

Purposing your survey:

archives as a market regulator, or how can archives connect supply and demand? by Laurence Horton, Alexia Katsanidou¹

GESIS

Abstract

What do researchers need from archives? What do archives need from researchers? These questions cover two types of researchers that encounter data archives: those who create the data (data creators) and those who re-use it (data re-users). These groups have different needs and archives mediate between them.

The role of an archive for creators and re-users is to support them in producing quality data, metadata and documentation and to facilitate wide and multipurpose data dissemination. By supporting multipurpose reuse, to the fullest extent possible, archives help realize the value of public investment in academic research.

This paper discusses the optimization of research data management training and support for research data creators, and data dissemination and long-term preservation for social science data archives. It outlines the GESIS plan to create a research data management and archive training centre for the European research area, to cater to both data supply and data demand.

The training centre will look to ensure excellence in the creation and long-term preservation of reusable data in the European Research area, contribute to promoting and to the adoption of standards in research data management, and promote data availability and reuse. Finally, the centre will provide and coordinate training on technologies and tools used by data professionals.

Keywords: Archives, research data management, incentives, sharing, training.

Introduction²

Social science data archives connect two primary audiences. One is data creators—those who bring social science data into being. In this category, we place principal investigators of studies as well as researchers

who work in data collection procedures. The other audience is data re-users. Here we mean researchers who either use data they themselves created some time ago or use data created by others to examine social phenomena.

The ligaments connecting these audiences are data archives: organizations that facilitate data ingest and dissemination. By accepting data into their catalogue for preservation and reuse, then furnishing the research community with that data, the archives establish a connection between the two audiences. However, it is a dynamic relationship fashioned by two forces: a movement towards data sharing for reuse and a set of resistances to data reuse.

In this paper, we discuss these forces and we highlight actions to promote data sharing and reuse. The basis of our perspective is a supply and demand model of data archives and thus the basis of our proposals are for both audiences. We focus on attempts to introduce practical policy suggestions to facilitate an easier relationship between creators, archives, and re-users primarily within the CESSDA-ERIC consortium of European social science data archives.

The Data Sharing Movement

The contemporary movement towards data sharing for reuse is a trend enabled and assisted by technological innovation. The means by which one can share data and collaborate on research have become cheaper and easier to utilize. Negating the barriers towards reuse and collaboration posed by time, distance, cost, and logistics are developments in instantaneous means of communication, large capacity data transfer, cheaper digital storage costs, and the power of data analysis software packages. Today we can do more research with more data in less time and at less cost. Indeed the range, scope and potential applications of data created, available, and analyzed can reach such a size that it may even

challenge the primacy of the experimental hypothesis approach in doing social research (Anderson, 2008).

In recognition of these phenomena, the European Commission commissioned a report on how to best direct this changing data environment towards scientific and economic innovation. Its High Level Expert Group on Scientific Data envisioned

... a scientific e-infrastructure that supports seamless access, use, re-use, and trust of data. In a sense [...] the data themselves become the infrastructure – a valuable asset, on which science, technology, the economy and society can advance (European Union, 2010 p.4)

The belief that technology is changing patterns of research and publications has a normative basis in the argument that publicly funded data is a public good and that funders can maximize the value of research they support with a requirement that data be shared to the fullest extent possible. This argument is based on the position of the Organization for Economic Co-Operation and Development (OECD) that publically funded research data should as far as possible be openly available to the research community for re-analysis, repurposing, and long-term preservation (OECD, 2007).

