Fecher, Benedikt; Friesike, Sascha; Hebing, Marcel

Working Paper

What drives academic data sharing?

SOEPpapers on Multidisciplinary Panel Data Research, No. 655

Provided in Cooperation with:
German Institute for Economic Research (DIW Berlin)

Suggested Citation: Fecher, Benedikt; Friesike, Sascha; Hebing, Marcel (2014) : What drives academic data sharing?, SOEPpapers on Multidisciplinary Panel Data Research, No. 655, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin

This Version is available at:
https://hdl.handle.net/10419/97160

SOEP Das Sozio-oekonomische Panel

655 2014

SOEPpapers
on Multidisciplinary Panel Data Research

SOEP — The German Socio-Economic Panel Study at DIW Berlin 655-2014

What Drives Academic Data Sharing?

Benedikt Fecher, Sascha Friesike and Marcel Hebing

DIW SOEP

SOEPpapers on Multidisciplinary Panel Data Research
at DIW Berlin

This series presents research findings based either directly on data from the German Socio-Economic Panel Study (SOEP) or using SOEP data as part of an internationally comparable data set (e.g. CNEF, ECHP, LIS, LWS, CHER/PACO). SOEP is a truly multidisciplinary household panel study covering a wide range of social and behavioral sciences: economics, sociology, psychology, survey methodology, econometrics and applied statistics, educational science, political science, public health, behavioral genetics, demography, geography, and sport science.

The decision to publish a submission in SOEPpapers is made by a board of editors chosen by the DIW Berlin to represent the wide range of disciplines covered by SOEP. There is no external referee process and papers are either accepted or rejected without revision. Papers appear in this series as works in progress and may also appear elsewhere. They often represent preliminary studies and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be requested from the author directly.

Any opinions expressed in this series are those of the author(s) and not those of DIW Berlin. Research disseminated by DIW Berlin may include views on public policy issues, but the institute itself takes no institutional policy positions.

The SOEPpapers are available at
http://www.diw.de/soeppapers

ISSN: 1864-6689 (online)

German Socio-Economic Panel Study (SOEP)
DIW Berlin
Mohrenstrasse 58
10117 Berlin, Germany

Contact: Uta Rahmann | soeppapers@diw.de

What Drives Academic Data Sharing?

Benedikt Fecher[a, b, *, 1], Sascha Friesike[b], Marcel Hebing[a]

[a] German Institute for Economic Research, Mohrenstraße 58, 10117 Berlin (Germany)

[*] Corresponding author; mail: bfecher@diw.de; tel.: +49 30 209 334 90; fax: +49 30 20 93 34 35

[b] Alexander von Humboldt Institute for Internet and Society, Unter den Linden 6, 10115 Berlin (Germany)

Abstract

Despite widespread support from policy makers, funding agencies, and scientific journals, academic researchers rarely make their research data available to others. At the same time, data sharing in research is attributed a vast potential for scientific progress. It allows the reproducibility of study results and the reuse of *old* data for *new* research questions. Based on a systematic review of 98 scholarly papers and an empirical survey among 603 secondary data users, we develop a conceptual framework that explains the process of data sharing from the primary researcher's point of view. We show that this process can be divided into six descriptive categories: *Data donor*, *research organization*, *research community*, *norms*, *data infrastructure*, and *data recipients*. Drawing from our findings, we discuss theoretical implications regarding knowledge creation and dissemination as well as research policy measures to foster academic collaboration. We conclude that research data cannot be regarded a knowledge commons, but research policies that better incentivize data sharing are needed to improve the quality of research results and foster scientific progress.

JEL

C81, C82, D02, H41, L17, Z13

Keywords

Data Sharing, Academia, Systematic Review, Research Policy, Knowledge Commons, Crowd Science, Commons-based Peer Production, SOEP

[1] A significant part of Benedikt Fecher's contribution to the paper was funded by the Deutsche Forschungsgemeinschaft (PI: Gert G. Wagner, DFG-Grant Number WA 547/6-1).

1. Introduction

The accessibility of research data has a vast potential for scientific progress. It facilitates the replication of research results and allows the application of *old* data in *new* contexts (Dewald et al., 1986; McCullough, 2009). It is hardly surprising that the idea of shared research data finds widespread support among academic stakeholders. The European Commission, for example, proclaims that access to research data will boost Europe's innovation capacity. To tap into this potential, data produced with EU funding should to be accessible from 2014 onwards (European Commission, 2012). Simultaneously, national research associations band together to promote data sharing in academia. The *Knowledge Exchange Group*, a joint effort of five major European funding agencies, is a good example for the cross-border effort to foster a culture of sharing and collaboration in academia (Knowledge Exchange Group, 2013). Journals such as *Atmospheric Chemistry and Physics, F1000Research, Nature,* or *PLoS One*, increasingly adopt data sharing policies with the objective of promoting public access to data (Silva, 2014).

In a study among 1,329 scientists employed in the US, 46 % reported they do not make their data electronically available to others (Tenopir et al., 2011). In the same study, around 60 % of the respondents, across all disciplines, agreed that the lack of access to data generated by others is a major impediment to progress in science. The results point to a striking dilemma in academic research, namely the mismatch between the general interest and the individual's behavior. At the same time, they raise the question of what exactly prevents researchers from sharing their data with others.

Still, little research devotes itself to the issue of data sharing in a comprehensive manner. In this article we offer a cross-disciplinary analysis of prevailing barriers and enablers, and propose a conceptual framework for data sharing in academia. The results are based on a) a systematic review of 98 scholarly papers on the topic and b) a survey among 603 secondary data users who are analyzing data from the German Socio-Economic Panel Study (hereafter SOEP). With this paper we aim to contribute to research practice through policy implications and to theory by comparing our results to current organizational concepts of knowledge creation, such as commons-based peer production (Benkler, 2002) and crowd science (Franzoni and Sauermann, 2014). We show that data sharing in today's academic world cannot be regarded a knowledge commons (Allen et al., 2014, Vlaeminck and Wagner, 2014).

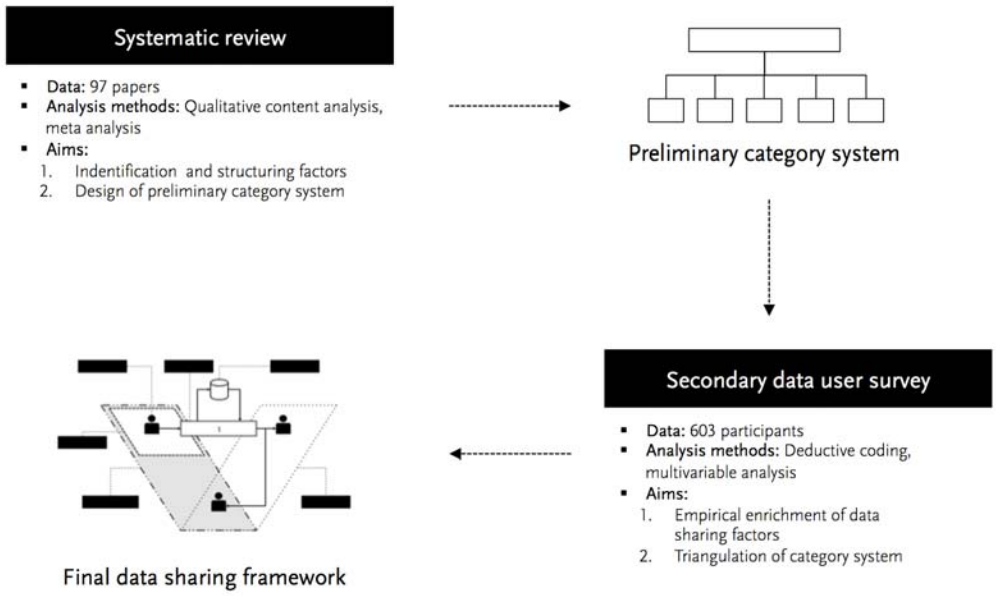
The remainder of this article is structured as follows: first, we explain how we methodologically arrived at our framework. Second, we will describe its categories and address the

predominant factors for sharing research data. Drawing from these results, we will in the end discuss theory and policy implications.

2. Methodology

In order to arrive at a framework for data sharing in academia, we used a systematic review of scholarly articles and an empirical survey among secondary data users (*SOEP User survey*). The first served to design a preliminary category system, the second to empirically revise it. In this section, we delineate our methodological approach as well as its limitations. Figure 1 summarizes the research design and methods of analysis.

Figure 1: Research Design



2.1. Systematic Review

Systematic reviews have proven their value especially in evidence-based medicine (Higging and Green, 2008). Here, they are used to systematically retrieve research papers from literature databases and analyze them according to a pre-defined research question. Today, systematic reviews are applied across all disciplines, reaching from educational policy (e.g., Davies, 2000; Hammersley, 2001) to innovation research (e.g., Dahlander and Gann, 2010). In our view, a systematic review

constitutes an elegant way to start an empirical investigation. It helps to gain an exhaustive overview of a research field, its leading scholars, and prevailing discourses and can be used to produce an analytical spadework for further inquiries.

2.1.1. Retrieval of Relevant Papers

In order to find the relevant papers for our research intent, we defined a research question (*Which factors influence data sharing in academia?*) as well as explicit selection criteria for the inclusion of papers (Booth, 2012). According to the criteria, the papers needed to address the perspective of the primary researcher, focus on academia, provide new insights, and stem from defined evaluation period. To ensure an as exhaustive first sample of papers as possible, we used a broad basis of multidisciplinary data banks (see table 1) and a search term ("*data sharing")* that generated a high number of search results. We did not limit our sample to research papers but also included for example discussion papers. In the first sample we included every paper that has the search term in either title or abstract.

The evaluation period spanned from December 1st 2001 to November 15th 2013, leading to a pre-sample of 9796 papers. We read the abstracts of every paper and selected only those those that a) address data sharing in academia and b) deal with the perspective of the primary researcher. In terms of intersubjective comprehensibility, we decided separately for every paper if it meets the defined criteria (Given, 2008). Only those papers were included in the final analysis sample that were approved by all three coders (yes/yes/yes in the coding sheet). Papers that received a no from every coder were dismissed; the others were discussed and jointly decided upon. The most common reasons for dismissing a paper were thematic mismatch (e.g., paper focusses on commercial data), and quality issues (e.g., a letter to the editor). Additionally, we conducted a small-scale expert poll on the social network for scientists *ResearchGate*. The poll resulted in five additional papers, three of which were not published in the defined evaluation period. We did, however, include them in the analysis sample due to their thematic relevance. In the end, we arrived at a sample of 98 papers. Table 1 shows the selected papers and the database in which we found them.

