

Data Sharing in Psychology: A Survey on Barriers and Preconditions

Bobby Lee Houtkoop¹, Chris Chambers², Malcolm Macleod³, Dorothy Bishop⁴,
Thomas Nichols⁴, & Eric-Jan Wagenmakers¹

1 University of Amsterdam

2 Cardiff University

3 University of Edinburgh

4 University of Oxford

Word count: 7,913 (excluding online appendix and references)

Address for correspondence:

Nieuwe Achtergracht 129B,

1018 WV, Amsterdam

The Netherlands

Abstract

Despite its potential to accelerate academic progress in psychological science, public data sharing remains relatively uncommon. In order to discover the perceived barriers to public data sharing and possible measures for lowering them, we conducted a survey with responses from 600 authors of articles in psychology. The results confirmed that data are shared only infrequently. Perceived barriers included that (1) sharing is not a common practice in the field; (2) sharing is preferred upon request; (3) sharing requires extra work; and (4) sharing requires new skills. In order to overcome these barriers, our survey suggests that strong encouragement from institutions, journals, and funders is particularly effective, combined with the development of educational material that demonstrates where and how data can be shared effectively.

Keywords: Public data sharing; open science; open practices; data availability.

Introduction

In every empirical discipline, scientific progress rests upon the availability of research data. The public sharing of such data holds great potential for scientific progress (Ceci & Walker, 1983; Fecher, Friesike & Hebing, 2015). Nevertheless, in many disciplines -- including psychology-- data sharing is still relatively rare (e.g., Wolins, 1962; Ceci, 1988; Wicherts, Borsboom & Molenaar, 2006; Reidpath & Allotey, 2001; Savage & Vickers, 2009; Vanpaemel, Vermorgen, Deriemaeker & Storms, 2015). In this article, we report the outcome of a survey designed to discover why researchers are reluctant to share data, and what can be done to make data sharing more attractive. Throughout this article and the corresponding survey, *data sharing* is defined as: “the activity of making primary research data available in an online repository upon publication of the article associated with the data”¹.

In order to put our survey into context, we first outline the recent initiatives that journals, funders, and researchers have developed in order to encourage the public sharing of primary research data. This list of initiatives underscores the fact that public data sharing is increasingly being recognized as an important component of the scientific process. Nevertheless it remains unclear how successful these initiatives will likely be, how their effectiveness can be increased, and how well these initiatives resonate with researchers.

¹ This definition does not mention the extent to which the online repository is publicly accessible. As is the case for the TOP guidelines, we assumed the online repository to be publicly accessible. However, for data that cannot be fully de-identified it may be necessary to impose restricted access. See also Gilmore, R., Kennedy, J., & Adolph, K. (this issue).

Recent Initiatives to Encourage Data Sharing

Initiatives from Journals. Academic journals may promote data sharing either by mandating it (i.e., stipulating sharing as a condition for publication) or by encouraging it (e.g., referring authors to appropriate sharing tools) (Alsheikh-Ali, Qureshi, Al-Mallah & Ioannidis, 2011).

A prominent example of stringent sharing policies are those employed by the collective journals of *The Royal Society*. To publish in a *Royal Society* journal, authors must make data, code, and research materials publicly available in appropriate and recognized repositories (The Royal Society, n.d.). At the Public Library of Science (PLOS), a similar data sharing policy applies to all journals (PLOS, n.d.).

Stringent journal policies appear to be effective. It may not come as a surprise that mandatory archiving policies substantially increase the probability of actually finding the data in an online archive (Vines et al., 2013). In addition, authors are more likely to share their data if their article was published in a high-impact journal (Piwowar and Chapman, 2010), possibly because high-impact journals often have more stringent sharing policies (Sturges et al., 2015).





To further encourage data sharing across a wide range of scientific journals, several initiatives have emerged. A prominent example is *the badges project*. Initiated by the Center for Open Science (COS), the badges project allows journals to incentivize open science practices by explicitly acknowledging them as such (Open Science Framework, n.d.). Journals can award badges to articles that feature different kinds of open practices (Open Science Framework, n.d.). For example, an Open Data badge is earned when researchers make publicly available the data that are necessary to reproduce the results as reported in the

1
2
3 publication. In terms of effectiveness, Kidwell et al. (2016) found that after the premier
4 journal *Psychological Science* had adopted the open data badges, data sharing rates increased
5 dramatically. Over the same time period, comparison journals showed little or no change (see
6
7 also Giofr , Cumming, Fresc, Boedker, and Tressoldi, 2017).
8
9

10
11 In addition to the badges project, COS developed another initiative to promote an open
12 research culture: the Transparency and Openness Promotion guidelines (TOP; Nosek et al.,
13 2015). The TOP guidelines provide concrete journal policies about eight transparency
14 standards, one of which is data sharing. Not every standard applies equally to all journals and
15 disciplines, and therefore the TOP guidelines come in three levels of strictness. The first level
16 is the most lenient; data-sharing is optional for authors, and cost-free for journals (i.e.,
17 “Article states whether data are available and, if so, where to access them”). The second level
18 is more strict: data sharing is mandatory for authors, but virtually cost-free for journals (i.e.,
19 “Data must be posted to a trusted repository. Exceptions must be identified at article
20 submission”). The third level of implementation is the most strict: data sharing is mandatory
21 for authors, and may entail costs in time and effort for journals (“Data must be posted to a
22 trusted repository, and reported analyses will be reproduced independently before
23 publication”).
24
25

