

# Open Research Data: From Vision to Practice

Heinz Pampel and Sünje Dallmeier-Tiessen

*To make progress in science, we need to be open and share.*  
—Neelie Kroes (2012)

**Abstract** “To make progress in science, we need to be open and share.” This quote from Neelie Kroes (2012), vice president of the European Commission describes the growing public demand for an Open Science. Part of Open Science is, next to Open Access to peer-reviewed publications, the Open Access to research data, the basis of scholarly knowledge. The opportunities and challenges of Data Sharing are discussed widely in the scholarly sector. The cultures of Data Sharing differ within the scholarly disciplines. Well advanced are for example disciplines like biomedicine and earth sciences. Today, more and more funding agencies require a proper Research Data Management and the possibility of data re-use. Many researchers often see the potential of Data Sharing, but they act cautiously. This situation shows a clear ambivalence between the demand for Data Sharing and the current practice of Data Sharing. Starting from a baseline study on current discussions, practices and developments the article describe the challenges of Open Research Data. The authors briefly discuss the barriers and drivers to Data Sharing. Furthermore, the article analyses strategies and approaches to promote and implement Data Sharing. This comprises an analysis of the current landscape of data repositories, enhanced publications and data papers. In this context the authors also shed light on incentive mechanisms, data citation practises and the interaction between data repositories and journals. In the conclusions the authors outline requirements of a future Data Sharing culture.

---

H. Pampel (✉)

GFZ German Research Centre for Geosciences, Potsdam, Germany  
e-mail: pampel@gfzpotsdam.de

S. Dallmeier-Tiessen

Scientific Information Service, CERN, Geneva, Switzerland

## The Vision of Open Research Data

Digitization has opened up new possibilities for scientists in their handling of information and knowledge. The potential of networked research was recorded in the “Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities” (2003). This declaration was signed by leading scientific organizations and is regarded as the central reference for the demands of access and sharing of scientific results in the digital age. Previous definitions of Open Access were related to free access to peer-reviewed literature,<sup>1</sup> whereas the “Berlin Declaration” considers this in a wider sense. Not only articles, but also “raw data and metadata, source materials, digital representations of pictorial and graphical materials and scholarly multimedia material” should be openly accessible and usable.

This demand is also evident on the political level. An example is the statement made in a publication of the Organisation for Economic Co-operation and Development (OECD) 2007, entitled “Principles and Guidelines for Access to Research Data from Public Funding”: “Sharing and Open Access to publicly funded research data not only helps to maximise the research potential of new digital technologies and networks, but provides greater returns from the public investment in research” (OECD 2007). The European Commission also strives for Open Access to research data. In the “Commission Recommendation on Access to and Preservation of Scientific Information” which was published in 2012, European member states are requested to ensure that “research data that result from publicly funded research become publicly accessible, usable and re-usable through digital e-infrastructures” (European Commission 2012a).

The discussion on the realisation of this aim is present in the scientific communities (Nature 2002, 2005, 2009a, b; Science 2011). The term Open Research Data can be applied on a cross-disciplinary layer. It covers the heterogeneity of the data with its diverse characteristics, forms and formats in the scientific disciplines. Further to this, the term Open Research Data is distinct to Open Data, which is mainly used in the context of Open Government initiatives and neglects the special requirements of science.

The two central arguments for Open Access to research data are a) the possibility to re-use data in a new connection and b) the verifiability it guarantees for ensuring good scientific practice. The OECD (2007) added a further argument: “Sharing and Open Access to publicly funded research data not only helps to maximise the research potential of new digital technologies and networks, but provides greater returns from the public investment in research.”

The vision of the High Level Expert Group on Scientific Data for the year 2030 is that, scientists in their role as data user, “are able to find, access and process the

---

<sup>1</sup> Compare: Budapest Open Access Initiative, 2002: <http://www.opensocietyfoundations.org/openaccess/read> & Bethesda Statement on Open Access Publishing, 2003: <http://www.earlham.edu/~peters/fos/bethesda.htm>

data they need”, and in their role as data producer “prefer to deposit their data with confidence in reliable repositories” (High Level Expert Group on Scientific Data 2010).