The *Riding the Wave* report (European Union, 2010) echoed an expectation of transparency in data creation. An expectation that the methods of generating and manipulating data be clear so data is comprehensible to others outside of, and remains comprehensible to as time passes, the original data creators themselves. In addition, there is an acceptance as the norm in good scientific research that findings be based on data that is available (where legally and ethically possible) for independent verification, analysis, and reuse. This is a movement accelerated by a requirement of some academic journals that publication of articles is dependent on the authors' making available the underlying data if it is not already accessible. We find an example of this trend in Dryad. Dryad is an open data repository for articles published in the natural sciences and lists a number of journals as partners for which it either holds, or works with, to preserve and disseminate data (Dryad, 2011). An additional example is European Data Watch Explained (EDaWaX) (European Data Watch Extended, 2011) This German-based project examines the absence of incentives in economics for the replication of results and data reuse with the intention of creating a publication data archive. A similar project for political science, but with narrower focus is the GESIS Data Infrastructure team's Data Policy availability project. This project empirically investigates data policies of all top academic journals in political science, analyses their content and finally proposes policy guidelines.

In an era of tight pressures on public spending, the political attraction of these arguments is clear. The European Commission has committed itself to an open data policy that it estimates would provide an extra €40 billion a year to the EU economy. "Taxpayers have already paid for this information, the least we can do is give it back to those who want to use it in new ways..." stated Commission Vice President Neelie Kroes. "Your data is worth more if you give it away" (European Commission, 2011a) she added. However, the EC policy is tied to public sector data, not publicly funded academic research data which remains exempt (European Commission, 2011b). Yet this too can be, and is, considered a public investment to be shared thereby maximizing its value. We find examples of this belief in the emergence of policies that mandate data sharing be addressed as an aspect of proposals seeking public funding.

The United States National Institutes of Health (NIH) enforced a data sharing policy in 2003, with a requirement for funding applications to include a plan for data sharing (National Institutes of Health, 2003). The National Science Foundation (NSF) followed in early 2011 by adopting a similar requirement to produce a data management plan for sharing (National Science Foundation, 2011).

In the American environment it is often institutions that provide a preservation and dissemination service. Examples include University of California-San Diego (2010), University of Illinois at Urbana-Champaign (2005), Cornell University (2005), Massachusetts Institute of Technology (2005), and University of Rochester (2008). However, these approaches have been institution-specific rather than national infrastructure tools as NIH and NSF aside, the United States lacks the regional, national and supranational level funding regime of European countries such as the United Kingdom and Germany.

Similar developments have occurred in Europe. In May 2011, Research Councils UK—the strategic partnership agency of the United Kingdom's seven main research councils—published a set of common principles on data policy intended to provide an overarching framework for individual council policies on data reuse. The principals include an explicit statement that:

Publicly funded research data are a public good, produced in the public interest, which should be made openly available with as few restrictions as possible in a timely and responsible manner that does not harm intellectual property. (Research Councils UK, 2011)

UK councils may vary in the specifics of data, but this principal holds across the field. The Economic and Social Research Council (ESRC), Natural Environment Research Council (NERC) and the British Academy all mandate research data be offered to data centers. In the case of the ESRC (Economic and Social Data Service, 2011) and NERC (Natural Environment Research Council, 2011), through council funded data centers. Other UK funders expect or encourage data sharing but do not mandate places of deposit. The Engineering and Physical Sciences Research Council (EPSRC) has introduced a policy (from May 2015) mandating that institutions ensure well documented data is preserved and available for a minimum of 10 years from last request for access by a third party (Engineering and Physical Sciences Research Council, 2011). From an institutional perspective, University of Edinburgh, followed by the University of Hertfordshire (2011), became the first UK universities to adopt an institutional Research Data Management Policy. This included, in Edinburgh's case, a commitment that:

Research data management plans must ensure that research data are available for access and re-use where appropriate and under appropriate safeguards (University of Edinburgh, 2010).

In Germany, the main publically funded research organizations have adopted a set of principles for the handling of research data. This 2010 agreement does not take as strong a tone as its RCUK equivalent; however, it does support long-term preservation and the "principle" of open access to research data, as well as the development of subject-specific requirements, standards, and metadata to facilitate interdisciplinary research and supporting infrastructure (Alliance of German Science Organisations, 2010). These principals drew, in part, from an earlier set of proposals submitted by the German Research Council (DFG) that encourage researchers to take into account data management issues. Reinforcement of this invitation is by guidelines promoting data sharing for experts on review panels.