26
27 The badges project and TOP guidelines are increasingly being adopted. Moreover, a
28 growing number of journals have outlined their own policies regarding data sharing. An
29 overview of these policies is provided in Box 1.
30
31

32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49 *Box 1. Data sharing guidelines of select journals with a clearly articulated data sharing policy.*
50

Journal	Data Sharing Policy
	<p>“Supporting data must be made available to editors and peer reviewers at the time of submission for the purposes of evaluating the manuscript. All manuscripts reporting original research published in Nature journals must include a data availability statement” (Nature, n.d.).</p>
	<p>“After publication, all data and materials necessary to understand, assess, and extend the conclusions of the manuscript must be available to any reader of Science. ... After publication, all reasonable requests for data or materials must be fulfilled. Any restrictions on the availability of data, codes, or materials, including fees and restrictions on original data obtained from other sources must be disclosed to the editors ... Unreasonable restrictions on data or material availability may preclude publication” (Science, n.d.).</p>
	<p>“To allow others to verify and build on the work published in Royal Society journals, it is a condition of publication that authors make available the data, code and research materials supporting the results in the article. Datasets and code should be deposited in an appropriate, recognised, publicly available repository. ... Exceptions to the sharing of data, code and materials may be granted at the discretion of the editor, especially for sensitive information such as human subject data or the location of endangered species. Authors must disclose upon submission of the manuscript any restrictions on the availability of data, code and research materials” (The Royal Society, n.d.).</p>
	<p>“PLOS journals require authors to make all data underlying the findings described in their manuscript fully available without restriction, with rare exception. When submitting a manuscript online, authors must provide a Data Availability Statement describing compliance with PLOS’s policy. If the article is accepted for publication, the data availability statement will be published as part of the final article. Refusal to share data and related metadata and methods in accordance with this policy will be grounds for rejection. PLOS journal editors encourage researchers to contact them if they encounter difficulties in obtaining data from articles published in PLOS journals. If</p>

1
2
3 restrictions on access to data come to light after publication, we reserve the right to post a
4 correction, to contact the authors' institutions and funders, or in extreme cases to retract the
5 publication" (PLOS, n.d.).
6
7
8
9

10
11
12
13 *Initiatives from Researchers.* Data sharing has also been promoted through collaborative
14 efforts of individual researchers. A prominent example is the Peer Reviewers Openness
15 initiative (PRO; Morey et al., 2016). PRO aims to encourage data sharing having its
16 signatories refuse to offer comprehensive reviews of manuscripts unless five requirements are
17 met. Included among these requirements is that the data are made publicly available in a
18 trusted repository. Authors who do not wish to comply with the requirements should provide
19 a justification.
20
21
22
23
24
25
26
27
28




29 The community-led Registered Reports (RR) initiative also includes an emphasis on
30 data sharing (<https://cos.io/rr/>). In contrast to conventional articles, Registered Reports are
31 reviewed before results are known, and in most cases before they even exist. Manuscripts that
32 survive pre-study review based purely on theoretical and methodological quality are then
33 provisionally accepted regardless of the results, provided various pre-specified quality checks
34 are met. Although public data archiving is not a necessary feature of the RR format, many
35 journals incorporate it into their policies to facilitate sharing and further enhance
36 transparency. To date, of 71 journals that have adopted RRs, 22 require data deposition
37 (Registered Reports, n.d.) including — among others — *Cortex*, *European Journal of*
38 *Neuroscience*, and *Nature Human Behaviour*.
39
40
41
42
43
44
45
46
47
48
49
50

Another community-led initiative has resulted in the so-called FAIR data principles, a set of guidelines that require data to be Findable, Accessible, Interoperable, and Reusable (Wilkinson, Dumontier, Aalbersberg, Appleton, Axton, Baak, ... & Bouwman, 2016). The FAIR guidelines are intended to help researchers implement good data management plans.

Initiatives from Funders. Research funders worldwide actively promote data sharing. Such funders include the National Science Foundation (NSF), the National Institutes of Health (NIH), and the Netherlands Organisation for Scientific Research (NWO). All three funding agencies require that applicants provide a detailed plan for data sharing (National Science Foundation, 2014; National Institutes of Health, 2015).

In addition, the German Research Foundation (DFG), the largest public funder in Germany, recently adopted a policy on data sharing which can be summarized as “publicly funded research, including the raw data, belongs to the public” (Schönbrodt, n.d.). To effectively apply this policy to various scientific disciplines, the DFG requested the development of discipline-specific implementation guidelines. In reply to this request, the German Psychological Society wrote a guideline in which they proposed a distinction between two levels of data sharing: “Type 1” data sharing refers to the sharing of the data needed to reproduce the results reported in the article, such that only a subset of the collected data may be publicly released; “Type 2” data sharing refers to the sharing of the entire data set.

Box 2. Data sharing policies of several funding organizations.