The demand for Open Research Data effects individual researchers and their data handling. In a report of The Royal Society (2012) entitled “Science as an open enterprise” which is worth reading, the recommendation is given that: “[s]cientists should communicate the data they collect and the models they create, to allow free and Open Access, and in ways that are intelligible, assessable and usable for other specialists in the same or linked fields wherever they are in the world. Where data justify it, scientists should make them available in an appropriate data repository. Where possible, communication with a wider public audience should be made a priority, and particularly so in areas where openness is in the public interest.” This recommendation makes it clear that diverse basic conditions must be created before Data Sharing can become a standard in scientific practice. Access and usage conditions must be defined. Murray-Rust et al.<sup>2</sup> for example demands the free accessibility in the public domain in their “Panton Principles”: “By open data in science we mean that it is freely available on the public Internet permitting any user to download, copy, analyse, re-process, pass them to software or use them for any other purpose without financial, legal, or technical barriers other than those inseparable from gaining access to the Internet itself.” The majority of the disciplines are still far away from the implementation of these “[p]rinciples for open data in science”, however. In addition, there are many cases in the life sciences and social science disciplines in which, because of data protection and personal rights, Data Sharing is not possible, or only possible under narrowly defined conditions.

## The Status of Data Sharing Today

In a consultation carried out in 2012, the European Commission determined that there were massive access barriers to research data. Of the 1,140 of those questioned, 87 % contradicted the statement that “there is no access problem to research data in Europe” (European Commission 2012b). In a revealing study made by Tenopir et al. (2011), 67 % of the more than 1,300 researchers pointed to a “lack of access to data generated by other researchers or institutions” that is a hindrance to advances in science. Scientists frequently see the potential offered by Open Research Data, but most are reticent with regard to the open accessibility of their own data. Tenopir et al. found, for example, that “only about a third (36 %) of the respondents agreed that others can access their data easily”. This is not in accordance with the researchers attitudes, as three-quarters state that they “share their data with others”. The study also sheds light on different disciplinary

---

<sup>2</sup> Panton Principles: <http://pantonprinciples.org/>

practices: Whereas 90 % of scientists working in atmospheric science were willing to sharing their data, only 58 % of the questioned social sciences scientists were ready to do this. The authors conclude: “there is a willingness to share data, but it is difficult to achieve or is done only on request.”

Further insights in disciplinary practices are presented by the studies Wicherts et al. (2006) in psychology and Savage and Vickers (2009) in medicine, for example. Wicherts et al. approached authors of 141 articles published in 2004 in journals of the American Psychological Association (APA) and requested access to data that was the basis of the articles. Within the next six months, they only received positive replies from one third of the authors, 73 % of the authors were not prepared to share their data. Savage & Vickers came to a similar result. They asked authors of ten articles that were published in the PLoS Medicine or PLoS Clinical Trials journals to allow them access to the underlying data of the articles. Despite the clear demands for Open Access to data in den editorial policies of each of the Open Access journals, only one author permitted access to the requested data. A further insight in the status of Data Sharing is presented by the analysis of Campbell et al. (2002) in genetics. In this study that involved about 1,800 life science scientists, they identified two central factors that hinder Data Sharing: “Lack of resources and issues of scientific priority play an important role in scientists’ decisions to withhold data, materials, and information from other academic geneticists.”

Alongside these very reserved attitudes, however, there are numerous examples which underline that open exchange of research data can successfully be realized. The “Bermuda Principles” that were adopted in human genetics in the framework of the Human Genome Project in 1996, for example. These principles require that “[a]ll human genomic sequence data generated by centers funded for large-scale human sequencing should be freely available and in the public domain to encourage research and development and to maximize the benefit to society” (Smith and Carrano 1996). Over time the pre-publication of data comes off as common practice, i.e. gene sequences are made openly accessible prior to the description of them in a peer reviewed article.<sup>3</sup> Alongside Data Sharing in large scientific projects, in which data is made openly available in trustworthy research data repositories, there are also examples of spontaneous Data Sharing. Research on a disease-causing strain of the *Escherichia coli* (O104:H4) bacteria is such a case. This caused more than 4,000<sup>4</sup> people to fall ill in Germany in 2011. The publication of sequence data under a Creative Commons licence and the use of the widely popular GitHub<sup>5</sup> as an exchange platform enabled scientists all over the

---

<sup>3</sup> See also the Fort Lauderdale Principles (Wellcome Trust 2003) and the Toronto Statement (Birney et al. 2009).