The DFG raise the issue of data management and demand secure preservation and visibility for those data publically funded and used for publications, but limit this demand to a ten-year period (Deutsche Forschungsgemeinschaft, 1998). Since then, greater effort has occurred to promote effective and consistent data management but not explicitly formulated in an official publication of the German Research Council.

Thus, the causes of a movement towards data preservation and sharing are clear: technology and financial benefit. Furthermore, the demand is there. Two of the largest data archives, the UK Data Archive (UKDA) as part of the Economic and Social Data Service (ESDS) and in the United States the Inter-University Consortium for Social and Political Research (ICPSR) (2011a), have both seen significant increases in orders for data they hold since offering online access to data (Economic and Social Data Service, various). A similar phenomenon is apparent in the GESIS Leibniz-Institute for Social Science's user statistics—specifically for Eurobarometer data, for which the number of datasets distributed has jumped between 2005 and 2009 (GESIS Leibniz Institute for the Social Sciences, 2010).

Resistances to Data Reuse

However, let us look at the supply side in the social sciences. Here there are still obstacles that prevent data sharing. Primary limitations are those placed by law and ethics. Neither data archives nor funding agencies believe in sharing all data with everyone, or even within the academic community. The policies and recommendations presented above recognize, as we do, that there has to be protection of intellectual property, professional credit, and critically—moral and ethical protection of research participants.

However, alongside these recognized limitations there are additional resistances to data sharing. Opposition remains to the idea of sharing research data. This phenomenon in the social sciences can draw on a range of arguments.

Low-level (researcher-level) ignorance as to why others would want to use their data. This was a reason cited by a small number of researchers interviewed for the UKDA's Data Management Planning for ESRC Centres and Programmes (UK Data Archive, 2010 pp. 17-21). It is not resistances to data reuse itself, but an inability to imagine that the type of data generated would be of interest to anyone else. We can overcome this problem through more interaction within the scientific community and open presentation of opportunities for data sharing.

Additional to the ignorance of researchers about potential reuse of their data, there are also epistemological concerns. These cover congruence, reflexivity, and context. Essentially, data creators holding this objection claim understanding and value of data can only exist in the specific context of their creation. They are concerned that their data, abstracted from the methodologies and ontologies adopted at the time of creation cannot adapt into a different research project. These problems of course need proper consideration particularly where the reflexive relationship between researcher and participant is critical to understanding the data, but given appropriate documentation, they should not prevent future reuse³

A clear problem is the lack of incentives to share data. As long as the main metric of career progression remains publications and citations of publications, data sharing will be a secondary concern. However, data creation requires the investment of a lot of scientific effort and expertise. Reusing an existing dataset builds on the scientific work of other researchers who should be not only acknowledged,

but also credited for their achievements. Widespread recognition and implementation of a system for acknowledgement of data citations as an indication of research quality and establishing them as equivalent to publication citations would remove a reservation against data sharing.

Data creators often have concerns as to the ethics of reuse concerning research participants. Specifically, a concern of compromised anonymity and confidentiality of participants emerges when disseminating data to other researchers. There are ways to anonymize data but some data are extremely sensitive and easily trackable. Thus, researchers can be reluctant to share on principle of protecting their participants' anonymity.

We propose that the character and structure of the current social science research environment determines attitudes to reuse. Outside of large-scale surveys, the concept of data reuse is not dispositional. There is still no established culture of archiving, sharing and reuse. The environment described above is situational. A strong situational determinist research environment should not only coerce researchers into creating reusable data, but also give them confidence to do so, thereby creating a researcher disposition towards creating reusable data. Using the colloquial metaphor that seems to be prevalent in research data management discussions, the current situation is mostly sticks and few carrots, and we need more carrots.