Table 1: Paper and databases of the final sample

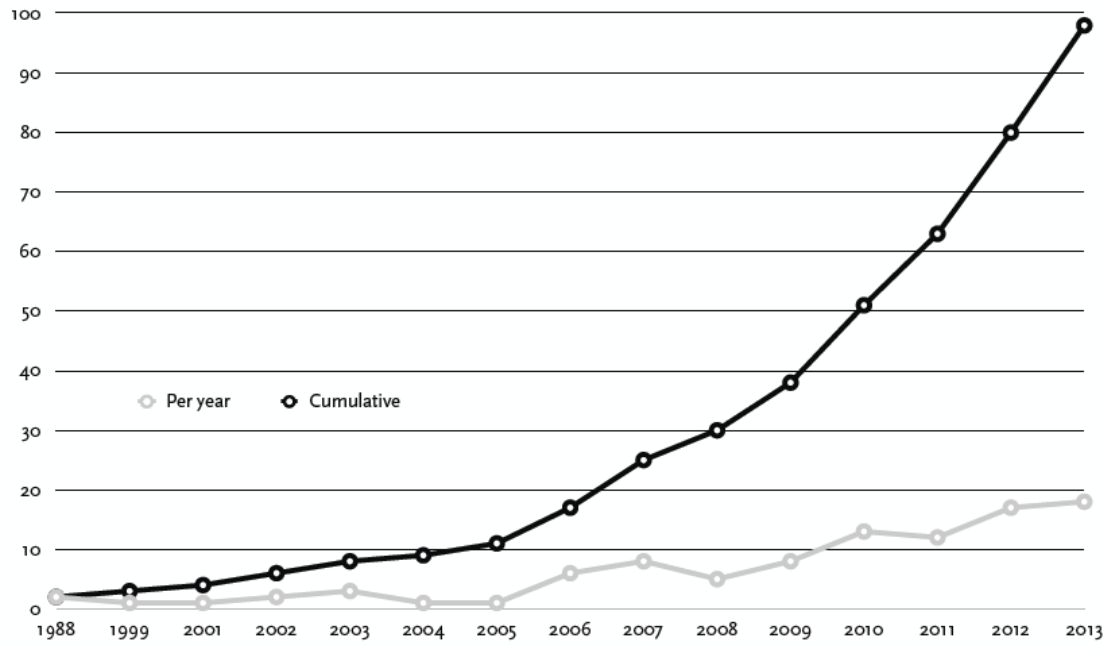
Database	Papers in analysis sample
Ebsco	Butler, 2007; Chokshi et al., 2006; De Wolf et al., 2005 (also JSTOR, ProQuest); De Wolf et al., 2006 (also ProQuest); Feldman et al., 2012; Harding et al., 2011; Jiang et al., 2013; Nelson, 2009; Perrino et al., 2013; Pitt and Tang, 2013; Teeters et al., 2008 (also Springer)
JSTOR	Anderson and Schonfeld, 2009; Axelsson and Schroeder, 2009 (also ProQuest); Cahill and Passamano, 2007; Cohn, 2012; Cooper, 2007; Costello, 2009; Fulk et al., 2004; Guralnick and Constable, 2010; Linkert et al., 2010; Ludman et al., 2010; Myneni and Patel, 2010; Parr, 2007; Resnik, 2010; Rodgers and Nolte, 2006; Sheather, 2009; Whitlock et al., 2010; Zimmerman, 2008
PLOS	Alsheikh-Ali et al,. 2011; Chandramohan et al., 2008; Constable et al., 2010; Drew et al., 2013; Haendel et al., 2012; Huang et al., 2013; Masum et al., 2013; Milia et al., 2012; Molloy, 2011; Noor et al., 2006; Piwowar, 2011; Piwowar et al., 2007; Piwowar et al., 2008; Savage and Vickers, 2009; Tenopir et al., 2011; Wallis et al., 2013; Wicherts et al., 2011;
ProQuest	Acord and Harley, 2012; Belmonte et al., 2007; Edwards et al., 2011; Eisenberg, 2006; Elman et al., 2010; Kim and Shanton, 2013; Nicholson and Bennett, 2011; Rai and Eisenberg, 2006; Reidpath and Allotey, 2001(also Wiley); Tucker, 2009
ScienceDirect	Anagnostou et al., 2013; Brakewood and Poldrack, 2013; Enke et al., 2011; Fisher and Fortman, 2010; Karami et al., 2013; Mennes et al., 2013; Parr and Cummings, 2005; Piwowar and Chapman, 2010; Rohlfing and Poline, 2012; Sayogo and Pardo, 2013; Van Horn and Gazzaniga, 2013; Wicherts and Bakker, 2012
Springer	Albert, 2012; Bezuidenhout, 2013; Breeze et al., 2012; Fernandez et al., 2012; Freymann et al., 2012; Gardner et al., 2003; Jarnevich et al., 2007; Jones et al., 2012; Pearce and Smith, 2011; Sansone and Rocca-Serra, 2012
Wiley	Borgman, 2012; Daiglesh et al., 2012; Delson et al., 2007; Eschenfelder and Johnson, 2011; Haddow, 2011; Hayman et al., 2012; Huang et al., 2012; Kowalcyk and Shankar, 2013; Levenson, 2010; NIH, 2002; NIH, 2003; Ostell, 2009; Piwowar, 2010; Rushby, 2013; Samson, 2008; Weber, 2013
From the expert poll	Campbell et al., 2002; Cragin et al., 2010; Overbey, 1999; Sieber, 1988; Stanley and Stanley 1988

2.1.2. Sample description

The 98 papers that made our final sample come from the following disciplines: Science, technology, engineering, and mathematics (60 papers), humanities (9), social sciences (6), law (1), interdisciplinary or no disciplinary focus (22). The distribution of the papers indicates that data sharing is an issue of relevance across all research areas, above all the STEM disciplines. The graph

of our analysis sample (see figure 2) indicates that academic data sharing is a topic that has received a considerable increase in attention during the last decade.

Figure 2: Papers in the sample by year



Further, we analyzed the references that the 98 papers cited. Table 2 lists the most cited papers in our sample and provides an insight into which articles and authors dominate the discussion. Two of the top three most cited papers come from the journal PLoS One. Among the most cited texts, Feinberg et al. (1985) is the only reference that is older than 2001.

Table 2: Most cited within the sample

Reference	# Citations
Piwowar, H.A., Day, R.S., Fridsma, D.B. 2007. Sharing detailed research data is associated with increased citation rate. PLoS ONE, 2(3): e308.	15
Campbell, E.G., Clarridge, B.R., Gokhale, M., Birenbaum, L., Hilgartner, S., Holtzman, N.A., Blumenthal, D. 2002. Data withholding in academic genetics: evidence from a national survey. JAMA, 287(4): 473–480.	15
Savage, C.J., Vickers, A.J. 2009. Empirical study of data sharing by authors publishing in PloS journals. PLoS ONE 4(9): e7078.	12
Feinberg, S.E., Martin, M.E., Straf, M.L. 1985. Sharing Research Data. Washington, DC: National Academy Press.	9
NIH National Institutes of Health. 2003. Final NIH statement on sharing research data Available.	9
Wicherts, J.M., Borsboom, D., Kats, J., Molenaar, D. 2006. The poor availability of psychological research data for reanalysis. American Psychologist, 61(7): 726–728.	9
Nelson, B. 2009. Data sharing: empty archives. Nature, 461: 160–163.	8
Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A.U., Wu, L., Read, E., Manoff, M., Frame, M. 2011. Data Sharing by Scientists: Practices and Perceptions. PloS ONE 6(6): e21101.	8
Gardner, D., Toga, A.W., Ascoli, G.A., Beatty, J.T., Brinkley, J.F., Dale, A.M., Fox, P.T., Gardner, E.P., George, J.S., Goddard, N., Harris, K.M., Herskovits, E.H., Hines, M.L., Jacobs, G.A., Jacobs, R.E., Jones, E.G, Kennedy, D.N., Kimberg, D.Y., Mazziotta, J.C., Perry L. Miller, Mori, S., Mountain, D.C., Reiss, A.L., Rosen, G.D., Rottenberg, D.A., Shepherd, G.M., Smalheiser, N.R., Smith, K.P., Strachan, T., Van Essen, D.C., Williams, R.W., Wong, S.T.C. 2003. Towards effective and rewarding data sharing. Neuroinformatics, 1(3): 289–295.	8
Borgman C.L. 2007. Scholarship in the Digital Age: Information, Infrastructure, and the Internet. Cambridge, MA: MIT Press.	8
Whitlock M.C. 2011. Data Archiving in Ecology and Evolution: Best Practices. Trends in Ecology and Evolution, 26(2): 61–65.	7

2.1.3. Preliminary Category System

In a consecutive, we applied a qualitative content analysis in order to build a category system and to condense the content of the literature in the sample (Mayring, 2000; Schreier, 2012). We defined the analytical unit compliant to our research question and copied all relevant passages in a *CSV* file.

After, we uploaded the file to the data analysis software *NVivo* and coded the units of analysis inductively. We decided for inductive coding as it allows building categories and establishing novel interpretative connections based on the data material, rather than having a conceptual pre-understanding. The preliminary category system allows allocating the identified factors to the involved individuals, bodies, regulatory systems, and technical components.

2.2. Survey Among Secondary Data Users

To empirically revise our preliminary category system, we further conducted a survey among secondary data users that analyze data from the German Socio-Economic Panel (SOEP). We specifically addressed secondary data users because this researcher group is familiar with the re-use of data and likely to offer informed responses.

The SOEP is a representative longitudinal study of private households in Germany conducted by the German Institute for Economic Research (Wagner et al., 2007). The data is available to researchers through a research data center. Currently, the SOEP has approximately 500 active user groups with more than 1,000 researchers per year analyzing the data. Researchers are allowed to use the data for their scientific projects and publish the results, but must neither re-publish the data nor syntax files as part of their publications.

The SOEP User Survey is a web-based usability survey among researchers who use the panel data. Beside an annually repeated module of socio-demographic and service related questions, the 2013 questionnaire included three additional questions on data sharing (see table 3). And the questionnaire includes the Big Five personality scale according to Richter et al. (2013) that we correlated with the willingness to share (see *3.1. Data Donor*).

When working with panel surveys like the SOEP, researchers expend serious effort to process and prepare the data for analysis (e.g., generating new variables). Therefore the questions were designed more broadly, including the willingness to share analysis scripts.

Table 3: Questions for secondary data users

Q1 We are considering giving SOEP users the possibility to make their baskets and perhaps also the scripts of their analyses or even their own datasets available to other users within the framework of SOEPinfo. Would you be willing to make content available here? ▪ Yes, I would be willing to make my own data and scripts publicly available ▪ Yes, but only on a controlled-access site with login and password ▪ Yes, but only on request ▪ No
Q2 What would motivate you to make your own scripts or data available to the research community? → open answer
Q3 What concerns would prevent you from making your own scripts or data available to the research community? → open answer

The web survey was conducted in November and December 2013, resulting in 603 valid response cases – of which 137 answered the open questions *Q2* and *Q3*. We analyzed the replies to these two open questions by applying deductive coding and using the categories from the preliminary category system. We furthermore used the replies to revise categories in our category system and add empirical evidence.

The respondents are on average 37 years old, 61 % of them are male. Looking at the distribution of disciplines among the researchers in our sample, the majority works in economics (46 %) and sociology/social sciences (39 %). For a study based in Germany it is not surprising that most respondents are German (76 %). Nevertheless, 24 % of the respondents are international data users.