Funder	Data Sharing Policy
	<p>“... The basic principle is therefore open access to data and that access to data is only limited when necessary. To make data that emerges from NWO-funded research as accessible and reusable as possible, NWO has decided to implement the data management policy in all NWO funding instruments with effect from 1 October 2016.</p> <p>In concrete terms this means that all calls for proposals published from 1 October 2106 onwards will include the data management protocol. The data management protocol consists of two steps:</p> <ol style="list-style-type: none"> 1. A data management section in the research proposal in which the researcher should answer a number of short questions; 2. A data management plan that must be submitted after the proposal has been awarded funding. The approval of this plan is a prerequisite for NWO disbursing the grant” (Netherlands Organisation for Scientific Research, 2016).
	<p>“Plans for data management and sharing of the products of research. Proposals must include a supplementary document of no more than two pages labeled "Data Management Plan". This supplementary document should describe how the proposal will conform to NSF policy on the dissemination and sharing of research results ...” (National Science Foundation, 2014).</p>
	<p>“The 2003 NIH Data Sharing Policy encourages NIH-funded researchers to share their final research data for use by other researchers in a timely way (i.e., no later than the acceptance for publication of the main findings from the final data set). The Policy expects applicants requesting \$500,000 or more in direct costs in funding from NIH for research for any one year to include a data sharing plan or state why data sharing is not possible. Supplemental guidance materials suggest that plans should describe (1) whether and how data will be made available to others, including provisions for protection of privacy, confidentiality, security, intellectual property, or other rights as appropriate; (2) items such as the data to be shared (e.g., genomic, clinical, or images), the expected timeline for when the data will be available, data formats, the format of the final data set, any query and/or analytic tools that will be provided, and the mode of data sharing</p>

(e.g., through a data archive or enclave or under the researcher's own auspices by mailing a disk or posting data on an institutional or personal website); and (3) procedures to request the data and any required data sharing agreements including the criteria for accessing data and any limitations placed on the use of data" (National Institute of Health, 2016).



"The German Research Foundation (DFG), the largest public funder of research in Germany, updated their policy on data sharing, which can be summarized in a single sentence: Publicly funded research, including the raw data, belongs to the public. Consequently, all research data from a DFG funded project should be made open immediately, or at least a couple of months after finalization of the research project [...]. Furthermore, the DFG asked all scientific disciplines to develop more specific guidelines which implement these principles in their respective discipline" (Schönbrodt, n.d.).

Finally, several funders in the United Kingdom, such as Wellcome and Research Councils UK, are signatories to the Concordat on Open Research Data 2016 (Wellcome, 2016). This concordat was developed by the UK research community and aims to "help to ensure research data gathered and generated by members of the UK research community is, wherever possible, made openly available for use by others in a manner consistent with relevant legal, ethical and regulatory frameworks and disciplinary norms, and with due regard to the costs involved" (Johnson, 2016, p.4).

The recent initiatives from journals, researchers, and funders highlight the change in awareness concerning the importance of data sharing. Nevertheless, in order to accomplish behavioral change efficiently, it is helpful to know why psychologists are reluctant to share their data: "knowing why authors choose not to be open would be tremendously useful to the

scientific community at large” (Morey et al., 2016). If the perceived barriers to sharing are not targeted by current initiatives, alternative initiatives can be designed. This also requires that we know what researchers consider important preconditions for data sharing.

There exists a small literature on barriers and conditions for data sharing, and we review it briefly before turning to our own study.

Barriers and Preconditions for Data Sharing

Two empirical studies have been conducted on factors that deter researchers from data sharing. Tenopir et al. (2011) found that the most common barriers reported by researchers were “insufficient time” and “a lack of funding”. Other reasons were “not having the rights to make data public”, “no place to put data”, “a lack of standards”, “sponsor does not require”, “do not need data”, and the belief that data should not be electronically available to other researchers. By means of a different questionnaire, Schmidt, Gemeinholzer and Treloar (2016) identified several other barriers for data sharing: (1) the desire to publish results before releasing the research data; (2) legal constraints; and (3) loss of credit or recognition².

Other studies with a smaller scope have also contributed to the understanding of barriers to data sharing. For example, Wicherts, Bakker & Molenaar found willingness to share research data was related to the internal consistency of statistical results and strength of evidence (2011). In a qualitative study, Cheah et al. (2015) identified three main factors that influenced researchers’ reluctance to share primary research data: (1) barriers relating to the

² Note that several options are available to credit the collector of the data, for example by referencing the original article, referencing a data paper, or directly citing a data set using a persistent identifier.

1
2
3 anonymity of participants; (2) barriers concerning worries about acknowledgement and the
4
5 fear that secondary researchers would publish findings from the original data before the
6
7 primary researcher; and (3) concerns related to the demand and resources that are required to
8
9 share data in an effective way (e.g., effort of time to prepare a data set for sharing and the
10
11 adequate curation of shared data). Additional barriers may include ethical and legal concerns.
12
13 These barriers are not the focus of the present work and are [discussed](#) elsewhere (Meyer, this
14
15 issue).
16
17