<sup>4</sup> See: [http://www.rki.de/DE/Content/Service/Presse/Pressemitteilungen/2011/11\\_2011.html](http://www.rki.de/DE/Content/Service/Presse/Pressemitteilungen/2011/11_2011.html) (Retrieved 20 August 2012).

<sup>5</sup> GitHub is a hosting service for the collaborative development of software. See: <https://github.com/ehec-outbreak-crowdsourced> (Retrieved 20 August 2012).

world to make a contribution to a rapid investigation of the bacterium (Kupferschmidt 2011; Turner 2011; Check Hayden 2012).

A further example of successful Data Sharing is the operation of the World Data System (WDS) of the International Council of Science (ICSU) which - even before the coming into being of the Internet - resulted from the International Geophysical Year (1957–1958). This network of disciplinary data centers ensures “full, open, timely, non-discriminatory and unrestricted access to metadata, data, products and services”.<sup>6</sup>

## Understanding the Barriers

So-called data policies have an increasing effect on scientists and how they handle research data.<sup>7</sup> Recommendations and mandatory requirements by funding agencies and scientific journals stand out here. They request the beneficiary of funds to ensure the preservation and accessibility of data created in the framework of a funded project or a publication. The National Institute of Health (NIH) was a pioneer in this respect. It anchored its “Data Sharing Policy” in 2003: Applicants for a grant upwards of 500,000 US dollar are requested to make statements on Data Sharing.<sup>8</sup> From 2011 on, the National Science Foundation (NSF) requires receivers of funds “to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants” (National Science Foundation 2011a). Measures for the implementation of this guideline must be specified in a “Data Management Plan” (National Science Foundation 2011b). This request is being increasingly taken up by scientific journals via editorial policies. Exemplary for these are the requirements of the Nature journals, in which “authors are required to make materials, data and associated protocols promptly available to readers without undue qualifications in material transfer agreements”. It is suggested that the data be made accessible “via public repositories”.<sup>9</sup>

It must be noted that implementation of the requirements formulated in the data policies will not run by itself (Pampel and Bertelmann 2011). To promote Data Sharing it is necessary to identify the barriers, which influence scientists with regard to the sharing of their own data. Surveys carried out by Kuipers and Van der Hoeven (2009) and Tenopir et al. (2011) allow the following barriers to be named:

---

<sup>6</sup> ICSU World Data System: [http://icsu-wds.org/images/files/WDS\\_Certification\\_Summary\\_11\\_June\\_2012\\_pdf](http://icsu-wds.org/images/files/WDS_Certification_Summary_11_June_2012_pdf)

<sup>7</sup> For details see: Pampel & Bertelmann (2011).

<sup>8</sup> National Institutes of Health: <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html>.

<sup>9</sup> Guide to Publication Policies of the Nature Journals: <http://www.nature.com/authors/gta.pdf>

“legal issues”, “misuse of data” and “incompatible data types” (Kuipers and Van der Hoeven 2009), as well as “insufficient time” and “lack of funding” (Tenopir et al. 2011). These barriers make it clear that a dedicated framework is required for the publication of research data. The conception and implementation of such a framework is being increasingly discussed under the Research Data Management term.<sup>10</sup> The aim is to develop organisational and technical measures to ensure a trustworthy infrastructure for permanent integrity and re-use of data. The centre of attention hereby is the operation of information infrastructures, such as research data repositories, in which research data can be permanently stored. To make re-use of the stored data possible, the Research Data Management framework must ensure that the data are described via metadata. Documentation of the instruments and methods used to obtain the data is necessary for reliable re-use of the data, for example. Such an enhanced documentation of data is often a time-consuming task that is competing with many other activities on the researchers priority list. Further to this, in many disciplines there are no standards in which the data can be described.