2.3. Limitations

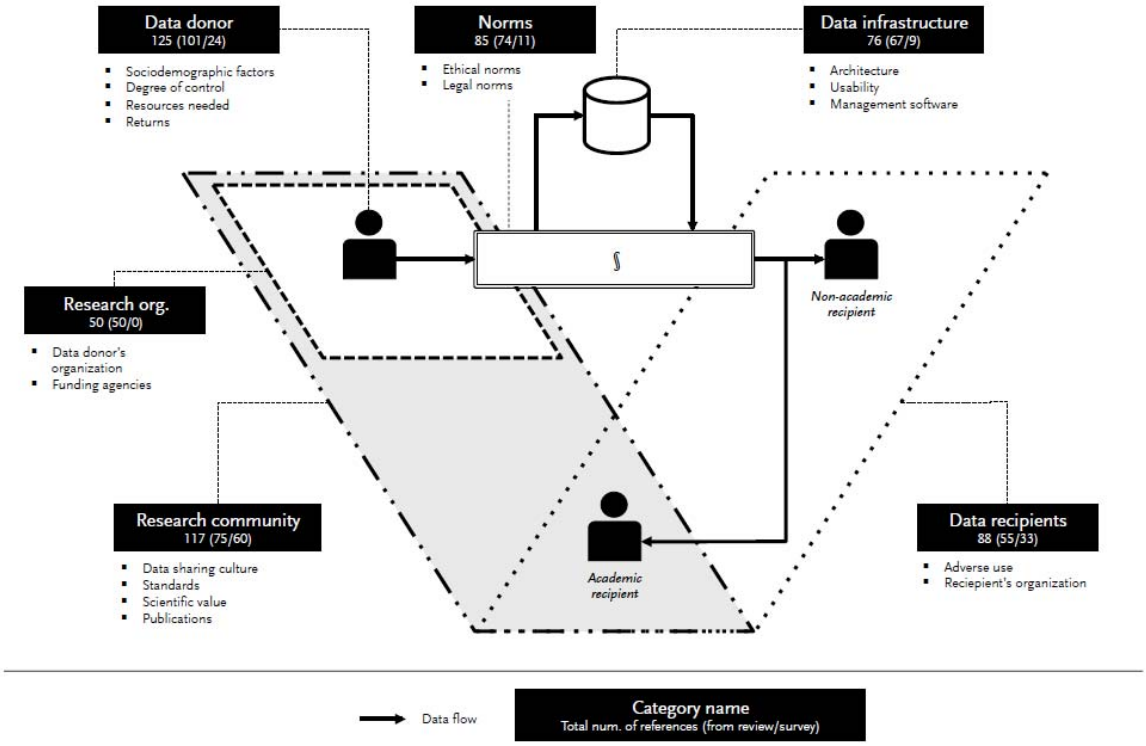
Our methodological approach goes along with common limitations of systematic reviews and qualitative methods. The sample of papers in the systematic review is limited to journal publications in well-known databases and excludes for example monographs, grey literature, and papers from public repositories such as preprints. Our sample does in this regard draw a picture of specific scope, leaving out for instance texts from open data initiatives or blog posts. Systematic reviews are furthermore prone to publication-bias (Torgerson, 2006), the tendency to publish positive results

rather than negative results. We tried to counteract a biased analysis by triangulating the derived category system with empirical data from a survey among secondary data users (Patton, 2002). For the analysis, we leaned onto quality criteria of qualitative research. Regarding the validity of the identified categories, an additional quantitative survey is however recommended.

3. Results

As a result of the systematic review and the survey we arrived at a framework that depicts academic data sharing in six descriptive categories. Figure 3 provides an overview of these six (*data donor*, *research organization*, *research community*, *norms*, *data infrastructure*, and *data recipients*) and highlights how often we found references for them in a) the literature review and b) in the survey (a/b). In total we found 541 references, 404 in the review and 137 in our survey. Furthermore, the figure shows the subcategories of each category.

Figure 3: Framework for academic data sharing



- *Data donor*, comprising factors regarding the individual researcher who is sharing data (e.g., invested resources, returns received for sharing)

- *Research organization*, comprising factors concerning the crucial organizational entities for the donating researcher, being the own organization and funding agencies (e.g., funding policies)
- *Research community*, comprising factors regarding the disciplinary data-sharing practices (e.g., formatting standards, sharing culture)
- *Norms*, comprising factors concerning the legal and ethical codes for data sharing (e.g., copyright, confidentiality)
- *Data recipients*, comprising factors regarding the third party reuse of shared research data (e.g., adverse use)
- *Data infrastructure*, comprising factors concerning the technical infrastructure for data sharing (e.g., data management system, technical support)

In the following, we explain the hindering and enabling factors for each category. A table that lists the identified sub-categories and data sharing factors summarizes each category. The tables further provide text references and direct quotes for selected factors. We translated most of the direct quotes from the survey from German to English (76 % German respondents).

3.1. Data Donor

The category *data donor* concerns the individual researcher who collects data. The sub-categories are *sociodemographic factors*, *degree of control*, *resources needed*, and *returns*.

Table 4: Overview category data donor

Sub-category	Factors	References	Exemplary Quotes
Sociodemographic factors	▪ Nationality ▪ Age ▪ Seniority and career prospects ▪ Character traits ▪ Research practice	Acord and Harley, 2012; Enke et al., 2011; Milia et al., 2012; Piwowar, 2011; Tenopir et al., 2011	*Age:* "There are some differences in responses based on age of respondent. Younger people are less likely to make their data available to others (either through their organization's website, PI's website, national site, or other sites.). People over 50 showed more interest in sharing data (...)" (Tenopir et al. 2011, p. 14)
Degree of control	▪ Knowledge about data requester ▪ Having a say in the data use ▪ Priority rights for	Acord and Harley, 2012; Belmonte et al., 2007; Bezuidenhout, 2013; Constable et al., 2010; Enke et al., 2011; Fisher and Fortman, 2010; Huang et al., 2013; Jarnevich et al.,	*Having a say in the data use:* "I have doubts about others being able to use my work without control from my side"

	publications	2007; Pearce and Smith, 2011; Pitt and Tang, 2013; Stanley and Stanley, 1988; Tenopir et al., 2011; Whitlock et al., 2010; Wallis et al., 2013	(Survey)
Resources needed	▪ Time and effort ▪ Skills and knowledge ▪ Financial resources	Acord and Harley, 2012; Axelsson and Schroeder, 2009; Breeze et al., 2012; Campbell et al., 2002; Cooper, 2007; Costello, 2009; De Wolf et al., 2006; De Wolf et al., 2005; Enke et al., 2011; Gardner et al., 2003; Guralnick and Constable, 2010; Haendel et al., 2012; Huang et al., 2013; Kowlcyk and Shankar, 2013; Nelson, 2009; Noor et al., 2006; Perrino et al., 2013; Piwowar et al., 2008; Reidpath and Allotey, 2001; Rushby, 2013; Savage and Vickers, 2009; Sayogo and Pardo, 2013; Sieber, 1988; Stanley and Stanley, 1988; Teeters et al., 2008; Van Horn and Gazzaniga, 2013; Wallis et al., 2013; Wicherts and Bakker, 2012	*Time and effort:* "The effort to collect data is immense. To collect data yourself becomes almost out of fashion." (Survey)
Returns	▪ Formal recognition ▪ Professional exchange ▪ Quality improvement	Acord and Harley, 2012; Costello, 2009, Daiglesh et al., 2012; Enke et al., 2011; Gardner et al., 2003; Mennes et al., 2013; Nelson, 2009; Ostelle, 2009; Parr, 2007; Perrino et al., 2013; Piwowar et al., 2007; Reidpath and Allotey, 2001; Rohlfing and Poline, 2012; Stanley and Stanley, 1988; Wallis et al., 2013; Wicherts and Bakker, 2012; Whitlock, 2011	*Formal recognition:* "… the science reward system has not kept pace with the new opportunities provided by the internet, and does not sufficiently recognize online data publication." (Costello, 2009, p. 426)

Sociodemographic factors. Frequently mentioned in the literature were the factors *age*, *nationality,* and *seniority in the academic system*. Enke et al. (2011), for instance, observe that German and Canadian scientists were more reluctant to share research data publicly than their US colleagues (which raises the question how national research policies influence data sharing). Tenopir et al. (2011) found that there is an influence of the researcher's *age* on the willingness to share data. Accordingly, younger people are less likely to make their data available to others. People over 50, on the other hand, were more likely to share research data. This result resonates with an assumed influence of *seniority in the academic system* and competitiveness on data sharing behavior (Milia et al., 2012). Data sets and other subsidiary products are awarded far less credit in tenure and promotion decisions than text

publications (Acord and Harley, 2012). Hence does competition, especially among non-tenured researchers, go hand in hand with a reluctance to share data. The perceived *right to publish first* (see *degree of control*) with the data further indicates that publications and not (yet) data is the currency in the academic system. Tenopir et al. (2011) point to an influence of the level of *research activity* on the willingness to share data. Individuals who work solely in research, in contrast to researchers who have time-consuming teaching obligations, are more likely to make their data available to other researchers. Acord and Harley (2012) further regard *character traits* as an influencing factor. This conjecture is not vindicated in our questionnaire. In contrast to our initial expectations, character traits (Big Five) are not able to explain much of the variation in *Q1*. In a logistic regression model on the willingness to share data and scripts in general (answer categories 1–3) and controlling for age and gender, only the *openness* dimension shows a significant influence. All other dimensions (*conscientiousness*, *extraversion*, *agreeableness*, *neuroticism*) do not have a considerable influence on the willingness to share.

Degree of control. A core influential factor on the individual data sharing behavior can be subsumed under the category *degree of control*. It denotes the researcher's need to *have a say* or at least *knowledge* regarding the access and use of the deposited data.

The relevance of this factor is emphasized by the results to question Q1 in our survey. Only a small number of researchers (18 %) categorically refuses to share scripts or research data. For those who are willing to share (82 %), control seems to be an important issue (summarized by the first three questions). 56 % are either demanding a context with access control or would only be willing to share on request. However, it has to be said that our sample comprises researchers that are familiar with secondary data and is therefore not representative for the academia in general.

Table 5: Question on sharing analysis scripts and data sets

Sharing analysis scripts and data sets	**Freq**	**Perc (valid)**
1. Willing to share publicly	120	25.9 %
2. Willing to share under access control	98	21.1 %
3. Willing to share only on request	163	35.1 %
4. Not willing	83	17.9 %
Sum	***464***	***100 %***

Sources: SOEP User Survey 2013, own calculations

Eschenfelder and Johnson (2011) suggest more control for researchers over deposited data (see also Haddow, 2011; Jarnevich et al., 2007; Pearce, 2011). According to some scholars, a *priority right* for publications, for example an embargo on data (e.g., Van Horn and Gazzaniga, 2013), would enable academic data sharing. Other authors point to a researcher's concern regarding the technical ability of the data requester to understand (Costello, 2009; Stanley and Stanley, 1988) and to interpret (Huang et al., 2013) a dataset (see also data recipients→adverse use). The need for control is also present in our survey among secondary data users. To the question why one would not share research data, one respondent replied: "I have doubts about others being able to use my work without control from my side" (Survey). Another respondent replied: "I want to know who uses my data." (Survey). The results in this category indicate a perceived ownership over the data on the part of the researcher, which is legally often not the case.