18
19 To complement the study of barriers to data sharing, it is important to know under
20
21 what conditions researchers would be more willing to share their data. Tenopir et al. (2011)
22
23 found that researchers would be more likely to share data if they could place conditions on the
24
25 access of the file, and that they wish secondary users of their data properly cite the primary
26
27 researcher(s) of the data. Using the same data set as Tenopir et al. (2011), Sayogo and Pardo
28
29 (2013) concluded that having data management skills and receiving support from the
30
31 institution in terms of funding, training, technical support, storing data, and data management,
32
33 positively influence the likelihood that researchers share data. A second important
34
35 determinant is the acknowledgement of the primary researcher, in the form of co-authorship
36
37 on publications, formal acknowledgment of the data providers, and collaboration opportunity.
38
39

40
41 In a qualitative study, Wallis, Rolando and Borgman (2013) found that the five most
42
43 important conditions were: (1) first rights to publish results; (2) proper attribution of the data
44
45 source; (3) familiarity between sharer and recipient; (4) funding agency expectations; and (5)
46
47 the amount of effort required to share. Following Wallis, Rolando & Borgman, Cheah et al.
48
49 (2015) found that solutions to data-sharing fell into three groups: (1) “ensuring that data are of
50
51

good quality” (refers to the idea that data sharing practices can be effective only if shared data sets are of high quality and are both well-managed and curated); (2) “high standards of consent” (solutions pertains to worries expressed by interviewed researchers regarding the adequate protection of the interests of research participants and communities), and (3) “better data governance” (solutions relates to the governance of data and addresses researchers’ concerns about the openness of their data).

The Current Study

In sum, despite promising initiatives and technological developments, the public sharing of research data remains uncommon. A handful of studies have investigated barriers to data sharing and remedial measures, but none has specifically focused on psychology, where distinctive issues may arise because of the use of human participants. In the current study, we addressed the question: *Why are researchers in psychology reluctant to share research data and what measures provide effective incentives to share?* We address this question using data from a questionnaire distributed to a sample of 14,184 unique email addresses belonging to researchers in the field of psychology.

Method

Sample

To collect email addresses of researchers active in psychology, metadata from a set of

articles was retrieved from Web of Science (WoS). WoS is a database for multidisciplinary research articles. In WoS, the following criteria were set: “Topic: Psychology”; “Research areas: Psychology”; “Type: Articles”; “Year published: 2005 - 2015”; and “Language: English”³. This search resulted in 21,116 research articles (collected on 6 July, 2016)⁴. In order to obtain metadata from these articles, including the authors’ email addresses, an automated Python script was used. This yielded a total number of 16,284 unique email addresses.

Note that this method of obtaining a sample of participants is subject to self-selection: we cannot force researchers to complete the survey. Consequently, researchers who already have a strong opinion about open science practices may be more likely to participate. Therefore, the sample of the current study may not adequately represent the population of researchers in psychology. We revisit this inevitable limitation in the Discussion section.

The 16,284 email addresses were used for three phases of this study; (1) the pilot study (to assess the quality of the questionnaire; 6 responses out of 100 emails), (2) the preliminary study (to formulate a proper analysis plan; 132 responses out of 2,000 emails), and (3) the primary study (780 responses out of the remaining 14,184 emails; see also Figure A1 in the online appendix).

³ The exact executed query in Web of Science is: “(from Web of Science Core Collection) You searched for: TOPIC: (psychology) Refined by: RESEARCH AREAS: (PSYCHOLOGY) AND DOCUMENT TYPES: (ARTICLE) Timespan: 2005-2015. Indexes: SCI-EXPANDED, SSCI, A&HCI, BKCI-S, BKCI-SSH, ESCI”.

⁴ The reason for this relatively modest number of articles is the fact that Web of Science is a selective, rather than comprehensive database. For example, Web of Science contains 280 journals with “psychology” in the title.

Materials

Our items were constructed based on earlier surveys (described in the introduction) and a discussion at a 2015 Oxford expert meeting on reproducibility. The adequacy of the resulting items and corresponding answer options were tested in a pilot study and in a preliminary study, where researchers were invited to provide feedback on the items. Subsequently, the data from the preliminary study was used to construct an analysis plan for the data gathered in the primary study. Both the preliminary study and primary study were preregistered.

The survey consisted of four subscales: (1) generic information; (2) perceived barriers for data sharing; (3) preconditions for data sharing; and (4) desirability and profitability of data sharing. See Appendix A for a full version of the questionnaire.

Generic Information. Seven items addressed: (1) age; (2) gender; (3) current academic position; (4) years active in academia; (5) percentage of research projects in which research data was shared; (6) how likely a researcher is to share research data for their next paper; and (7) how likely a researcher is to share research data if the current academic system remains unchanged. The questionnaire for the preliminary data collection (and the initial phase of the primary data collection) also included a question about the university at which respondents are employed, but in order to guarantee anonymity we decided to remove this question from the questionnaire.

Barriers. Thirty items addressed barriers to data sharing. One item pertained to legal

constraints. The remaining 29 items contain a statement relating to a possible barrier for data sharing and were answered on a 7-point Likert scale, varying from 1 (disagree strongly) to 7 (agree strongly). Examples of items in the barrier scale are “It is unfair for other researchers to profit from my hard work; they should collect their own data”, and “I am afraid that other researchers will perform alternative analyses on my data and argue that my conclusions are invalid”.