Promoting reuse: Cognition vs. Emotion

There is a case to be made, and has been made by funding councils and institutions, that data management and reuse be addressed as a mandatory requirement in any funding application. The reasoned argument for data management stands clear: it is fundamental to transparent, high quality sustainable data generation. Therefore, in psychological terms, data management for reuse is a "cold", cognitive task – an intellectually conscious, controlled process based on explicit learning (Kahneman, 2003). However, often the resistance to reuse draws not so much on logic, but sources that are more emotive. Drawing on movements within political psychology, what we feel should not happen is to dismiss emotive impulses.

We believe that emotions should be brought into the discussion between data creators and re-users. This is predicated on the belief that emotional responses are great motivators. Emotions can be harnessed to aid decisions, for example, the emotion to care. Ambition, incentives, professional acclimation can all be connected with data sharing and help researchers reach their decision to share. Researchers make an effort in collecting and working with data, and therefore they should develop an affective relationship with them. They are their intellectual creators and they should be given reason and tools to present them to the community in the same way they do with publications. If we can tie good research data management and data sharing into recognized career advancement, we can bring with it esteem of peers not just for the publications but the data underpinning publications. If we can instill professional pride in replication and peer scrutiny of data creation like the academic community has instilled in journal publications, then by sharing data researchers will be a more "important" with wider recognition than those who do not share because they will help advance the state of their discipline. Those who chose not to, however, will have another emotion to manage – fear: the fear of professional irrelevance (King 1995, p.445). For without emotions such as care, or fear, what incentive—and as we have suggested, incentives are currently lacking—is there to think of the consequences of actions? Through

this, we could hope to see a dispositional environment towards data sharing emerge.

Support for data sharing procedures is an important factor in facilitating sharing as lack of awareness can be a serious obstacle. While resources exist to support data creators in generating reusable data, they are often not discipline-specific. For example, the first versions of the Digital Curation Centre's (DCC) Data Management Planning Tool (Digital Curation Centre, 2011) or the Australian National Data Service's data management planning advice (Australian National Data Service, 2011) offer detailed but generic support. Although discipline-specific focuses are emerging, promoted in part through programs like JISC's Managing Research Data (Joint Information Systems Commission, 2009), as most are either generic tools or pure data management projects, these resources do not occupy the brokerage positions that data archives can assume.

The "brokerage" role of data archives – the supply and demand model

The responsibility of a broker is as a third-person facilitator to bring "sellers" and "buyers" together. We can therefore think of the brokerage role for an archive in terms of facilitating the "buying" (acquisition) and "selling" (dissemination) of data between data creator and data re-user. Archives know their "market" for data, and have established relations with creators "sellers" and re-users "buyers"; they are institutions that talk to both communities from acquisition to dissemination via ingest. Consequently, they become important regulators of this data market. They regulate the inflow and the quality of data on the supply side by encouraging data creators to share, leading the move to professional credit for sharing by making data citation possible and advising and supporting data creators on avoiding unnecessary obstacles to creating shareable data. However, they also regulate the output of data towards the demand side by disseminating them, increasing their visibility, and providing a service for responsible reuse of data.

To highlight four cases, the UKDA (2011), the ICPSR (2011b), in the Netherlands the DANS (Data Archiving and Networked Services, 2011a) and the IQDA (Irish Qualitative Data Archive, 2010) are national archives that have produced resources to aid data creators as well as providing data and dissemination support.

However, archives do not only regulate supply and demand. Through division of labor and specialization, they also add value to the data life cycle. Archives undertake tasks that enhance data quality and data survival in an uncertain technological world. Though not exhaustively, data archives provide long-term preservation of data with a strategy to ensure readability as file formats and technologies change. In addition, archives add value to data through structured metadata, catalogue records, and harmonization with comparative data collections. Archives develop networks for secure and easier access of data for reuse.

Nevertheless, to provide high quality data, archives must adopt modern technologies and standards, ensure cooperation between same-discipline archives across countries, and promote dialogue with archives operating in other disciplines. Through systematic interaction, archives can be the critical ligament that facilitates data sharing.