Recently, it can be observed that libraries, data centers and other institutions are increasingly collaborate and begin to build up information infrastructures to support scientists in the handling of their data and so also to promote Data Sharing (Pampel et al. 2010; Osswald and Strathmann 2012; Reilly 2012).

Van der Graaf and Waaijers (2011) have formulated four central fields of action for the realization of a “collaborative data infrastructure” which enables the “use, re-use and exploit research data to the maximum benefit of science and society”. Incentives must be given to stimulate Data Sharing (1); in addition, the education and training of scientists and service providers on and around the handling of data must be intensified (2). Further to these the authors point to the importance on the structuring and networking of research data infrastructures that serve for a permanent and reliable data storage (3) and point out the challenge of the long-term financing of these infrastructures (4).

## Overcoming the Barriers

A central barrier to of the pervasiveness of Data Sharing is the lack of incentives for the individual scientist to make his data openly accessible. In particular in projects, in which data management was not already discussed in the preparatory phase, the individual scientist has good reasons for not making his or her data openly accessible, as there are no incentive mechanisms for the sharing of research data in the competitive scientific system (Borgman 2010; Klump 2012).

---

<sup>10</sup> For details see: Büttner et al. (2011) and Pryor (2012).

The slogan “[c]redit where credit is overdue” (Nature Biotechnology 2009) clearly expresses that: Data Sharing will only be successful when it is worthwhile for a scientist to make his data openly accessible. Against this background, a growing number of publication strategies are appearing with a view to the implementation of Data Sharing on the basis of the established scientific reputation system. Three of these strategies are as follows<sup>11</sup>:

1. The publication of research data as an independent information object in a research data repository.
2. The publication of research data as a textual documentation in the form of a so-called data paper.
3. The publication of research data as enrichment of an article, a so-called “enriched publication”.

Whereas the first named practice has long been established in the life sciences with the use of data repositories such as GenBank (Benson et al. 2012)<sup>12</sup>, the second named data paper strategy has been gaining more and more attention recently. Chavan and Penev (2011) define this publication type as follows: “a journal publication whose primary purpose is to describe data, rather than to report a research investigation. As such, it contains facts about data, not hypotheses and arguments in support of those hypotheses based on data, as found in a conventional research article.” Experience with data papers has been made, among others, in the geosciences<sup>13</sup> and ecology.<sup>14</sup> The use of this model has recently been widened to include so-called data journals. The pioneer of this development is the Open Access journal Earth System Science Data (ESSD). It has published descriptions of geosciences data sets since 2008. The data sets themselves are published on a “reliable repository” (Pfeiffenberger and Carlson 2011). The data sets and descriptive publications described are permanently persistently addressed by means of a digital object identifier (DOI) which also facilitate data citation. Thanks to this procedure that was developed within the Publication and Citation of Scientific Primary Data (STD-DOI) project (Klump et al. 2006) and expanded by DataCite (Brase and Farquhar 2011), it is possible to link publications and the underlying data. This procedure also supports the visibility of the data. Some publishing houses, for example, have therefore already integrated freely accessible research data in their platforms (Reilly et al. 2011). A number of data journals

---

<sup>11</sup> The following categorization is based on Dallmeier-Tiessen (2011).

<sup>12</sup> For the GenBank history see: Cravedi (2008).

<sup>13</sup> The AGU Journals have published data papers for many years. See: [http://www.agu.org/pubs/authors/policies/data\\_policy.shtml](http://www.agu.org/pubs/authors/policies/data_policy.shtml) (Retrieved 20 August 2012).

<sup>14</sup> See the Data Papers of the Journals Ecological Archives of the Ecological Society of America (ESA): [http://esapubs.org/archive/archive\\_D.htm](http://esapubs.org/archive/archive_D.htm) (Retrieved 20 August 2012).

have been brought into being in the meantime.<sup>15</sup> It must be noted here that the establishment of data journals is only feasible when data, metadata and the corresponding text publication are freely accessible, as only then can a barrier free re-use of the data be possible.