A subset of eight barrier items were judged to be particularly sensitive to dishonest responding. In order to address this problem, these items were stated in two ways; in one formulation the respondent evaluates himself, and in the second formulation the respondent evaluates others (e.g., John, Loewenstein, & Prelec, 2012; see also Anderson, Martinson, & de Vries, 2007).

An example of an item in which a respondent is required to evaluate himself is: “I feel that by sharing my data I lose control over my intellectual property”. This same item is also formulated for the evaluation of others: “I think that other researchers are afraid that by sharing their data, they lose control over their intellectual property”.

Preconditions. Twenty items addressed the conditions under which psychologists would be willing to share their primary research data. Each item started with “How likely are you to share your research data if ...”, and were answered using a 7-point Likert scale varying from 1 (not at all likely) to 7 (very likely). Examples of items are “... minimal time is required to share”, and “... you had the first rights to publish the results from the data”.

1
2
3 *Desirability and Profitability.* Five items addressed respondents' general opinion about data
4 sharing. Four items related to the desirability and profitability of data sharing for researchers'
5 field and their current research project. These items were answered using a 5-point Likert
6 scale varying from very undesirable (1) to very desirable (5) for the desirability questions, and
7 from none at all (1) to a great deal (5) for the profitability questions. An example question is
8 "How desirable do you think data sharing is for your current research project?". The
9 remaining item queried researchers about the main benefit of data sharing. The listed response
10 options were derived from the benefits described by Borgman (2012).
11
12
13
14
15
16
17
18
19
20
21
22

23 **Procedure**

24
25 For the primary study, the questionnaire was sent to 14,184 unique email addresses
26 using the Qualtrics Mailer service. When no response was recorded within 2 weeks, we sent a
27 single reminder.
28
29
30

31
32 The questionnaire contained two instructions. The first instruction stated the definition
33 of data sharing used for the questionnaire, that is, "data sharing is the activity of making
34 primary research data available in an online repository upon publication of the article
35 associated with the data".
36
37
38
39

40 The second instruction covered legal constraints relating to the sharing of data.
41
42 Participants were instructed that when legal constraints in their field prohibit data sharing,
43 they should disregard this issue when filling out the remainder of the questionnaire.
44
45
46

47 The median duration for filling out the questionnaire was 8 minutes and 56 seconds,
48 with an interquartile range of 5 minutes and 21 seconds.
49
50

Analysis Plan Based on Preliminary Data

In order to analyze the survey data appropriately, part of this project concerned the collection of preliminary data. Towards this purpose, the questionnaire was distributed to 2,000 email addresses randomly drawn from a total sample of 16,184 email addresses (for more information, consult the “Sample Size” section). At this exploratory stage, we initially entertained factor analysis and network analysis.

However, bar plots of the preliminary data appeared to yield the most valuable information (the full results of the preliminary data collection are uploaded to the OSF project, [see the Disclosures](#)). Consequently, the data from the primary data collection phase are presented using bar plots.

The results of the preliminary data analysis yielded two results that we hoped to confirm in the analysis of the primary data set by means of an explicit test. Specifically, based on the preliminary data, we wished to test whether there is a difference in Likert scores between (a) fear evaluated for self vs. for others; and (b) the desirability of sharing for the research field vs. the current research project. To evaluate the evidence we will execute one-sided Bayesian t-tests. The direction of the tests were provided by the preliminary data. The alternative hypothesis has a default folded Cauchy prior with a mode at 0 and a width of 0.707 (the current default value; e.g., Wagenmakers et al., in press). These analyses were preregistered through AsPredicted.org.

Disclosures

Preregistration

Both the preliminary study (<https://osf.io/9rh5a/>) and the primary study (<https://osf.io/xjh4p/>) were preregistered through AsPredicted.org.

Data, Materials, and Online Resources

The documents that are available on the Open Science Framework are included in Table 1.

Table 1. Documents and direct URLs to documents on the OSF.

Document	URL
Online appendix	https://osf.io/26bgq/
Data set that excludes respondents who did not complete the survey (N = 600)	https://osf.io/gv98z/
Data set that includes respondents who did not complete the survey (N = 780)	https://osf.io/2q79y/
Preliminary data	https://osf.io/khg7j/
Survey	https://osf.io/nts79/
Invitation and reminder emails	https://osf.io/uw9bh/
Rcode of the figures	https://osf.io/r876q/
Entire OSF project page	https://osf.io/xm8dh/

Measures

1
2
3 We report how we determined our sample size, all data exclusions, all manipulations,
4
5 and all measures in the study.
6
7

8 9 **Conflicts of Interest**

10
11 None.
12
13
14
15

16 **Author Contributions**

17
18 EJW, CC, DB, MM, and TN generated the idea for the study. BLH conducted the
19
20 literature study. BLH, EJW, CC, DB, MM, and TN created the survey. BLH gathered and
21
22 analyzed the data for the preliminary study and described the results. BLH gathered and
23
24 analyzed the data for primary study. BLH wrote the first draft of the manuscript. EJW
25
26 critically edited the first manuscript. CC, DB, MM, and TN gave feedback to the manuscript
27
28 which BLH and EJW processed. EJW and BLH wrote the manuscript for first submission to
29
30 AMPPS. EJW and BLH wrote the revised manuscript. All authors approved the final
31
32 submitted version of the manuscript.
33
34
35
36
37