Resources needed. Here we subsume factors relating to the researcher's investments in terms of *time* and *costs* as well as their *knowledge* regarding data sharing. "Too much effort!" was a blunt answer we found in our survey as a response to the question why researchers do not share data. In the literature we found the argument *time and effort* 19 times and seven times in the survey. Besides the actual sharing effort (Acord and Harley, 2012; Enke et al., 2011; Gardner et al., 2003; Van Horn and Gazzaniga, 2013), scholars utter concerns regarding the effort required to help others to make sense of the data (Wallis et al., 2013). The *knowledge* factor becomes apparent in Sieber's study (1988) in which most researchers stated that data sharing is advantageous for science, but that they had not thought about it until they were asked for their opinion. Missing knowledge further relates to poor curation and storing skills (Haendel et al., 2012, Van Horn and Gazzaniga, 2013) and missing knowledge regarding adequate repositories (Enke et al., 2011; Wallis et al., 2013). In general, missing knowledge regarding the existence of databases and know-how to use them is described as a hindering factor for data sharing. Several scholars, for instance Piwowar et al. (2008) and Teeters et al. (2008), hence suggest to integrate data sharing in the curriculum. Others mention the *financial effort* to share data and suggest forms of financial compensation for researchers or their organizations (De Wolf et al., 2006; Sieber, 1988).

Returns. Within the examined texts we found 26 references that highlight the issue of missing returns in exchange for sharing data, 12 more came from the survey. The basic attitude of the references describes a lack of *recognition* for data donors (Daiglesh et al., 2012; Parr, 2007; Stanley and Stanley, 1988). Both sources – review and survey – argue that donors do not receive enough formal *recognition* to justify the individual efforts and that a safeguard against uncredited use is necessary

(Gardner et al., 2003; Ostelle, 2009). The form of attribution a donor of research data should receive remains unclear and ranges from a mentioning in the acknowledgements to citations and co-authorships (Enke et al., 2011). Several authors explain that impact metrics need to be adapted to foster data sharing (Costello, 2009; Parr, 2007). Yet, there is also literature that reports positive individual returns from shared research data. Kim and Shanton (2013) for instance explain that shared data can highlight the quality of a finding and thus indicate sophistication. Piwowar et al. (2007) report an increase in citation scores for papers, which feature supplementary data. Further, *quality improvements* in the form of *professional exchange* are mentioned: "Seeing how others have solved a problem may be useful.", "I can profit and learn from other people's work." (Both quotes are from our survey). Enke et al. (2013) also mention an increased visibility within the research community as a possible positive return.

3.2. Research Organization

The category *research organization* comprises the most relevant organizational entities for the donating researcher. These are the data donor's own organization as well as funding agencies.

Table 6: Overview category research organization

Sub-category	Factors	References	Exemplary Quotes
Data donor's organization	▪ Data sharing policy and organizational culture ▪ Data management	Belmonte et al., 2013; Breeze et al., 2012; Cragen et al., 2010; Enke et al., 2013; Huang et al., 2013; Masum et al., 2013; Pearce and Smith, 2011; Perrino et al., 2013; Savage and Vickers, 2009; Sieber, 1988	*Data sharing policy and organizational culture:* "Only one-third of the respondents reported that sharing data was encouraged by their employers or funding agencies." (Huang et al., 2013, p. 404)
Funding agencies	▪ Funding policy (grant requirements) ▪ Financial compensation	Axelsson and Schroeder, 2009; Borgman, 2012; Eisenberg, 2006; Enke et al., 2011; Cohn, 2012; Fernandez et al., 2012; Huang et al., 2013; Mennes et al., 2013; Nelson, 2009; NIH, 2002; NIH, 2003; Perrino et al., 2013; Pitt and Tang, 2013, Piwowar et al., 2008; Sieber, 1988; Stanley and Stanley, 1988; Teeters et al., 2008; Wallis, 2013; Wicherts and Bakker, 2012	*Funding policy (grant requirements):* "… until data sharing becomes a requirement for every grant […] people aren't going to do it in as widespread of a way as we would like." (Nelson, 2009, p. 161)

Data donor's organization. An individual researcher is generally placed in an organizational context, for example a university, a research institute or a research and development department of a company. The respective organizational affiliation impinges on his or her data sharing behavior especially through internal policies, the organizational culture as well as the available data infrastructure. Huang et al. (2013, p. 404) for instance, in an international survey on biodiversity data sharing found out that "only one-third of the respondents reported that sharing data was encouraged by their employers or funding agencies". The respondents whose organizations or affiliations encourage data sharing were more willing to share. Huang et al. (2013) view the *organizational policy* as a core adjusting screw. They suggest detailed data management and archiving instructions as well as recognition for data sharing (i.e., career options). Belmonte et al. (2007) and Enke et al. (2011) further emphasize the importance of intra-organizational *data management,* for instance consistent data annotation standards in laboratories (see also *3.6. Data Infrastructure*). Cragen et al. (2010, p. 4025) see data sharing rather as a community effort in which the single organizational entity plays a minor role: "As a research group gets larger and more formally connected to other research groups, it begins to function more like big science, which requires production structures that support project coordination, resource sharing and increasingly standardized information flow".

Funding agencies. Besides journal policies, the *policies* of funding agencies are named as a key adjusting screw for academic data sharing throughout the literature (e.g., Enke et al., 2011; Wallis, 2013). Huang et al. (2013) argue that making data available is no obligation with many funding agencies and that they do not provide sufficient *financial compensation* for the efforts needed in order to share data. Perrino et al. (2013) argue that funding policies show varying degrees of enforcement when it comes to data sharing and that binding policies are necessary to convince researchers to share. The National Science Foundation of the US, for instance, has long required data sharing in its grant contracts (see Cohn, 2012), "but has not enforced the requirements consistently" (Borgman 2012, p. 1060).

3.3. Research Community

The category research community subsumes the sub-categories data sharing culture, standards, scientific value, and publications.

Table 7: Overview category research community

Sub-category	Factors	References	Exemplary Quotes
Data sharing culture	▪ Disciplinary practice	Costello, 2009; Enke et al., 2011; Haendel et al., 2012; Huang et al., 2013; Milia et al., 2012; Nelson, 2009; Tenopir et al., 2011	*Disciplinary practice:* "The main obstacle to making more primary scientific data available is not policy or money but misunderstandings and inertia within parts of the scientific community." (Costello, 2009, p. 419)
Standards	▪ Metadata ▪ Formatting standards ▪ Interoperability	Axelsson and Schroeder, 2009; Costello, 2009; Delson et al., 2007; Edwards et al., 2011; Enke et al., 2011; Freymann et al., 2012; Gardner et al., 2003; Haendel et al., 2012; Jiang et al., 2013; Jones et al., 2012; Linkert et al., 2010; Milia et al., 2012; Nelson, 2009; Parr, 2007; Sayogo and Pardo, 2013; Teeters et al., 2008; Tenopir et al., 2011; Wallis et al., 2013; Whitlock, 2011	*Metadata:* "In our experience, storage of binary data [...] is based on common formats [...] or other formats that most software tools can read [...]. The much more challenging problem is the metadata. Because standards are not yet agreed upon." (Linkert et al., 2010, p. 779)
Scientific value	▪ Scientific progress ▪ Exchange ▪ Review ▪ Synergies	Alsheikh-Ali et al., 2011; Bezuidenhout, 2013; Butler, 2007, Chandramohan et al., 2008; Chokshi et al., 2006; Costello, 2009; De Wolf et al., 2005; Huang et al., 2013; Ludman et al., 2010; Molloy, 2011; Nelson, 2009; Perrino et al., 2013; Pitt and Tang, 2013; Piwowar et al., 2008; Stanley and Stanley, 1988; Teeters et al., 2008; Tenopir et al., 2011; Wallis et al., 2013; Whitlock et al., 2010	*Exchange:* "The main motivation for researchers to share data is the availability of comparable data sets for comprehensive analyses (72%), while networking with other researchers (71%) was almost equally important." (Enke et al., 2011, p. 28)
Publications	▪ Journal policy	Alsheikh-Ali et al., 2011; Anagnostou et al., 2013; Chandramohan et al., 2008; Costello, 2009; Enke et al., 2011; Huang et al., 2012; Huang et al., 2013; Milia et al., 2012; Noor et	*Data sharing policy:* "It is also important to note that scientific journals may benefit from adopting stringent sharing data rules since papers whose datasets are available without

	al., 2006; Parr, 2007; Pearce, 2011; Piwowar, 2010; Piwowar, 2011; Piwowar and Chapman, 2009; Tenopir et al., 2011; Wicherts et al., 2011; Whitlock 2011	restrictions are more likely to be cited than withheld ones." (Milia et al., 2012, p. 3)

Data sharing culture. The literature reports a substantial variation in academic data sharing across *disciplinary practices* (Nelson, 2009; Milia et al., 2012). Even fields, which are closely related like medical genetics and evolutionary genetics, show substantially different sharing rates (Milia et al., 2012). Medical research and social sciences are reported to have an overall low data sharing culture (Tenopir et al., 2011), which possibly relates to the fact that these disciplines work with individual-related data. Costello (2009) goes so far as to describe the data sharing culture as the main obstacle to academic data sharing (see table 7 for quote).

Standards. When it comes to the interoperability of data sets, many scholars see the absence of *metadata* standards and *formatting standards* as an impediment for sharing and reusing data; lacking standards hinder *interoperability* (e.g., Axelsson and Schroeder, 2009; Costello, 2009; Enke et al., 2011; Jiang et al., 2013; Linkert et al., 2010; Nelson, 2009; Parr, 2007; Teeters et al., 2008; Tenopir et al., 2011). However, there were no references in the survey for the absence of formatting standards.

Scientific value. In this subcategory we subsume all findings that bring value to the scientific community. It is a very frequent argument that data sharing can enhance *scientific progress*. A contribution hereto is often considered an intrinsic motivation for participation. This is supported by our survey: We found sixty references for this subcategory, examples are "Making research better", "Feedback and exchange", "Consistency in measures across studies to test robustness of effect", "Reproducibility of one's own research". Huang et al. (2013) report that 90 % of their "respondents indicated the desire to contribute to scientific progress" (Huang et al., 2013, p. 401). Tenopir et al. (2011) report "(67%) of the respondents agreed that lack of access to data generated by other researchers or institutions is a major impediment to progress in science" (Tenopir et al., 2011, p. 7). Other scholars argue that data sharing accelerates scientific progress because it helps find *synergies* and avoid repeating work (Wallis et al., 2013). It is also argued that shared data increases quality assurance and makes the *review* process better (Enke et al., 2011), and that it increases the networking and the *exchange* with other researchers (Enke et al., 2011). Wicherts and Bakker (2012) argue that researchers who share data commit less errors and that data sharing encourages more research (secondary analyses).

Publications. In most research disciplines publications are the primary currency. Promotions, grants, and recruitments are often based on publications records. The demands and offers of publication outlets therefore have an impact on the individual researcher's data sharing disposition (Breeze et al., 2012). Enke et al. (2011) describe *journal policies* to be the major motivator for data sharing, even before funding agencies. A study conducted by Huang et al. (2013) shows that 74 % of researchers would accept leading journals' data sharing policies. However, other research indicates that today's journal policies for data sharing are all but binding (Alsheikh-Ali et al., 2011; Noor et al., 2006; Pearce and Smith, 2011; Tenopir et al., 2011). Several scholars argue that more stringent data sharing policies are needed to make researchers share (Huang et al., 2012; Parr, 2007; Pearce and Smith, 2011). At the same time they argue that publications that include or link to the used dataset receive more citations. And therefore both journals and researchers should be incentivized to follow data sharing policies (Parr, 2007).

