Data Management Policies: http://www.science.gc.ca/default.asp?lang=En&n=1E116DB8-1

Exemplary discipline-specific data policies by funders

- RCUK Common Principles on Data Policy: http://www.rcuk.ac.uk/research/datapolicy
- The EAGDA funders on incentivising data sharing: Establishing incentives and changing cultures to support data access: Response from the EAGDA funders: http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/EAGDA/WTP056496.htm
- BBSRC Data Sharing Policy: http://www.bbsrc.ac.uk/about/policies/policy-foi/policy/data-sharing-policy/;
- ESRC Research Data Policy: http://www.esrc.ac.uk/funding/guidance-for-grant-holders/research-data-policy/
- DFG's guidelines on the "Handling of Research Data": http://www.dfg.de/en/research_funding/proposal_review_decision/applicants/submitting_proposal/research_data/index.html

Exemplary data policies by universities

- Heidelberg University: https://www.uni-heidelberg.de/universitaet/profil/researchdata/
- Oxford University: http://researchdata.ox.ac.uk
- University of Cambridge: http://www.data.cam.ac.uk/research-data-policies

Research on discipline-specific data publication

- Data Sharing by Scientists: Practices and Perceptions: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0021101
- Data sharing, small science and institutional repositories: http://rsta.royalsocietypublishing.org/content/368/1926/4023
- Digital Curation Centre (2010): Data Dimensions: Disciplinary Differences in Research Data Sharing, Reuse and Long term Viability: http://www.dcc.ac.uk/news/data-dimensions-disciplinary-differences-research-data-sharing-reuse-and-long-term-viabilityhttps://www.uni-heidelberg.de/universitaet/profil/researchdata/
- Knowledge Exchange Group Report (2014): Sowing the seed: incentives and motivations for sharing research data: researcher's perspectives: http://knowledge-exchange.info/event/sowing-the-seed
- Open Data in Economics: The Basis of Reproducible Research: http://blogs.lse.ac.uk/impactofsocialsciences/2013/10/23/open-data-in-economics-the-basis-of-reproducible-research/
- RIN Report on data practices across disciplines (2008): To Share or not to Share: Publication and Quality Assurance of Research Data Outputs: http://www.rin.ac.uk/our-work/data-management-and-curation/share-or-not-share-research-data-outputs
- RIN Report on data practices in life sciences (2009): Patterns of information use and exchange: case studies of researchers in the life sciences: http://www.rin.ac.uk/our-work/using-and-accessing-information-resources/patterns-information-use-and-exchange-case-studie
- RIN Report on data practices in humanities (2011): Reinventing Research? Information Practices in the humanities: http://www.rin.ac.uk/our-work/using-and-accessing-information-resources/information-use-case-studies-humanities
- Researcher Perspectives on Publication and Peer Review of Data: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0117619;

- The conundrum of sharing research data: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1869155
- What Drives Academic Data Sharing?: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0118053;
- The Problem of Data: http://www.clir.org/pubs/reports/pub154
- When Data Sharing Gets Close to 100%: What Human Paleogenetics Can Teach the Open Science Movement: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0121409
- Wiley survey on the extend of data sharing across disciplines: Researcher data sharing insights: http://exchanges.wiley.com/blog/2014/11/03/how-and-why-researchers-share-data-and-why-they-dont/

Discipline-specific repositories and journals

- The Registry of Research Data Repositories provides a comprehensive overview of repositories with open licenses: http://www.re3data.org/
- Different data access models adopted by the UK Data Service: https://www.ukdataservice.ac.uk/get-data/data-access-policy
- American Economic Review: https://www.aeaweb.org/aer/data.php
- Ubiquity Press: http://www.ubiquitypress.com/
- Public Library of Science: https://www.plos.org/

Study on licensing research data

- Safe to be open - Study on the protection of research data and recommendations for access and usage: https://www.openaire.eu/search/publication?articleId=od_____323::d052becf1674cfad70faca45f07fa5fa

Exemplary organisations/projects providing discipline-specific support

- B2SHARE: https://b2share.eudat.eu/?ln=en
- DataOne: https://www.dataone.org
- DataOne for librarians: https://www.dataone.org/for-librarians
- Data Documentation Initiative: http://www.ddialliance.org/
- Disciplinary research data management training by the Digital Curation Centre: http://www.dcc.ac.uk/training/train-trainer/disciplinary-rdm-training/disciplinary-rdm-training
- Making Data Count: Developing a Data Metrics Pilot: http://blogs.plos.org/tech/making-data-count-plos-cdl-and-dataone-join-forces-to-build-incentives-for-data-sharing/
- Thomson Reuters' Data Citation Index: http://wokinfo.com//products_tools/multidisciplinary/dci/?utm_source=false&utm_medium=false&utm_campaign=false