38 **Acknowledgements**

39
40 We are grateful to C. H. J. Hartgerink from Tilburg University for collecting the set of
41
42 author email addresses used in this survey.
43
44
45

46 **Results**

47 48 **Respondents**

1
2
3 For the primary data collection, an invitation to the survey was sent to a total of
4
5 14,184 email addresses. After accounting for 2,155 bounces (i.e., automatic failures to deliver
6
7 the email, for instance because an address was no longer active), the total number of requests
8
9 equalled 12,029. The questionnaire was filled out by 826 researchers, but not every
10
11 respondent completed the survey. Removing incomplete surveys resulted in a total of 600
12
13 responses (i.e., a response rate of 4.99%).
14
15

16 Of these 600 respondents, 411 reported they were male and 189 female. The average
17
18 age of the respondents was 45.9 years (SD = 13.25). Most had relatively senior academic
19
20 positions: the sample contained 35 PhD students, 61 postdocs, 96 assistant professors, 131
21
22 associate professors/readers, 228 professors, and 49 respondents with other academic
23
24 positions. On average, respondents had been active in academia for 19.42 years (SD = 12.38).
25
26

27 The first question concerned the percentage of research projects in which data was
28
29 shared in an online repository. The majority of researchers indicated that in less than 10% of
30
31 their projects they shared the primary data in an online repository (consult Figure A2 in the
32
33 online appendix). In the group of researchers younger than 45, 25.23% indicated to have
34
35 shared data in more than 10% of their research projects. In the age group of researchers 45
36
37 and older, 8.87% indicated to have shared data in more than 10% of their research projects.
38
39 This difference may be explained by a lack of availability of suitable repositories in the past.
40
41

42 Regarding future data sharing, many researchers indicated to be somewhat or
43
44 extremely likely to share the research data of their next paper in an online repository (consult
45
46 Figure A3 in the online appendix). Across researchers younger than 45, 48.33% indicated to
47
48 be somewhat likely or extremely likely to share data for their next paper, whereas this
49
50

percentage was 37.64% across researchers aged 45 and older.

Additionally, we assessed how likely researchers in psychology are to share their data in the current academic system. The response pattern shows an equal distribution of answers to this question (consult Figure A4 in the online appendix).

Furthermore, we asked researchers whether they experience legal constraints related to the sharing of research data. Of the total of 600 respondents, 169 respondents (28.17%) indicated to strongly agree, somewhat agree, or neither agree nor disagree with the statement that they are not allowed to share research data (Figure A5 in the online appendix). These 169 respondents were presented a follow-up question about the type of legal constraint they experience. The results show that: (1) eight respondents experienced constraints set by their Institutional Review Board which prohibited them from sharing their data (4.73%); (2) 47 respondents were constrained by a consent form which stated that the data will not be shared (27.81%); (3) 34 respondents indicated that anonymity cannot be guaranteed if the data is shared (20.12 %); (4) 27 respondents were prohibited from sharing data due to the interests of stakeholders (e.g., company or institution; 15.98%); and (5) 14 respondents indicated that their funder/adviser/boss did not allow them to share data (8.28 %). The remaining 39 respondents (23.08%) experienced other legal constraints.

Desirability and Profitability

Questions relating to the general opinion on data sharing showed that the vast majority of respondents consider data sharing desirable (Figure 1a) and profitable (Figure 1b) for their own research field. In relation to researchers' current research project, these figures were

found to be slightly lower (Figure 4a, Figure 1b.).