The linking of articles and data is also addressed in the third named enriched publication strategy (Woutersen-Windhouver et al. 2009). The aim is to build and sustain a technical environment to relate all relevant information objects around an article so that a knowledge space is created, in which the research data that are the basis of the article can be made freely accessible.<sup>16</sup>

The implementation of the three strategies requires trustworthy repositories on which the data can be made permanently accessible. A differentiation must be made here between institutional, disciplinary, multi-disciplinary and project-specific infrastructure (Pampel et al. 2012). Prominent examples of disciplinary research data repositories are GenBank in genetics and PANGAEA in geosciences and Dryad in biodiversity research.<sup>17</sup> A look at the access conditions of repositories highlights some differences: GenBank states that there are “no restrictions on the use or distribution of the GenBank data, PANGAEA licences the data under the “Creative Commons Licence Attribution” and Dryad makes the data accessible under the “Creative Commons License CC0” in the public domain.

A number of studies have been published that show the impact of Data Sharing on citation rates. Articles for which the underlying data is shared are more frequently cited than articles for which this is not the case. This is substantiated in studies from genetics (Piwowar et al. 2007; Botstein 2010), astronomy (Henneken and Accomazzi 2011; Dorch 2012) and paleoceanography Sears (2011). Such results need to be considered when discussing the lack of incentives for Data Sharing. The same holds true for data citation and data papers, which could contribute to the researchers publication profile and thus current research assessments and incentive systems.

---

<sup>15</sup> Examples: Atomic Data and Nuclear Data Tables (Elsevier); Biodiversity Data Journal (Pensoft Publishers); Dataset Papers in Biology (Hindawi Publishing Corporation); Dataset Papers in Chemistry (Hindawi Publishing Corporation); Dataset Papers in Ecology (Hindawi Publishing Corporation); Dataset Papers in Geosciences (Hindawi Publishing Corporation); Dataset Papers in Materials Science (Hindawi Publishing Corporation); Dataset Papers in Medicine (Hindawi Publishing Corporation); Dataset Papers in Nanotechnology (Hindawi Publishing Corporation); Dataset Papers in Neuroscience (Hindawi Publishing Corporation); Dataset Papers in Pharmacology (Hindawi Publishing Corporation); Dataset Papers in Physics (Hindawi Publishing Corporation); Earth System Science Data—ESSD (Copernicus Publications); Geoscience Data Journal (Wiley); GigaScience (BioMed Central); Nuclear Data Sheets (Elsevier); Open Archaeology Data (Ubiquity Press); Open Network Biology (BioMed Central). Please note that the majority of the journals are still developing and a narrow definition of the type of publication is difficult because of this early development stage.

<sup>16</sup> Potential offered by this strategy under use of Linked Open Data.

<sup>17</sup> An overview of existing data repositories is offered by re3data.org (<http://re3data.org>).



## Translating Vision into Practice

The developments in recent years have shown that numerous initiatives have emerged in Data Sharing. The hesitation among researchers in many disciplines is met by new strategies that work on barriers such as the lack of incentives. A professionalization of the Research Data Management, which supports scientists in the sharing of their data, is necessary to ensure the permanent accessibility, however. In this context, priority must be given to the structuring and networking of the research data repositories and their long-term financing.

A more detailed analysis for the identification and overcoming of barriers to Data Sharing has been created in the framework of the EU-project Opportunities for Data Exchange (ODE).<sup>18</sup> This project takes the various players involved in scholarly communication and data management (policy-makers, funders, researchers, research and education organisations, data centres and infrastructure service providers and publishers) into consideration, names variables that have an effect on the sharing and points out strategies for overcoming barriers to Open Access (Dallmeier-Tiessen et al. 2012). Many of the strategies that are outlined show that, to counter the diverse challenges, close cooperation is necessary between the players named above. As an example, the successful implementation of data policies of supporting organizations requires a Research Data Management and infrastructures that support scientists and create a regulatory framework. All of these measures will only lead to success, however, when scholarly societies and other disciplinary players who support the anchoring in the disciplinary communities take part. All players in the scientific process are therefore requested to make their contribution to Open Access of research data.