3.4. Norms

In the category *norms* we subsume all ethical and legal codes that impact a researcher's data sharing behavior.

Table 8: Overview category norms

Sub-category	Factors	References	Exemplary Quotes
Ethical norms	▪ Confidentiality ▪ Informed consent ▪ Potential harm	Axelsson and Schroeder, 2009; Brakewood and Poldrack, 2013; Cooper, 2007; De Wolf et al., 2006; Fernandez et al., 2012; Freymann et al., 2012; Haddow, 2011; Jarnevich et al., 2007; Levenson, 2010; Ludman et al., 2010; Mennes et al., 2013; Pearce and Smith, 2011; Perrino et al., 2013; Resnik, 2010; Rodgers and Nolte, 2006; Sieber, 1988; Tenopir et al., 2011	*Informed consent:* "Autonomous decision-making means that a subject [patient] needs to have the ability to think about his or her choice to participate or not and the ability to actually act on that decision." (Brakewood and Poldrack, 2013, p. 673)
Legal norms	▪ Ownership and right of use ▪ Privacy ▪ Contractual consent ▪ Copyright	Axelsson and Schroeder, 2009; Brakewood and Poldrack, 2013; Breeze et al., 2012; Cahill and Passamo, 2007; Cooper, 2007; Costello, 2009; Chandramohan et al., 2008; Chokshi et al., 2006; Daiglesh et al., 2012; De Wolf et al., 2005; De Wolf et al., 2006; Delson et al., 2007; Eisenberg, 2006; Enke et al., 2011; Freymann et al., 2012; Haddow, 2011; Kowalcyk and Shankar, 2013; Levenson, 2010; Mennes et al., 2013; Nelson, 2009; Perrino et al., 2013; Pitt and Tang, 2013; Rai and Eisenberg, 2006; Reidpath and Allotey, 2001; Resnik, 2010; Rohlfing and Poline, 2012; Teeters et al., 2008; Tenopir et al., 2011	*Ownership and right of use:* "In fact, unresolved legal issues can deter or restrain the development of collaboration, even if scientists are prepared to proceed." (Sayogo and Pardo, 2013, p. 21)

Ethical norms. As ethical norms we regard moral principles of conduct from the data collector's perspective. Brakewood and Poldrack (2013, p. 673) regard the respect for persons as a core principle in data sharing and emphasize the importance of informed consent and *confidentiality*, which is particularly relevant in the context of individual-related data. The authors demand that a patient "needs to have the ability to think about his or her choice to participate or not and the ability to actually act on that decision". De Wolf et al. (2005), Harding et al. (2011), and Mennes et al. (2013) take the same line regarding the necessity of *informed consent* between researcher and study subject. Axelsson and Schroeder (2009) describe the maxim to act upon public trust as an important

precondition for database research. Regarding data sensitivity, Cooper (2007) emphasizes the need to consider if the data being shared could harm people. Similarly, Enke et al. (2011) point to the possibility that some data could be used to *harm* environmentally sensitive areas. Often ethical considerations in the context of data sharing concern adverse use of data, we specify that under adverse use in the category data recipients.

Legal norms. Legal uncertainty can deter data sharing, especially in disciplines that work with sensitive data, example are corporate or personal data (Sayogo and Pardo, 2013). Under legal norms we subsume *ownership and rights of use*, *privacy, contractual consent* and *copyright*. These are the most common legal issues regarding data sharing. The sharing of data is restricted by the national privacy acts. In this regard, Freyman et al. (2012) and Pitt and Tang (2013) emphasize the necessity for de-identification as a pre-condition for sharing individual-related data. Pearce (2011) on the other hand states that getting rid of identifiers is often not enough and pleads for restricted access. Many authors point to the necessity of *contractual consent* between data collector and study participant regarding the terms of use of personal data (e.g., Anderson and Schonfeld, 2009; Cooper, 2007; De Wolf et al., 2005; De Wolf et al., 2006; Haddow, 2011; Ludman et al., 2010; Mennes et al., 2013; Resnik, 2010). While *privacy* issues apply to individual-related data, issues of *ownership and rights of use* concern all kinds of data. Enke et al. (2011) states that the legal framework concerning the ownership of research data before and after deposition in a database is complex and involves many uncertainties that deter data sharing (see also Costello, 2009; Delson et al., 2007). Eisenberg (2006) even regards the absence of adequate intellectual property rights, especially in the case of patent-relevant research, as a barrier for data sharing and therefore innovation (see also Cahill and Passamo, 2007; Milia et al., 2012). Chandramohan et al. (2008) emphasize that data collection financed by tax money is or should be a public good.

3.5. Data Recipients

In the category data recipients, we subsumed influencing factors regarding the use of data by the data recipient and the recipient's organizational context.

Table 9: Overview category data recipients

Sub-category	Factors	References	Exemplary Quotes
Adverse Use	▪ Falsification ▪ Commercial misuse ▪ Competitive misuse ▪ Flawed interpretation ▪ Unclear intent	Acord and Harley, 2012; Anderson and Schonfeld, 2009; Cooper, 2007; Costello, 2009; De Wolf et al., 2006; Enke et al., 2011; Fisher and Fortman, 2010; Gardner et al., 2003; Harding et al., 2011; Hayman et al., 2012; Huang et al., 2012; Huang et al., 2013; Molloy, 2011; Nelson, 2009; Overbey, 1999; Parr and Cummings, 2005; Pearce and Smith, 2011; Perrino et al., 2013; Reidpath and Allotey, 2001; Sieber, 1988; Stanley and Stanley, 1988; Tenopir et al., 2011; Wallis et al., 2013; Zimmerman, 2008	*Falsification:* "I am afraid that I made a mistake somewhere that I didn't find myself and someone else finds." (Survey) *Competitive Use:* "Furthermore I have concerns that I used my data exhaustively before I publish it." (Survey)
Recipient's organization	▪ Data security conditions ▪ Commercial or public organization	Fernandez et al., 2012; Tenopir et al., 2011	*Data security conditions:* "Do the lab facilities of the receiving researcher allow for the proper containment and protection of the data? Do the [...] security policies of the receiving lab/organization adequately reduce the risk that an internal or external party accesses and releases the data in an unauthorized fashion?" (Fernandez et al., 2012, p. 138)

Adverse use. A multitude of hindering factors for data sharing in academia can be assigned to a presumed *adverse use* on the part of the data recipient. In detail, these are *falsification*, *commercial misuse*, *competitive misuse*, *flawed interpretation*, and *unclear intent*. For all of these factors, references can be found in both, the literature and the survey. Regarding the fear of *falsification*, one respondent states: "I am afraid that I made a mistake somewhere that I didn't find myself and someone else finds." In the same line, Costello (2009, p. 421) argues that "[a]uthors may fear that their selective use of data, or possible errors in analysis, may be revealed by data publication". Similarly, Acord and Harley (2012), Costello (2009), Pearce (2011), Sieber (1988), and Stanley and Stanley (1988) describe the fear of falsification as a reason to withhold data. Few authors see a potential "*commercialization* of research findings" (Ludman 2010, p. 6) as a reason not to share data (see also: Anderson and Schonfeld,

2009; Costello, 2009; Gardner et al., 2003). The most frequently mentioned withholding reason regarding the third party use of data is *competitive misuse*; the fear that someone else publishes with *my* data before *I* can (16 survey references, 13 text references). This indicates that at least from the primary researcher's point of view, withholding data is a common competitive strategy in a publication-driven system. Costello (2009, p. 421) encapsulates this issue: "If I release data, then I may be scooped by somebody else producing papers from them." Many other authors in our sample examine the issue of *competitive misuse* (e.g., Acord and Harley, 2012; Daiglesh et al., 2012; Masum et al., 2013; Milia et al., 2012; Teeters et al., 2008; Tenopir et al., 2011). Another issue regarding the recipient's use of data concerns a possible *flawed interpretation* (e.g., Cooper, 2007; Enke et al., 2011; Perrino et al., 2013; Zimmerman, 2008). Perrino et al. (2013, p. 439), regarding a dataset from psychological studies, state: "The correct interpretation of data has been another concern of investigators. This included the possibility that the [data recipient] might not fully understand assessment measures, interventions, and populations being studied and might misinterpret the effect of the intervention". The issue of flawed interpretation is closely related to the factor data documentation (see *3.6 data infrastructure*). Associated with the need for control (as outlined in *data donor*), authors and respondents alike mention a declaration of intent as an enabling factor, as one of the respondents states "missing knowledge regarding the purpose and the recipient" is a reason not to share (see also Sayogo and Pardo, 2013; Stanley and Stanley, 1988).

Recipient's organization. According to Fernandez et al. (2012) the recipient's organization, its type (*commercial or public*) and *data security conditions* have some impact on academic data sharing. Fernandez et al. (2012, p. 138) summarize the potential uncertainties: "Do the lab facilities of the receiving researcher allow for the proper containment and protection of the data? Do the physical, logical and personnel security policies of the receiving lab/organization adequately reduce the risk that [someone] release[s] the data in an unauthorized fashion?" (see also Tenopir et al., 2011).

3.6. Data Infrastructure

In the category data infrastructure we subsume all factors concerning the technical infrastructure to store and retrieve data. It is comprised of the sub-categories *architecture*, *usability*, and *management software*.

Table 10: Overview category data infrastructure

Sub-category	Factors	References	Exemplary Quotes
Architecture	▪ Access ▪ Performance ▪ Storage ▪ Data quality ▪ Data security	Axelsson and Schroeder, 2009; Constable et al., 2010; Cooper, 2007; De Wolf et al., 2006; Enke et al., 2011; Haddow, 2011; Huang et al., 2012; Jarnevich et al., 2007; Kowalcyk and Shankar, 2013; Linkert et al., 2010; Resnik, 2010; Rodgers and Nolte, 2006; Rushby, 2013; Sayogo and Pardo, 2013; Teeters et al., 2008	*Access:* "It is not permitted, for example, for a faculty member to obtain the data for his or her own research project and then "lend" it to a graduate student to do related dissertation research, even if the graduate student is a research staff signatory, unless this use is specifically stated in the research plan." (Rodgers and Nolte, 2006, p. 90)
Usability	▪ Tools and applications ▪ Technical support	Axelsson and Schroeder, 2009; Haendel et al., 2012; Mennes et al., 2013; Nicholson and Bennett, 2011; Ostell, 2009; Teeters et al., 2008	*Tools and applications:* "At the same time, the platform should make it easy for researchers to share data, ideally through a simple one-click upload, with automatic data verification thereafter." (Mennes et al. 2012, p. 688)
Management software	▪ Data documentation ▪ Metadata standards	Acord and Hartley, 2012; Axelsson and Schroeder, 2009; Breese et al., 2012; Constable et al., 2010; Delson et al., 2007; Edwards et al., 2011; Enke et al., 2011; Gardner et al., 2003; Jiang et al., 2013; Jones et al., 2012; Karami et al., 2013; Linkert et al., 2010; Myneni and Patel, 2010; Nelson, 2009; Parr, 2007; Teeters et al., 2008; Tenopir et al., 2011	*Data documentation:* "The authors came to the conclusion that researchers often fail to develop clear, well-annotated datasets to accompany their research (i.e., metadata), and may lose access and understanding of the original dataset over time." (Tenopir et al. 2011, p. 2)

Architecture. A common rationale within the surveyed literature is that restricted access, for example through a registration system, would contribute to *data security* and counter the perceived

loss of control (Cooper, 2007; De Wolf et al., 2006; Resnik, 2010; Rodgers and Nolte, 2006). Some authors even emphasize the necessity to edit the data after it has been *stored* (Enke et al., 2011). There is however disunity if the infrastructure should be centralized (e.g., Linkert et al., 2010) or decentralized (e.g., Constable et al., 2010). Another issue is that *data quality* is maintained after it has been archived (Brakewood and Poldrack, 2013). In this respect, Teeters et al. (2008) suggest that data infrastructure should provide technical support and means of indexing.