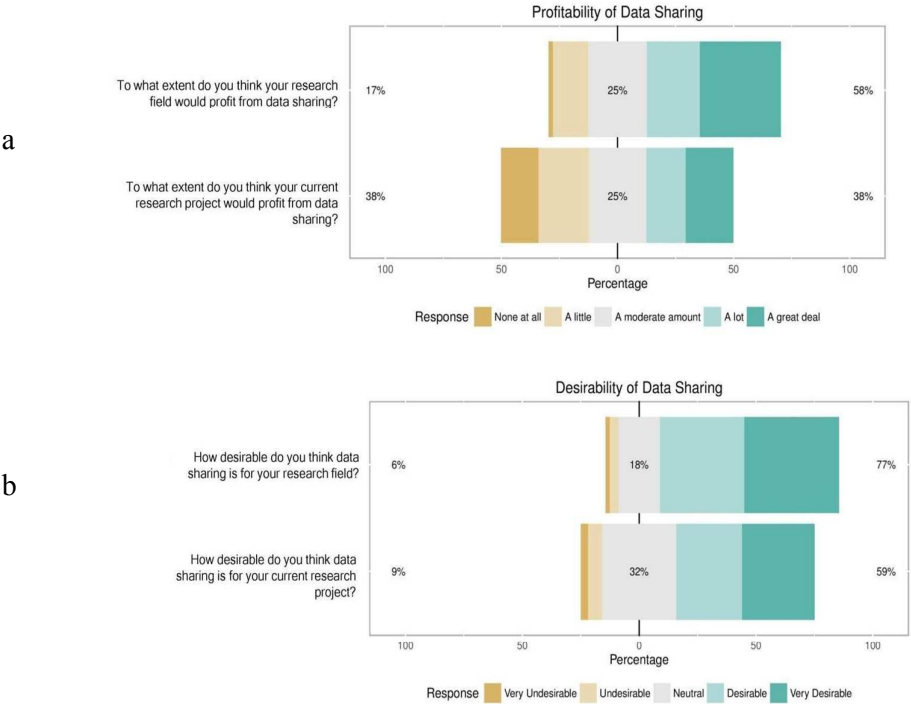


Figure 1. Responses to questions relating to the desirability (a) and profitability (b) of data sharing for the researchers' own field and their own current project. For each statement, the percentages on the left represent the proportion of researchers who responded with "none at all" and "a little" for plot a, and "very undesirable" and "undesirable" for plot b. The percentages in the center show the proportion of researchers who responded with "a moderate amount" for plot a, "neutral" for plot b. The percentages on the right represent the proportion of researchers who responded with "a lot" and "a great deal for plot a, and "desirable" and "very desirable" for plot b. This figure was created using the "likert" package in R (Bryer & Speerschneider, 2015).

To confirm statistically whether researchers consider data sharing as more desirable and profitable to their research field than to the research project they are currently engaged in themselves, we conducted one sided Bayesian paired samples t-tests. The results show extreme evidence in support of the alternative hypotheses, indicating that researchers are indeed more convinced of the desirability ($BF_{+0} = 9.059e + 19$) and profitability ($BF_{+0} = 9.656e + 49$) of data sharing for the research field in which they are active, than for their current research project (Table A3 in the online appendix).

We also asked researchers to indicate what they consider to be the greatest benefit of data sharing. Based on Borgman (2012), respondents could select one of four benefits of data sharing: (1) the ability to reproduce or verify research (chosen by 207 respondents or 34.5%); (2) make results of publicly funded research available to the public (chosen by 44 respondents or 7.33%); (3) enable other researchers to ask new questions of extant data (chosen by 231 respondents, or 38.5%); and (4) advance the state of research and innovation (chosen by 118 respondents or 19.67%).

Barriers

To assess the reasons why researchers in psychology are reluctant to share research data, three sets of questions were analyzed: (1) barriers for data sharing; (2) self-evaluation of fears related to barriers; and (3) evaluation of others' fears related to barriers.

Respondents were instructed to evaluate to what extent each barrier keeps them from sharing data. Every statement in the questionnaire represents a possible barrier. Figure 2 shows the ordered results for these statements. The items on top were items that researchers

evaluated as the most relevant barriers for data sharing.

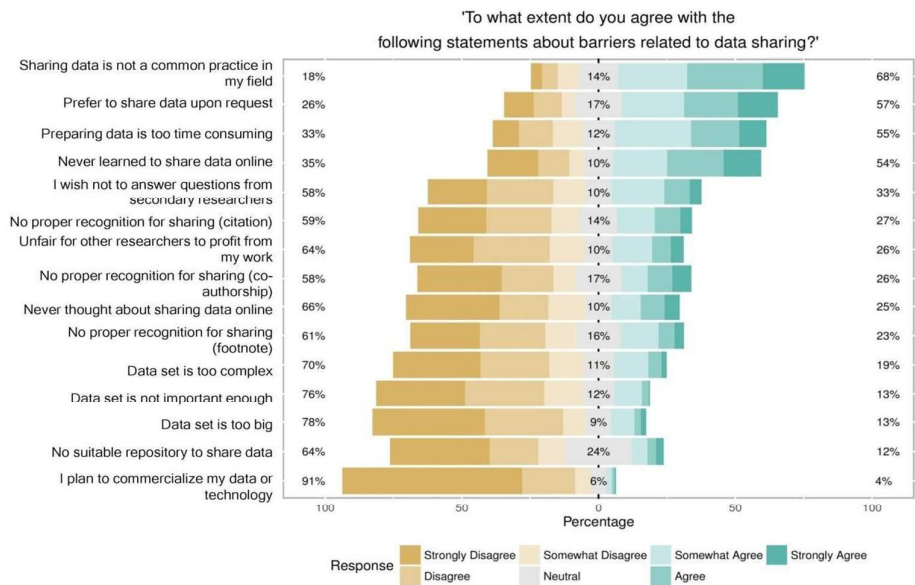


Figure 2. Results to the question to what extent several barriers keep researchers away from sharing research data. For each statement, the percentages on the left represent the proportion of researchers who responded with “strongly disagree”, “disagree”, and “somewhat disagree”. The percentages in the center show the proportion of researchers who responded with “neutral”. The percentages on the right represent the proportion of researchers who responded with “somewhat agree”, “agree”, and “strongly agree”. Figure was created using the “likert” package in R (Bryer & Speerschneider, 2015).

Out of all measured barriers for data sharing, the four barriers that researchers considered most relevant were: (1) the online sharing of research data is not a common practice in the researcher’s field; (2) researchers prefer to share their data sets upon request,

suggesting that they want to control what they distribute and whom they distribute it to; (3) researchers consider preparing data to make them suitable for sharing too time consuming; and (4) researchers have never learned to share their data online. These findings are highly similar to the findings in the preliminary data set.