Incentives

The role of the archive is to build incentives for both audiences to adopt best practices when dealing with data. From the supply side, it is important to increase the cognitive and emotional incentives for data sharing. We have already stated the important enticement for

creators in making data available for reuse is their publications record, as their rewards and career advancements depend on that. The first step is then to make data citable. To do so, we need to provide the infrastructure and technology that allow the efficient referencing of data files. The most commonly used form of identifier is the Digital Object Identifier (DOI®) System (International DOI Foundation, 2011). These persistent identifiers are codes that connect a digital object such as a dataset, with accompanying metadata that includes author names, year of data collection and other important information of relevance. DOIs digitally identify journal articles, thus researchers are already familiar with their basic uses and functions. By having a DOI allocated to a dataset, the researcher can be sure that by using that specific DOI they refer to the same dataset. Therefore, referencing a dataset within the publication used to create it becomes effective. A reader of this publication can then identify the very same dataset with no alterations and replicate the analysis. This ensures research quality and the primary investigator is acknowledged.

GESIS is a data archive that has a project providing persistent identifiers for data files in its collection. One example, hosted by GESIS, is the *da|ra* project (GESIS Leibniz Institute for the Social Sciences, 2011) GESIS's registration agency for social science research data. The *da|ra* infrastructure lays foundations for permanent identification, storage, and localizing to create citable research data. Initiated in 2010 with a pilot phase, on entering 2012 the project is now in an upgrade phase. An expansion phase from 2013 to 2014 will centre on the development of useful services like user statistics, citation indexing, peer review possibilities for data, and registering of other data formats.

Another project of note is the effort by DANS (Data Archiving and Networked Services, 2011b) to produce a competitive alternative to the DOI. The Dutch archive is involved in the design and implementation of a persistent identifier (PI) infrastructure in cooperation with the infrastructure-oriented SURFfoundation [sic], and Koninklijke Bibliotheek (National Library of the Netherlands). This collaboration seeks to establish a mechanism called the National Resolver that would translate the PI into the current URL of the object.

In highlighting the projects and arguments we have presented thus far, it is our main goal to encourage researchers to take pride in their data creation activity, not just the outputs, and to invest time in making it reusable and archivable. We also aim to encourage researchers to value the work of other researchers who collect data, and to acknowledge this process as important and equal to other publication activities. To do that we focus on a new innovative data management training facility which we are involved in developing at GESIS: the Archiving and Data Management Training and Information Center (GESIS, 2012)

A concept for training

A new development that builds on the supply and demand model is the GESIS plan to create a research data management and archive training centre for the CESSDA-ERIC European area. This area is inclusive of data archives in twenty European nations (CESSDA, 2011)

The training centre will provide a central reference point for European researchers and archives, containing original resources and links to significant external resources, with the aim to ensure excellence in the creation and long-term preservation of reusable data, contribute to promoting the adoption of standards in research data management, and to advance data availability and reuse. The centre will also provide and coordinate training on technologies and tools used by data professionals.

By networking, and through surveys of demand for training needs, we are identifying themes and developing training concepts through potential collaborations with expert instructors. These concepts will be the basis on which courses are developed. The idea is to build resources around them using mixed and matched smaller thematic units depending on the needs of each specific course.

Our website will hold resources created by us, links to external resources, and will host information on the training center's consulting activities. The Virtual Centre of Competence will allow for consultation on best practice in research data management and archiving, including personal development and the promotion of skills training, provide information on our training activities, and offer structured teaching and self-learning materials.

Specifically, the centre will support data creators in implementing international standards of metadata and documentation. Information for data creators about the importance and uses of persistent identifiers and will be given, plus advice on ethics and consent, details on issues of data ownership, and an overview of archiving software systems.

The main support for data reuse is through the training of data archive staff to provide quality user support and to deal with increased volume of support requests. In addition, there will be information for archive professionals about new projects, new technologies, data discovery, and dissemination tools. Furthermore, the presentation of projects on data harmonization will enable archive professionals to add to the value of data for their users and create an online user community engaged in task of harmonization.