The publication strategies outlined show that there really is a possibility for the anchoring of Data Sharing in the scientific reputation system. Further innovation is desirable, though. The implementation of the increasing demand for Open Science from society<sup>19</sup> and academic policy (Kroes 2012), as is assumed, for example, by the federation of national academies ALLEA - ALL European Academies (2012), needs a culture of sharing. The establishment of this culture is a far reaching challenge. It appears that implementation of it can only then be successful when changes are made in the scientific reputation system. Scientific performances should in the future be valued with a “sharing factor” that not only judges the citation frequency in the scientific community, but also rates the implementation of sharing of information and knowledge for the good of society.

The demand for openness in science is loud and clear. All players in the scientific area should direct their practices to this demand. The publication strategies for research data have up to now been important approaches towards Open Science. The following citation from the “Berlin Declaration” (2003) makes it

---

<sup>18</sup> See: <http://ode-project.eu>

<sup>19</sup> See here, for example, the Vision of the Open Knowledge Foundation (OKF): <http://okfn.org/about/vision/> (Retrieved 20 August 2012).

clear, that further steps are necessary for the realization of Open Science: “Our mission of disseminating knowledge is only half complete if the information is not made widely and readily available to society.”

**Open Access** This chapter is distributed under the terms of the Creative Commons Attribution Noncommercial License, which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- All European Academies. (2012). Open Science for the 21st century, Rome, Italy. Available at: <http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/allea-declaration-1.pdf>.
- Benson, D.A., et al. (2012). GenBank. *Nucleic Acids Research*, 40(D1), D48–D53. doi:10.1093/nar/gkr1202.
- Birney, H., et al. (2009). Prepublication data sharing. *Nature*, 461(7261), 168–170. doi:10.1038/461168a.
- Borgman, C. L. (2010). Research data: Who will share what, with whom, when, and why? *SSRN Electronic Journal*. doi:10.2139/ssrn.1714427.
- Botstein, D. (2010). It’s the data! *Molecular Biology of the Cell*, 21(1), 4–6. doi:10.1091/mbc.E09-07-0575.
- Brase, J. & Farquhar, A. (2011). Access to research data. *D-Lib Magazine*, 17(1/2). doi:10.1045/january2011-brase.
- Büttner, S., Hobohm, H.-C., & Müller, L. (eds.) (2011). *Handbuch Forschungsdatenmanagement*, Bad Honnef: Bock + Herchen.
- Campbell, E.G., et al. (2002). Data withholding in academic genetics: Evidence from a national survey. *JAMA: The Journal of the American Medical Association*, 287(4), 473–480.
- Chavan, V., & Penev, L. (2011). The data paper: a mechanism to incentivize data publishing in biodiversity science. *BMC Bioinformatics*, 12(Suppl 15), S2. doi:10.1186/1471-2105-12-S15-S2.
- Check Hayden, E. (2012). Open-data project aims to ease the way for genomic research. *Nature*. doi:10.1038/nature.2012.10507.
- Cravedi, K. (2008). GenBank celebrates 25 years of service with two day conference. Leading scientists will discuss the DNA database at April 7-8 Meeting. National Institutes of Health. Available at: <http://www.nih.gov/news/health/apr2008/nlm-03.htm>.
- Dallmeier-Tiessen, S., et al. (2012). Compilation of results on drivers and barriers and new opportunities. Available at: <http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2012/08/ODE-CompilationResultsDriversBarriersNewOpportunities1.pdf>.
- Dorch, B. (2012). On the citation advantage of linking to data: Astrophysics. Available at: [http://hprints.org/docs/00/71/47/34/PDF/Dorch\\_2012a.pdf](http://hprints.org/docs/00/71/47/34/PDF/Dorch_2012a.pdf).
- European Commission. (2012a). Commission recommendation on access to and preservation of scientific information. C(2012) 4890 final, Available at: [http://ec.europa.eu/research/science-society/document\\_library/pdf\\_06/recommendation-access-and-preservation-scientific-information\\_en.pdf](http://ec.europa.eu/research/science-society/document_library/pdf_06/recommendation-access-and-preservation-scientific-information_en.pdf).
- European Commission. (2012b). Online survey on scientific information in the digital age, Luxembourg: Publications Office of the European Union. Available at: [http://ec.europa.eu/research/science-society/document\\_library/pdf\\_06/survey-on-scientific-information-digital-age\\_en.pdf](http://ec.europa.eu/research/science-society/document_library/pdf_06/survey-on-scientific-information-digital-age_en.pdf).
- Henneken, E.A. & Accomazzi, A. (2011). Linking to data—effect on citation rates in astronomy. *Digital Libraries; Instrumentation and Methods for Astrophysics*. Available at: <http://arxiv.org/abs/1111.3618v1>.