Usability. The topic of usability comes up multiple times in the literature. The authors argue that service providers need to make an effort to simplify the sharing process and the involved *tools* (Mennes et al., 2013, Teeters et al., 2008). Authors also argue that guidelines are needed besides a *technical support* that makes it easy for researchers to share (Nicholson and Bennett, 2011).

Management system. We found 24 references for the *management system* of the *data infrastructure*; these were concerned with *data documentation* and *metadata standards* (Axelsson and Schroeder, 2009; Linkert et al., 2010; Tenopir et al., 2011). The *documentation* of data remains a troubling issue and many disciplines complain about missing standards (Acord and Harley, 2012). At the same time other authors explain that detailed *metadata* is needed to prevent misinterpretation (Gardner et al., 2003).

4. Discussion and Conclusion

The accessibility of research data holds great potential for scientific progress. It allows the verification of study results and the reuse of data in new contexts (Dewald et al., 1986; McCullough, 2009). Despite its potential and prominent support, sharing data is not yet common practice in academia. With the present paper we explain that data sharing in academia is a multidimensional effort that includes a diverse set of stakeholders, entities and individual interests. To our knowledge, there is no overarching framework, which puts the involved parties and interests in relation to one another. In our view, the conceptual framework with its empirically revised categories has theoretical and practical use. In the remaining discussion we will elaborate possible implications for theory, and research practice. We will further address the need for future research.

4.1. Theoretical Implications: Data is Not a Knowledge Commons

Concepts for the production of immaterial goods in a networked society frequently involve the dissolution of formal entities, modularization of tasks, intrinsic motivation of participants, and the absence of ownership. Benkler's (2006) *commons-based peer production,* to certain degree *wisdom of the crowds* (Surowiecki, 2004) and *collective intelligence* (Lévy, 1997) are examples for organizational theories that embrace novel forms of networked collaboration. Frequently mentioned empirical cases for these forms of collaboration are the open source software community (von Hippel and von Krogh, 2003) or the online encyclopedia Wikipedia (Benkler, 2006; Kittur and Kraut, 2008). In both cases, the product of the collaboration can be considered a commons. The production process is inherently inclusive.

In many respects, Franzoni and Sauermann's (2014) theory for *crowd science* resembles the concepts commons-based peer production and crowd intelligence. The authors dissociate crowd science from traditional science, which they describe as largely closed. In traditional science, researchers retain exclusive use of key intermediate inputs, such as data. Crowd science on the other hand is characterized by its inherent openness with respect to participation and the disclosure of intermediate inputs such as data. Following that line of thought, data could be as well considered a knowledge commons; a good that can be accessed by everyone and whose consumption is non-rivalry (Hess and Ostrom, 2006). Crowd-science is in that regard a commons-driven principle for scholarly knowledge. In many respects, academia indeed fulfills the requirements for crowd science, be it the immateriality of knowledge products, the modularity of research, and public interest.

In the case of data sharing in academia, however, the theoretical depiction of a crowd science (Franzoni and Sauermann, 2014) or an open science (Nielsen, 2011), and with both the accessibility of data, does not meet the empirical reality. The core difference to the model of a commons-based peer production like we see it in open source software or crowd science lies in the motivation for participation. Programmers do not have to release their code under an open source license. And many do not. The same is true for Wikipedia, where a rather small but active community edits pages. Both systems run on voluntariness, self-organization, and intrinsic motivation. Academia however, contradicts to different degrees these characteristics of a knowledge exchange system.

In an ideal situation every researcher would publish data alongside a research paper to make sure the results are reproducible and the data is reusable in a new context. Yet today, most researchers remain reserved to share their research data. This indicates that their efforts and

perceived risks outweigh the potential individual benefits they expect from data sharing. Research data is in large parts not a knowledge commons. Instead, our results points to a *perceived ownership of data* (reflected in the right to publish first) and a *need for control* (reflected in the fear of data misuse). Both impede a commons-based exchange of research data. When it comes to data, academia remains neither accessible nor participatory. As data publications lack sufficient *formal recognition* (e.g., citations, co-authorship) in comparison to text publications, researchers find furthermore too few incentives to share data. While altruism and with it the idea to contribute to a common good, is a sufficient driver for some researchers, the majority currently remains incentivized not to share. If data sharing leads to better science and simultaneously, researchers are hesitant to share, the question arises how research policies can foster data sharing among academic researchers.

4.2. Policy Implications: Towards more Data Sharing

Worldwide, research policy makers support the accessibility of research data. This can be seen in the US with efforts by the National Institutes of Health (NIH, 2002; NIH, 2003) and also in Europe, with the EU's Horizon 2020 programme (European Commission, 2012). In order to develop consequential policies for data sharing, policy makers need to understand and address the involved parties and their perspectives. The framework that we present in this paper helps to gain a better understanding of the prevailing issues and provides insights into the underlying dynamics of academic data sharing. Considering that research data is far from being a commons, we believe that research policies should work towards an efficient exchange system in which as much data is shared as possible. Strategic policy measures could therefore go into two directions: First, they could provide incentives for sharing data and second impede researchers not to share. Possible incentives could include adequate *formal recognition* in the form of data citation and career prospects. In academia, a largely non-monetary knowledge exchange system, research policy should be geared towards making intermediate products count more. Furthermore could forms of *financial reimbursement*, for example through additional person hours in funding, help to increase the individual effort to make data available. As long as academia remains a publication-centered business, journal policies further need to adopt mandatory data sharing policies (McCullough, 2009; Savage and Vickers, 2009) and provide easy-to use data management systems. Impediments regarding sharing supplementary data could include clear and elaborate reasons to opt out. In order to remove risk aversion and ambiguity, an understandable and clear legal basis regarding the rights of use is needed

to inform researchers on what they can and cannot do with data they collected. This is especially important in medicine and in the social sciences where much data comes from individuals. Clear guidelines that explain how consent can be obtained and how data can be anonymized are needed. Educational efforts within data-driven research fields on proper data curation, sharing culture, data documentation, and security could be fruitful in the intermediate-term. Ideally these become part of the curriculum for university students (Piwowar et al., 2008). Infrastructure investments are needed to develop efficient and easy-to-use data repositories and data management software, for instance as part of the Horizon 2020 research infrastructure endeavors.

4.3. Future Research

We believe that more research needs to address the discipline-specific barriers and enablers for data sharing in academia in order to make informed policy decisions. Regarding the framework that we introduce in this paper, the identified factors need further empirical revision. In particular, we regard the intersection between academia and industry worth investigating. For instance: A study among German life scientists showed that those who receive industry funding are more likely to deny others' requests for access to research materials (Czarnitzki et. al, 2014). In the same line, Haeussler (2010), in a comparative study on information sharing among scientists, finds that the likelihood of sharing information decreases with the competitive value of the requested information. It increases when the inquirer is an academic researcher. Following this, future research could address data sharing between industry and academia. Open enterprise data, for example, appears to be a relevant topic for legal scholars as well as innovation research.

References

Acord, S.K., Harley, D., 2012. Credit, time, and personality: The human challenges to sharing scholarly work using Web 2.0. New Media & Society 15, 379–397. doi:10.1177/1461444812465140

Albert, H., 2012. Scientists encourage genetic data sharing. SH News 1, 1–2. doi:10.1007/s40014-012-1434-z

Allen, L., et al., 2014. Credit where credit is due. Nature 508. 312-313.

Alsheikh-Ali, A.A., Qureshi, W., Al-Mallah, M.H., Ioannidis, J.P.A., 2011. Public Availability of Published Research Data in High-Impact Journals. PLoS ONE 6, e24357. doi:10.1371/journal.pone.0024357

Anagnostou, P., Capocasa, M., Milia, N., Bisol, G.D., 2013. Research data sharing: Lessons from forensic genetics. Forensic Science International: Genetics. doi:10.1016/j.fsigen.2013.07.012

Anderson, J.R., Toby L. Schonfeld, 2009. Data-Sharing Dilemmas: Allowing Pharmaceutical Company Access to Research Data. IRB: Ethics and Human Research 31, 17–19. doi:10.2307/25594876

Axelsson, A.-S., Schroeder, R., 2009. Making it Open and Keeping it Safe: e-Enabled Data-Sharing in Sweden. Acta Sociologica 52, 213–226. doi:10.2307/25652127

Belmonte, M.K., Mazziotta, J.C., Minshew, N.J., Evans, A.C., Courchesne, E., Dager, S.R., Bookheimer, S.Y., Aylward, E.H., Amaral, D.G., Cantor, R.M., Chugani, D.C., Dale, A.M., Davatzikos, C., Gerig, G., Herbert, M.R., Lainhart, J.E., Murphy, D.G., Piven, J., Reiss, A.L., Schultz, R.T., Zeffiro, T.A., Levi-Pearl, S., Lajonchere, C., Colamarino, S.A., 2007. Offering to Share: How to Put Heads Together in Autism Neuroimaging. Journal of Autism and Developmental Disorders 38, 2–13. doi:10.1007/s10803-006-0352-2

Benkler, Y., 2002. Coase's Penguin, or, Linux and "The Nature of the Firm". The Yale Law Journal 112, 369. doi:10.2307/1562247

Bezuidenhout, L., 2013. Data Sharing and Dual-Use Issues. Sci Eng Ethics 19, 83–92. doi:10.1007/s11948-011-9298-7

Booth, A., Papaioannou, D., Sutton, A., 2012. Systematic approaches to a successful literature review. Sage, Los Angeles; Thousand Oaks, Calif.

Borgman, C.L., 2007. Scholarship in the digital age: information, infrastructure, and the Internet. MIT Press, Cambridge, Mass.

Borgman, C.L., 2012. The conundrum of sharing research data. Journal of the American Society for Information Science and Technology 63, 1059–1078. doi:10.1002/asi.22634

Breeze, J., Poline, J.-B., Kennedy, D., 2012. Data sharing and publishing in the field of neuroimaging. GigaSci 1, 1–3. doi:10.1186/2047-217X-1-9

Butler, D., 2007. Data sharing: the next generation. Nature 446, 10–11. doi:10.1038/446010b

Cahill, J.M., Passamano, J.A., 2007. Full Disclosure Matters. Near Eastern Archaeology 70, 194–196. doi:10.2307/20361332

Campbell, E.G., Clarridge, B.R., Gokhale, M., Birenbaum, L., Hilgartner, S., Holtzman, N.A., Blumenthal, D., 2002. Data withholding in academic genetics: evidence from a national survey. JAMA 287, 473–480.