Finally, and perhaps most importantly, the training centre looks to support other archives, libraries and repositories in ensuring state of the art data-related functions and in keeping up with the constant development of new technologies. This feature is not only useful for institutions either in a formative stage or that are not specialized social science resources, it is essential for all institutions operating in the data world to keep up with innovation, establish clear workflows, and strive for internationally accepted standards.

The centre seeks to bring together the best examples and expert individuals to provide training. Training will not only have the traditional form of workshops. It will be an active form of community building and incentive development through all communication channels provided to us by the new technologies. The core of our training concept is to negate all the reasons outlined in this text that allow researchers to sit on their data without sharing, and this can only be done with systematic incentive building.

This training centre is only one way to augment the incentives of data sharing by bringing the subject closer to researchers' hearts. However, the other driving factors mentioned and analyzed in this paper have to be pushed forward in order to ensure the emotive connection of researchers to sharing data, and to establish it an integral part of the scientific contribution. In the world of data, the imperative to share is clear. We have enough sticks; it is time to cultivate the carrots.

References

Alliance of German Science Organisations (2010) "Principles for the Handling of Research Data" http://www.allianzinitiative.de/en/core_activities/research_data/principles/

- Anderson, C. (2008) "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete", *Wired Magazine*, 16(8) http://www.wired.com/science/discoveries/magazine/16-07/pb_theory
- Australian National Data Service (2011) "Data Management Planning" <http://ands.org.au/guides/data-management-planning-awareness.html>
- Bishop, L. (2009) "Ethical Sharing and Reuse of Qualitative Data", *Australian Journal of Social Issues*, 44(3), pp.261-267
- CESSDA: Council for European Social Science Data Archives (2011) "Member Organisations" <http://www.CESSDA.org/about/members/>
- Cornell University (2005) "Registry of Digital Collections" <http://rdc.library.cornell.edu/search/index.php?mode=browse&type=Collect ion>
- Data Archiving and Networked Services (2011a) "Data Management Plan" <http://www.dans.knaw.nl/en/content/categorieen/diensten/data-management-plan>
- Data Archiving and Networked Services (2011b), "Persistent Identifiers" <http://www.dans.knaw.nl/en/content/categorieen/diensten/persistent-identifiers>
- Deutsche Forschungsgemeinschaft (1998) "Proposals for Safeguarding good scientific practice Wiley-VCH" http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_0198.pdf
- Digital Curation Centre (2011) "DMP Online" <https://dmponline.dcc.ac.uk/>
- Dryad (2011) "Dryad Partners" <http://datadryad.org/partners>
- Economic and Social Data Service (2011) "About the Economic and Social Data Service" <http://www.esds.ac.uk/about/about.asp>
- Economic and Social Data Service (various) "Annual Reports" <http://www.esds.ac.uk/news/publications.asp>
- Engineering and Physical Sciences Research Council (2011) "Implementing the Delivery Plan" <http://www.epsrc.ac.uk/PLANS/IMPLEMENTINGDELIVERYPLAN/Pages/default.aspx>
- European Commission (12 December 2011a) "Digital Agenda: Turning government data into gold" <http://europa.eu/rapid/pressReleasesAction.do?reference=IP/11/1524&format=HTML&aged=0&language=EN&guiLanguage=en>
- European Commission (12 December 2011b) "Digital Agenda: Commission's Open Data Strategy, Questions & answers" <http://europa.eu/rapid/pressReleasesAction.do?reference=MEMO/11/891&format=HTML&aged=0&language=EN&guiLanguage=en>
- European Data Watch Extended (2011) "About EDaWaX" <http://www.edawax.de/about/>
- European Union (2010) "Riding the Wave: How Europe can gain from the rising tide of scientific data. Final report of the High level Expert Group on Scientific Data" <http://cordis.europa.eu/fp7/ict/e-infra-structure/docs/hlg-sdi-report.pdf>
- GESIS Leibniz Institute for the Social Sciences (2010) "Eurobarometer Data Service – International Data Sets Distributed via Archive Networks, 2005-2009" http://www.gesis.org/fileadmin/upload/dienstleistung/daten/umfragedaten/eurobarometer/contacts/eb-user-stat_archives_2005-2009_v2.pdf
- GESIS Leibniz Institute for the Social Sciences (2011) "Über da|ra" <http://www.gesis.org/dara/home/ueber-dara/>
- GESIS Leibniz Institute for the Social Sciences (2012) "Archiving and Data Management Training and Information Center" <http://www.gesis.org/archive-and-data-management-training-and-information-centre>
- Inter-university Consortium for Political and Social Research (2011a) "ICPSR Usage Statistics" <http://www.icpsr.umich.edu/icpsrweb/ICPSR/curation/usage.jsp>