- High Level Expert Group on Scientific Data. (2010). Riding the wave. How Europe can gain from the rising tide of scientific data. Available at: <http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf>.
- Klump, J. (2012). Offener Zugang zu Forschungsdaten. In U. Herb (ed), *Open Initiatives: Offenheit in der digitalen Welt und Wissenschaft*. Saarbrücken: Universaar, pp. 45–53. Available at: <http://hdl.handle.net/10760/17213>.
- Klump, J., et al. (2006). Data publication in the Open Access initiative. *Data Science Journal*, 5, 79–83. doi:10.2481/dsj.5.79.
- Kroes, N. (2012). Opening Science Through e-infrastructures. Available at: [europa.eu/rapid/pressReleasesAction.do?reference=SPEECH/12/258](http://europa.eu/rapid/pressReleasesAction.do?reference=SPEECH/12/258).
- Kuipers, T. & van der Hoeven, J. (2009). Insight into digital preservation of research output in Europe. Available at: [http://www.parse-insight.eu/downloads/PARSE-Insight\\_D3-4\\_Survey\\_Report\\_final\\_hq.pdf](http://www.parse-insight.eu/downloads/PARSE-Insight_D3-4_Survey_Report_final_hq.pdf).
- Kupferschmidt, K. (2011). Scientists rush to study genome of lethal *E. coli*. *Science*, 332(6035), 1249–1250. doi:10.1126/science.332.6035.1249.
- National Science Foundation. (2011a). Award and administration guide. Chapter VI other post award requirements and considerations, Available at: [http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/aag\\_6.jsp#VID4](http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/aag_6.jsp#VID4).
- National Science Foundation. (2011b). Proposal and award policies and procedures guide. Grant proposal guide. Chapter II proposal preparation instructions. Available at: [http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/gpg\\_2.jsp#dmp](http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/gpg_2.jsp#dmp).
- Nature. (2002). How to encourage the right behaviour. *Nature*, 416(6876), 1. doi:10.1038/416001b.
- Nature. (2005). Let data speak to data. *Nature*, 438(7068), 531. doi:10.1038/438531a.
- Nature. (2009a). Data for the masses. *Nature*, 457(7226), 129. doi:10.1038/457129a.
- Nature. (2009b). Data's shameful neglect. *Nature*, 461(7261), 145. doi:10.1038/461145a.
- Nature Biotechnology. (2009). Credit where credit is overdue. *Nature Biotechnology*, 27(7), 579. doi:10.1038/nbt0709-579.
- Organisation for Economic Co-operation and Development (OECD). (2007). OECD principles and guidelines for access to research data from public funding. Retrieved from <http://www.oecd.org/dataoecd/9/61/38500813.pdf>.
- Osswald, A., & Strathmann, S. (2012). The role of libraries in curation and preservation of research data in Germany: Findings of a survey. In *78th IFLA General Conference and Assembly*. Available at: <http://conference.ifla.org/sites/default/files/files/papers/wlic2012/116-osswald-en.pdf>.
- Pampel, H., & Bertelmann, R. (2011). "Data policies" im Spannungsfeld zwischen Empfehlung und Verpflichtung. In S. Büttner (ed.), *Handbuch Forschungsdatenmanagement*. Bad Honnef: Bock + Herchen, pp. 49–61.
- Pampel, H., Bertelmann, R., & Hobohm, H.-C. (2010). "Data librarianship" Rollen, Aufgaben, Kompetenzen. In U. Hohoff & C. Schmiedeknecht (eds.), *Ein neuer Blick auf Bibliotheken*. Hildesheim: Olms, pp. 159–176. Available at: <http://econpapers.repec.org/paper/rswrswwps/rswrswwps144.htm>.
- Pampel, H., Goebelbecker, H.-J., & Vierkant, P. (2012). *re3data.org: Aufbau eines Verzeichnisses von ForschungsdatenRepositoryen. Ein Werkstattbericht (Forthcoming)*. Jülich: Forschungszentrum Jülich.
- Pfeiffenberger, H., & Carlson, D. (2011). "Earth system science data" (ESSD) a peer reviewed journal for publication of data. *D-Lib Magazine*, 17(1/2). doi:10.1045/january2011-pfeiffenberger.
- Piwowar, H.A., Day, R.S., & Fridsma, D.B. (2007). Sharing detailed research data is associated with increased citation rate J. Ioannidis, ed. *PLoS ONE*, 2(3), 308. doi:10.1371/journal.pone.0000308.
- Pryor, G. (2012). *Managing research data*. London: Facet Publishing.