Chandramohan, D., Shibuya, K., Setel, P., Cairncross, S., Lopez, A.D., Murray, C.J.L., Żaba, B., Snow, R.W., Binka, F., 2008. Should Data from Demographic Surveillance Systems Be Made More Widely Available to Researchers? PLoS Medicine 5, e57. doi:10.1371/journal.pmed.0050057

Chokshi, D.A., Parker, M., Kwiatkowski, D.P., 2006. Data sharing and intellectual property in a genomic epidemiology network: policies for large-scale research collaboration. Bulletin of the World Health Organization 84.

Cochrane Collaboration, 2008. Cochrane handbook for systematic reviews of interventions, Cochrane book series. Wiley-Blackwell, Chichester, England ; Hoboken, NJ.

Cohn, J.P., 2012. DataONE Opens Doors to Scientists across Disciplines. BioScience 62, 1004. doi:10.1525/bio.2012.62.11.16

Constable, H., Guralnick, R., Wieczorek, J., Spencer, C., Peterson, A.T., The VertNet Steering Committee, 2010. VertNet: A New Model for Biodiversity Data Sharing. PLoS Biology 8, e1000309. doi:10.1371/journal.pbio.1000309

Cooper, M., 2007. Sharing Data and Results in Ethnographic Research: Why This Should not be an Ethical Imperative. Journal of Empirical Research on Human Research Ethics: An International Journal 2, 3–19.

Costello, M.J., 2009. Motivating Online Publication of Data. BioScience 59, 418–427.

Cragin, M.H., Palmer, C.L., Carlson, J.R., Witt, M., 2010. Data sharing, small science and institutional repositories. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 368, 4023–4038. doi:10.1098/rsta.2010.0165

Czarnitzki, D., Grimpe, C., Pellens, M., 2014. Access to Research Inputs: Open Science Versus the Entrepreneurial University. SSRN Electronic Journal. doi:10.2139/ssrn.2407437

Dahlander, L., Gann, D.M., 2010a. How open is innovation? Research Policy 39, 699–709. doi:10.1016/j.respol.2010.01.013

Dahlander, L., Gann, D.M., 2010b. How open is innovation? Research Policy 39, 699–709. doi:10.1016/j.respol.2010.01.013

Dalgleish, R., Molero, E., Kidd, R., Jansen, M., Past, D., Robl, A., Mons, B., Diaz, C., Mons, A., Brookes, A.J., 2012. Solving bottlenecks in data sharing in the life sciences. Human Mutation 33, 1494–1496. doi:10.1002/humu.22123

Davies, P., 2000. The Relevance of Systematic Reviews to Educational Policy and Practice. English 26, 365 – 378.

De Wolf, V.A., Sieber, J.E., Steel, P.M., Zarate, A.O., 2005. Part I: What Is the Requirement for Data Sharing? IRB: Ethics and Human Research 27, 12–16.

De Wolf, V.A., Sieber, J.E., Steel, P.M., Zarate, A.O., 2006. Part III: Meeting the Challenge When Data Sharing Is Required. IRB: Ethics and Human Research 28, 10–15.

Delson, E., Harcourt-Smith, W.E.H., Frost, S.R., Norris, C.A., 2007. Databases, Data Access, and Data Sharing in Paleoanthropology: First Steps. Evolutionary Anthropology 16, 161–163.

Dewald, W., Thursby, J.G., Anderson, R.G., 1986. Replication in empirical economics: The Journal of Money, Credit and Banking Project. American Economic Review 76, 587–603.

Drew, B.T., Gazis, R., Cabezas, P., Swithers, K.S., Deng, J., Rodriguez, R., Katz, L.A., Crandall, K.A., Hibbett, D.S., Soltis, D.E., 2013. Lost Branches on the Tree of Life. PLoS Biology 11, e1001636. doi:10.1371/journal.pbio.1001636

Duke, C.S., 2006. Data: Share and Share Alike. Frontiers in Ecology and the Environment 4, 395.

Edwards, P.N., Mayernik, M.S., Batcheller, A.L., Bowker, G.C., Borgman, C.L., 2011. Science friction: Data, metadata, and collaboration. Social Studies of Science 41, 667–690. doi:10.1177/0306312711413314

Eisenberg, R.S., 2006. Patents and data-sharing in public science. Industrial and Corporate Change 15, 1013–1031. doi:10.1093/icc/dtl025

Elman, C., Kapiszewski, D., Vinuela, L., 2010. Qualitative Data Archiving: Rewards and Challenges. PS: Political Science & Politics 43, 23–27. doi:10.1017/S104909651099077X

Eschenfelder, K., Johnson, A., 2011. The Limits of sharing: Controlled data collections. Proceedings of the American Society for Information Science and Technology 48, 1–10. doi:10.1002/meet.2011.14504801062

European Commission, 2012a. Scientific data: open access to research results will boost Europe's innovation capacity. http://europa.eu/rapid/press-release_IP-12-790_en.htm.

European Commission, 2012b. Scientific data: open access to research results will boost Europe's innovation capacity. European Commission Press Release.

Feinberg, S.E., Martin, M.E., Straf, M.L., 1985. Sharing Research Data. National Academy Press, Washington, DC.

Feldman, L., Patel, D., Ortmann, L., Robinson, K., Popovic, T., 2012. Educating for the future: another important benefit of data sharing. The Lancet 379, 1877–1878. doi:10.1016/S0140-6736(12)60809-5

Fernandez, J., Patrick, A., Zuck, L., 2012. Ethical and Secure Data Sharing across Borders, in: Blyth, J., Dietrich, S., Camp, L.J. (Eds.), Financial Cryptography and Data Security, Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 136–140.

Foley, P., Alfonso, X., Sakka, M.A., 2006. Information sharing for social inclusion in England: A review of activities, barriers and future directions. Journal of Information, Communication and Ethics in Society 4, 191–203. doi:10.1108/14779960680000292

Franzoni, C., Sauermann, H., 2014. Crowd science: The organization of scientific research in open collaborative projects. Research Policy 43, 1–20. doi:10.1016/j.respol.2013.07.005

Freymann, J., Kirby, J., Perry, J., Clunie, D., Jaffe, C.C., 2012. Image Data Sharing for Biomedical Research—Meeting HIPAA Requirements for De-identification. J Digit Imaging 25, 14–24. doi:10.1007/s10278-011-9422-x

Fulk, J., Heino, R., Flanagin, A.J., Monge, P.R., Bar, F., 2004. A Test of the Individual Action Model for Organizational Information Commons. Organization Science 15, 569–585. doi:10.2307/30034758

Gardner, D., Toga, A.W., et al., 2003. Towards Effective and Rewarding Data Sharing. Neuroinformatics 1, 289–295.

Given, L.M., 2008. The Sage encyclopedia of qualitative research methods. Sage Publications, Thousand Oaks, Calif.

Guralnick, R., Constable, H., 2010. VertNet: Creating a Data-sharing Community. BioScience 60, 258–259. doi:10.1525/bio.2010.60.4.2

Haddow, G., Bruce, A., Sathanandam, S., Wyatt, J.C., 2011. "Nothing is really safe": a focus group study on the processes of anonymizing and sharing of health data for research purposes. Journal of Evaluation in Clinical Practice 17, 1140–1146. doi:10.1111/j.1365-2753.2010.01488.x

Haendel, M.A., Vasilevsky, N.A., Wirz, J.A., 2012. Dealing with Data: A Case Study on Information and Data Management Literacy. PLoS Biology 10, e1001339. doi:10.1371/journal.pbio.1001339

Haeussler, C., 2011. Information-sharing in academia and the industry: A comparative study. Research Policy 40, 105–122. doi:10.1016/j.respol.2010.08.007

Hammersley, M., 2001. On "Systematic" Reviews of Research Literatures: A "narrative" response to Evans & Benefield. British Educational Research Journal 27, 543–554. doi:10.1080/01411920120095726

Harding, A., Harper, B., Stone, D., O'Neill, C., Berger, P., Harris, S., Donatuto, J., 2011. Conducting Research with Tribal Communities: Sovereignty, Ethics, and Data-Sharing Issues. Environmental Health Perspectives 120, 6–10. doi:10.1289/ehp.1103904

Hayman, B., Wilkes, L., Jackson, D., Halcomb, E., 2012. Story-sharing as a method of data collection in qualitative research. Journal of Clinical Nursing 21, 285–287. doi:10.1111/j.1365-2702.2011.04002.x

Hess, C., Ostrom, E. (Eds.), 2007. Understanding knowledge as a commons: from theory to practice. MIT Press, Cambridge, Mass.

Higging, J.P.., Green, S., n.d. Cochrane handbook for systematic reviews of interventions, Cochrane Book Series. Wiley-Blackwell, Chichester, England ; Hoboken, NJ.

Hippel, E. von, Krogh, G. von, 2003. Open Source Software and the "Private-Collective" Innovation Model: Issues for Organization Science. Organization Science 14, 209–223. doi:10.1287/orsc.14.2.209.14992

Huang, X., Hawkins, B.A., Lei, F., Miller, G.L., Favret, C., Zhang, R., Qiao, G., 2012. Willing or unwilling to share primary biodiversity data: results and implications of an international survey. Conservation Letters 5, 399–406. doi:10.1111/j.1755-263X.2012.00259.x

Jarnevich, C., Graham, J., Newman, G., Crall, A., Stohlgren, T., 2007. Balancing data sharing requirements for analyses with data sensitivity. Biol Invasions 9, 597–599. doi:10.1007/s10530-006-9042-4

Jiang, X., Sarwate, A.D., Ohno-Machado, L., 2013. Privacy Technology to Support Data Sharing for Comparative Effectiveness Research: A Systematic Review. Medical Care 51, S58–S65. doi:10.1097/MLR.0b013e31829b1d10

Jones, R., Reeves, D., Martinez, C., 2012. Overview of Electronic Data Sharing: Why, How, and Impact. Curr Oncol Rep 14, 486–493. doi:10.1007/s11912-012-0271-7

Karami, M., Rangzan, K., Saberi, A., 2013. Using GIS servers and interactive maps in spectral data sharing and administration: Case study of Ahvaz Spectral Geodatabase Platform (ASGP). Computers & Geosciences 60, 23–33. doi:10.1016/j.cageo.2013.06.007

Kim, Y., Stanton, J.M., 2012. Institutional and Individual Influences on Scientists' Data Sharing Practices. Journal of Computational Science Education 3, 47–56.

Kittur, A., Kraut, R.E., 2008. Harnessing the wisdom of crowds in wikipedia: quality through coordination. ACM Press, p. 37. doi:10.1145/1460563.1460572

Knowledge Exchange, 2013. Knowledge Exchange Annual Report and Briefing 2013 (Annual Report). Copenhagen, Denmark.

Levenson, D., 2010. When should pediatric biobanks share data? American Journal of Medical Genetics Part A 152A, fm vii–fm viii. doi:10.1002/ajmg.a.33287

Lévy, P., 1997. Collective intelligence: mankind's emerging world in cyberspace. Perseus Books, Cambridge, Mass.