- Inter-university Consortium for Political and Social Research (2011b)
 "Guidelines for Effective Data Management Plans" <http://www.icpsr.umich.edu/icpsrweb/ICPSR/dmp/index.jsp>
- International DOI Foundation (2011) "Welcome to the DOI® System" <http://www.doi.org/>
- Irish Qualitative Data Archive (2010) "Preparing Qualitative Data for Archiving" <http://www.iqda.ie/content/preparing-qualitative-data-archiving>
- Joint Information Systems Commission (2009) "Managing Research Data (JISCMRD)" <http://www.jisc.ac.uk/whatwedo/programmes/mrd.aspx>
- Kahneman, D. (2003) "A perspective on judgment and choice: Mapping bounded rationality" *American Psychologist*, 58(9), pp.697-720. doi: 10.1037/0003-066X.58.9.697
- King, G. (1995) "Replication, Replication" *PS: Political Science & Politics*, 28, pp.444-452
- Massachusetts Institute of Technology, "DSpace@MIT" <http://dspace.mit.edu/>
- National Institutes of Health (2003) "Final NIH Statement on Sharing Research Data" <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html>
- National Science Foundation (2011) "NSF Data Management Plan Requirements" <http://www.nsf.gov/eng/general/dmp.jsp>
- Natural Environment Research Council (2011) "Data Centres" <http://www.nerc.ac.uk/research/sites/data/>
- Organisation for Economic Co-operation and Development (2007) "OECD Principles and Guidelines for Access to Research Data from Public Funding" <http://www.oecd.org/dataoecd/9/61/38500813.pdf>
- Research Councils UK (2011) "RCUK Common Principles on Data Policy" <http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx>
- The Engineering and Physical Sciences Research Council (2011) "EPSRC Policy Framework on Research Data" <http://www.epsrc.ac.uk/about/standards/researchdata/Pages/impact.aspx>
- UK Data Archive (2010) "Data Management Practices in the Social Sciences" http://www.data-archive.ac.uk/media/203597/datamanagement_socialsciences.pdf pp.17-21
- UK Data Archive (2011) "Create and Manage Data" <http://www.data-archive.ac.uk/create-manage>
- University of California-San Diego (2010) "Digital Library Program" <http://libraries.ucsd.edu/about/digital-library/index.html>
- University of Edinburgh (2010) "Research Data Management Policy" <http://www.ed.ac.uk/schools-departments/information-services/about/policies-and-regulations/research-data-policy>
- University of Hertfordshire (2011), "Data Management Policy" <http://sitem.herts.ac.uk/secreg/upr/IM12.htm>
- University of Illinois at Urbana-Champaign (2005) "IDEALS" <http://www.ideals.illinois.edu/>
- University of Rochester (2008) "UR Research" <https://urresearch.rochester.edu/home.action>

Notes

1. GESIS-Leibniz Institute for the Social Sciences, Unter Sachsenhausen 6-8, 50667 Cologne, Germany, Tel. +49-221-47694 494. Email: laurence.horton@gesis.org and alexia.katsanidou@gesis.org
2. All online literature references available 29 March 2012.
3. A persuasive case for qualitative data reuse is made by Bishop (2009)