- Reilly, S., et al. (2011). *report on integration of 15 data and publications*, Available at: [http://www.alliancepermanentaccess.org/wpcontent/uploads/downloads/2011/11/ODE-ReportOnIntegrationOfDataAndPublications-1\\_1.pdf](http://www.alliancepermanentaccess.org/wpcontent/uploads/downloads/2011/11/ODE-ReportOnIntegrationOfDataAndPublications-1_1.pdf).
- Reilly, S. (2012). The role of libraries in supporting data exchange. In *78th IFLA General Conference and Assembly*. Available at: <http://conference.ifla.org/sites/default/files/files/papers/wlic2012/116-reilly-en.pdf>.
- Savage, C.J. & Vickers, A.J. (2009). Empirical study of data sharing by authors publishing in PLoS Journals. C. Mavergames, ed. *PLoS ONE*, 4(9), e7078. doi:10.1371/journal.pone.0007078.
- Science. (2011). Challenges and opportunities. *Science*, 331(6018), 692–693. doi:10.1126/science.331.6018.692.
- Sears, J.R., (2011). Data sharing effect on article citation rate in paleoceanography. IN53B1628. In *AGU Fall Meeting 2011*. Available at: <http://static.coreapps.net/agu2011/html/IN53B-1628.html>.
- Smith, D., & Carrano, A. (1996). International large-scale sequencing meeting. *Human Genome News*, 6(7). Available at: [http://www.ornl.gov/sci/techresources/Human\\_Genome/publicat/hgn/v7n6/19intern.shtml](http://www.ornl.gov/sci/techresources/Human_Genome/publicat/hgn/v7n6/19intern.shtml).
- Tenopir, C., et al. (2011). Data sharing by scientists: Practices and perceptions. C. Neylon, ed. *PLoS ONE*, 6(6), p.e21101. doi:10.1371/journal.pone.0021101.
- The Royal Society. (2012). Science as an open enterprise. The Royal Society Science Policy Centre report, Available at: [http://royalsociety.org/uploadedFiles/Royal\\_Society\\_Content/policy/projects/sape/2012-06-20-SAOE.pdf](http://royalsociety.org/uploadedFiles/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf).
- Turner, M. (2011). The German *E. coli* outbreak: 40 lives and hours of crowdsourced sequence analysis later. *Nature News Blog*. Available at: [http://blogs.nature.com/news/2011/06/the\\_german\\_e\\_coli\\_outbreak\\_40.html](http://blogs.nature.com/news/2011/06/the_german_e_coli_outbreak_40.html).
- van der Graaf, M. & Waaijers, L. (2011). A Surfboard for riding the wave. Towards a four country action programme on research data, Wellcome Trust. Available at: <http://www.knowledge-exchange.info/Admin/Public>.
- Wellcome Trust. (2003). Sharing data from largescale biological research projects. A system of tripartite responsibility. Available at: <http://www.genome.gov/Pages/Research/WellcomeReport0303.pdf>.
- Wicherts, J. M., et al. (2006). The poor availability of psychological research data for reanalysis. *The American psychologist*, 61(7), 726–728. doi:10.1037/0003-066X.61.7.726.
- Woutersen-Windhower, S., et al. (2009). *Enhanced publications: Linking publications and research data in digital repositories*. Amsterdam: Amsterdam University Press. Available at: <http://dare.uva.nl/aup/nl/record/316849>.