Linkert, M., Curtis T., Burel, J.-M., Moore, W., Patterson, A., Loranger, B., Moore, J., Neves, C., MacDonald, D., Tarkowska, A., Sticco, C., Hill, E., Rossner, M., Eliceiri, K.W., Swedlow, J.R., 2010. Metadata matters: access to image data in the real world. Metadata matters: access to image data in the real world 189, 777–782.

Ludman, E.J., Fullerton, S.M., Spangler, L., Trinidad, S.B., Fujii, M.M., Jarvik, G.P., Larson, E.B., Burke, W., 2010. Glad You Asked: Participants' Opinions Of Re-Consent for dbGap Data Submission. Journal of Empirical Research on Human Research Ethics 5, 9–16. doi:10.1525/jer.2010.5.3.9

Masum, H., Rao, A., Good, B.M., Todd, M.H., Edwards, A.M., Chan, L., Bunin, B.A., Su, A.I., Thomas, Z., Bourne, P.E., 2013. Ten Simple Rules for Cultivating Open Science and Collaborative R&D. PLoS Computational Biology 9, e1003244. doi:10.1371/journal.pcbi.1003244

Mayring, P., 2000. Qualitative Content Analysis. Forum: Qualitative Social Research 1.

McCullough, B.D., 2009. Open Access Economics Journals and the Market for Reproducible Economic Research. Economic Analysis & Policy 39, 118–126.

Mennes, M., Biswal, B.B., Castellanos, F.X., Milham, M.P., 2013. Making data sharing work: The FCP/INDI experience. NeuroImage 82, 683–691. doi:10.1016/j.neuroimage.2012.10.064

Milia, N., Congiu, A., Anagnostou, P., Montinaro, F., Capocasa, M., Sanna, E., Bisol, G.D., 2012. Mine, Yours, Ours? Sharing Data on Human Genetic Variation. PLoS ONE 7, e37552. doi:10.1371/journal.pone.0037552

Molloy, J.C., 2011. The Open Knowledge Foundation: Open Data Means Better Science. PLoS Biology 9, e1001195. doi:10.1371/journal.pbio.1001195

Myneni, S., Patel, V.L., 2010. Organization of biomedical data for collaborative scientific research: A research information management system. International Journal of Information Management 30, 256–264. doi:10.1016/j.ijinfomgt.2009.09.005

Nelson, B., 2009. Data sharing: Empty archives. Nature 461, 160–163. doi:10.1038/461160a

Nicholson, S.W., Bennett, T.B., 2011. Data Sharing: Academic Libraries and the Scholarly Enterprise. portal: Libraries and the Academy 11, 505–516.

Nielsen, M.A., 2012. Reinventing discovery: the new era of networked science. Princeton University Press, Princeton, N.J.

NIH, 2002. NIH Data-Sharing Plan. Anthropology News 43, 30–30. doi:10.1111/an.2002.43.4.30.2

NIH, 2003. NIH Posts Research Data Sharing Policy. Anthropology News 44, 25–25. doi:10.1111/an.2003.44.4.25.2

Noor, M.A.F., Zimmerman, K.J., Teeter, K.C., 2006. Data Sharing: How Much Doesn't Get Submitted to GenBank? PLoS Biology 4, e228. doi:10.1371/journal.pbio.0040228

Ostell, J., 2009. Data Sharing: Standards for Bioinformatic Cross-Talk. Human Mutation 30, vii–vii. doi:10.1002/humu.21013

Overbey, M.M., 1999. Data Sharing Dilemma. Anthropology Newsletter April.

Parr, C., Cummings, M., 2005. Data sharing in ecology and evolution. Trends in Ecology & Evolution 20, 362–363. doi:10.1016/j.tree.2005.04.023

Parr, C.S., 2007. Open Sourcing Ecological Data. BioScience 57, 309. doi:10.1641/B570402

Patton, M.Q., 2002. Qualitative research and evaluation methods, 3 ed. ed. Sage Publications, Thousand Oaks, Calif.

Pearce, N., Smith, A., 2011. Data sharing: not as simple as it seems. Environ Health 10, 1–7. doi:10.1186/1476-069X-10-107.

Perrino, T., Howe, G., Sperling, A., Beardslee, W., Sandler, I., Shern, D., Pantin, H., Kaupert, S., Cano, N., Cruden, G., Bandiera, F., Brown, H., 2013. Advancing Science Through Collaborative Data Sharing and Synthesis. Perspectives on Psychological Science 8, 433–444.

Pitt, M.A., Tang, Y., 2013. What Should Be the Data Sharing Policy of Cognitive Science? Topics in Cognitive Science 5, 214–221.

Piwowar, H.A., 2010. Who shares? Who doesn't? Bibliometric factors associated with open archiving of biomedical datasets. ASIST 2010, October 22–27, 2010, Pittsburgh, PA, USA. 1–2.

Piwowar, H.A., 2011. Who Shares? Who Doesn't? Factors Associated with Openly Archiving Raw Research Data. PLoS ONE 6, e18657. doi:10.1371/journal.pone.0018657

Piwowar, H.A., Becich, M.J., Bilofsky, H., Crowley, R.S., on behalf of the caBIG Data Sharing and Intellectual Capital Workspace, 2008. Towards a Data Sharing Culture: Recommendations for Leadership from Academic Health Centers. PLoS Medicine 5, e183. doi:10.1371/journal.pmed.0050183

Piwowar, H.A., Chapman, W.W., 2010. Public sharing of research datasets: A pilot study of associations. Journal of Informetrics 4, 148–156. doi:10.1016/j.joi.2009.11.010

Piwowar, H.A., Day, R.S., Fridsma, D.B., 2007. Sharing Detailed Research Data Is Associated with Increased Citation Rate. PLoS ONE 2, e308. doi:10.1371/journal.pone.0000308

Rai, A.K., Eisenberg, R.S., 2006. Harnessing and Sharing the Benefits of State-Sponsored Research: Intellectual Property Rights and Data Sharing in California's Stem Cell Initiative. Duke Law School Science, Technology and Innovation Research Paper Series 1187–1214.

Richter, D., Metzing, M., Weinhardt, M., Schupp, J., 2013. SOEP scales manual, SOEP Survey Papers / Series C - Data Documentations (Datendokumentationen). Berlin.

Rohlfing, T., Poline, J.-B., 2012. Why shared data should not be acknowledged on the author byline. NeuroImage 59, 4189–4195. doi:10.1016/j.neuroimage.2011.09.080

Rushby, N., 2013. Editorial: Data-sharing. British Journal of Educational Technology 44, 675–676. doi:10.1111/bjet.12091

Samson, K., 2008. NerveCenter: Data sharing: Making headway in a competitive research milieu. Annals of Neurology 64, A13–A16. doi:10.1002/ana.21478

Sansone, S.-A., Rocca-Serra, P., 2012. On the evolving portfolio of community-standards and data sharing policies: turning challenges into new opportunities. GigaSci 1, 1–3. doi:10.1186/2047-217X-1-10

Sarathy, R., Muralidhar, K., 2006. Secure and useful data sharing. Decision Support Systems 42, 204–220. doi:10.1016/j.dss.2004.10.013

Savage, C.J., Vickers, A.J., 2009. Empirical Study of Data Sharing by Authors Publishing in PLoS Journals. PLoS ONE 4, e7078. doi:10.1371/journal.pone.0007078

Sayogo, D.S., Pardo, T.A., 2013. Exploring the determinants of scientific data sharing: Understanding the motivation to publish research data. Government Information Quarterly 30, S19–S31. doi:10.1016/j.giq.2012.06.011

Schreier, M., 2012. Qualitative content analysis in practice. Sage Publications, London; Thousand Oaks, Calif.

Sieber, J.E., 1988. Data sharing: Defining problems and seeking solutions. Law and Human Behavior 12, 199–206. doi:10.1007/BF01073128

Silva, L., 2014. PLOS' New Data Policy: Public Access to Data. EverONE PLOS ONE community blog.

Stanley, B., Stanley, M., n.d. Data Sharing: The Primary Researcher's Perspective. Law and Human Behavior 12, 173–180.

Surowiecki, J., 2004. The wisdom of crowds: why the many are smarter than the few and how collective wisdom shapes business, economies, societies, and nations, 1st ed. ed. Doubleday, New York.

Teeters, J.L., Harris, K.D., Millman, K.J., Olshausen, B.A., Sommer, F.T., 2008. Data Sharing for Computational Neuroscience. Neuroinformatics 6, 47–55. doi:10.1007/s12021-008-9009-y

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A.U., Wu, L., Read, E., Manoff, M., Frame, M., 2011. Data Sharing by Scientists: Practices and Perceptions. PLoS ONE 6, e21101. doi:10.1371/journal.pone.0021101

Torgerson, C.J., 2006. Publication Bias: The Achilles' Heel of Systematic Reviews? British Journal of Educational Studies 54, 89–102. doi:10.1111/j.1467-8527.2006.00332.x

Tucker, J., 2009. Motivating Subjects: Data Sharing in Cancer Research, Dissertation submitted to the Faculty of the Virginia Polytechnic Institute and State University.

Van Horn, J.D., Gazzaniga, M.S., 2013. Why share data? Lessons learned from the fMRIDC. NeuroImage 82, 677–682. doi:10.1016/j.neuroimage.2012.11.010

Vlaeminck, S., Wagner, G. G., 2014. On the role of research data centres in the management of publication-related research data – Results of a survey among scientific infrastructure service providers in the field of social sciences, in: LIBER Quarterly 23, 336–357.

Wagner, G. G., Frick, J.R., Schupp, J., 2007. The German Socio-Economic Panel Study (SOEP) - Evolution, Scope and Enhancements. Schmollers Jahrbuch 127, 139–169.

Wallis, J.C., Rolando, E., Borgman, C.L., 2013. If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology. PLoS ONE 8, e67332. doi:10.1371/journal.pone.0067332

Weber, N.M., 2013. The relevance of research data sharing and reuse studies. Bulletin of the American Society for Information Science and Technology 39, 23–26. doi:10.1002/bult.2013.1720390609

Whitlock, M.C., 2011. Data archiving in ecology and evolution: best practices. Trends in Ecology & Evolution 26, 61–65. doi:10.1016/j.tree.2010.11.006

Whitlock, M.C., McPeek, M.A., Rausher, M.D., Rieseberg, L., Moore, A.J., 2010. Data Archiving. The American Naturalist 175, 145–146. doi:10.1086/650340

Wicherts, J.M., Bakker, M., 2012. Publish (your data) or (let the data) perish! Why not publish your data too? Intelligence 40, 73–76. doi:10.1016/j.intell.2012.01.004

Wicherts, J.M., Bakker, M., Molenaar, D., 2011. Willingness to Share Research Data Is Related to the Strength of the Evidence and the Quality of Reporting of Statistical Results. PLoS ONE 6, e26828. doi:10.1371/journal.pone.0026828

Wicherts, J.M., Borsboom, D., Kats, J., Molenaar, D., 2006. The poor availability of psychological research data for reanalysis. American Psychologist 61, 726–728. doi:10.1037/0003-066X.61.7.726