Next, researchers evaluated barriers related to six fears: (1) the fear of being scooped, meaning that other researchers publish results from the data before the primary researcher can; (2) the fear that alternative analysis on shared data will lead to the rejection of conclusions; (3) the fear of losing control of intellectual property; (4) the fear that others will discover errors in shared data; (5) the fear that secondary users will misinterpret the data; and (6) the fear that others will use shared data for unintended purposes. Researchers evaluated these fear related barriers for both themselves (Figure 3a), as well as for others (Figure 3b). Possible differences between these two evaluations may expose the presence of socially desirable responding.

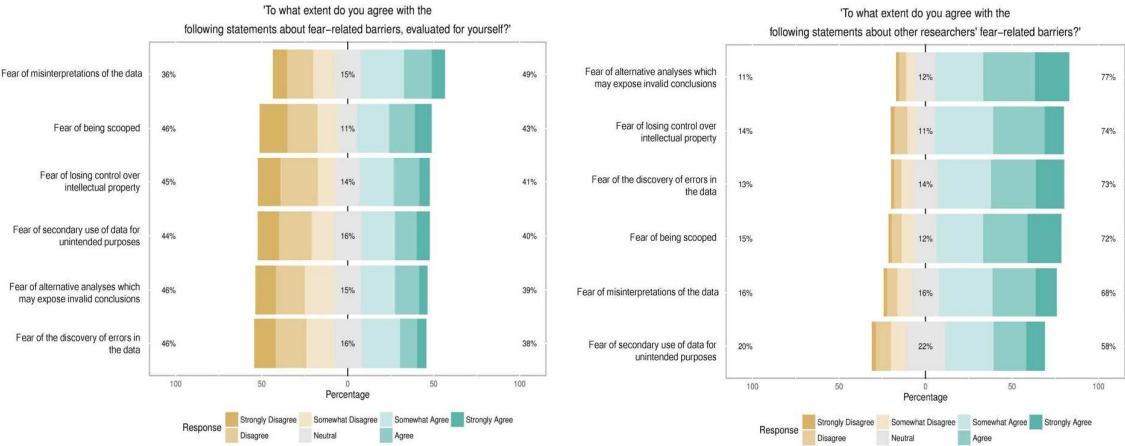


Figure 3. Results to the question to what extent several fear-related barriers keep researchers

1
2
3 themselves (a) and other researchers (b) away from sharing research data. For each
4
5 statement, the percentages on the left represent the proportion of researchers who responded
6
7 with “strongly disagree”, “disagree”, and “somewhat disagree”. The percentages in the
8
9 center show the proportion of researchers who responded with “neutral”. The percentages on
10
11 the right represent the proportion of researchers who responded with “somewhat agree”,
12
13 “agree”, and “strongly agree”. Figure created using the “likert” package in R (Bryer &
14
15 Speerschneider, 2015).
16
17

18
19 The results show that researchers often agree that the aforementioned fears keep other
20
21 researchers from sharing research data (Figure 3b). The fear of alternative analyses that may
22
23 expose invalid conclusions and the fear of the discovery of errors in the data, were evaluated
24
25 as the most prominent barriers for others. When researchers evaluated these fear related
26
27 barriers for themselves (Figure 3a), they indicated the fear of misinterpretation of the data and
28
29 the fear of being scooped as the most relevant barriers for data sharing. This shows a
30
31 remarkable difference between how researchers evaluate barriers for themselves and how they
32
33 do so for others.
34
35

36
37 Another noteworthy difference between the evaluations of others and researchers’
38
39 evaluation of themselves, is that respondents seem to think that other researchers are more
40
41 strongly impacted by barriers to data sharing than they are themselves. To statistically test this
42
43 difference, we conducted one-sided Bayesian paired samples t-tests for every fear that was
44
45 evaluated. Based on the results of the preliminary data, we tested the alternative hypotheses
46
47 that all six evaluations of the fears of other researchers have higher values than researchers’
48
49 own evaluations of these fears. The alternative hypothesis has a default folded Cauchy prior
50
51

with a mode at 0 and a width of 0.707 (the current default value). For every pair of items, there is extreme evidence in support of the alternative hypotheses, which state that the fear-related barriers are seen as more relevant for others than for researchers themselves (consult Table A2 in the online appendix for details). These results suggest that when researchers evaluate the relevance of these barriers for themselves, these fears do not convincingly limit their decision to share research data. However, when evaluated for others, these fears are seen as important reasons not to share research data. This discrepancy suggests that responding was sensitive to socially desirable responding. An alternative account is that the self-selected group of respondents were relatively fearless.

Preconditions

After the barriers, researchers were asked how likely they would to share their research data under 22 different conditions for data sharing. Figure 4 contains a summary of the results. The results of the questionnaire suggest that fulfilling any of the conditions would lead to a substantially increased willingness to share research data.

Every condition from the questionnaire yielded an increased tendency to share research data, but five conditions were considered most effective and three of these relate to the mandatory sharing of data. If research funders, journals, and institutions were to mandate data sharing, researchers indicated that they would comply. This suggests that mandatory data sharing would not lead to researchers seeking alternative funders or publishers who do not have this requirement.

The fourth condition relates to the financial encouragement of data sharing. Researchers indicated they would be more likely to share their data if, as an incentive, their

research grants would increase. A fifth condition that researchers consider effective is encouragement by funders to share research data.

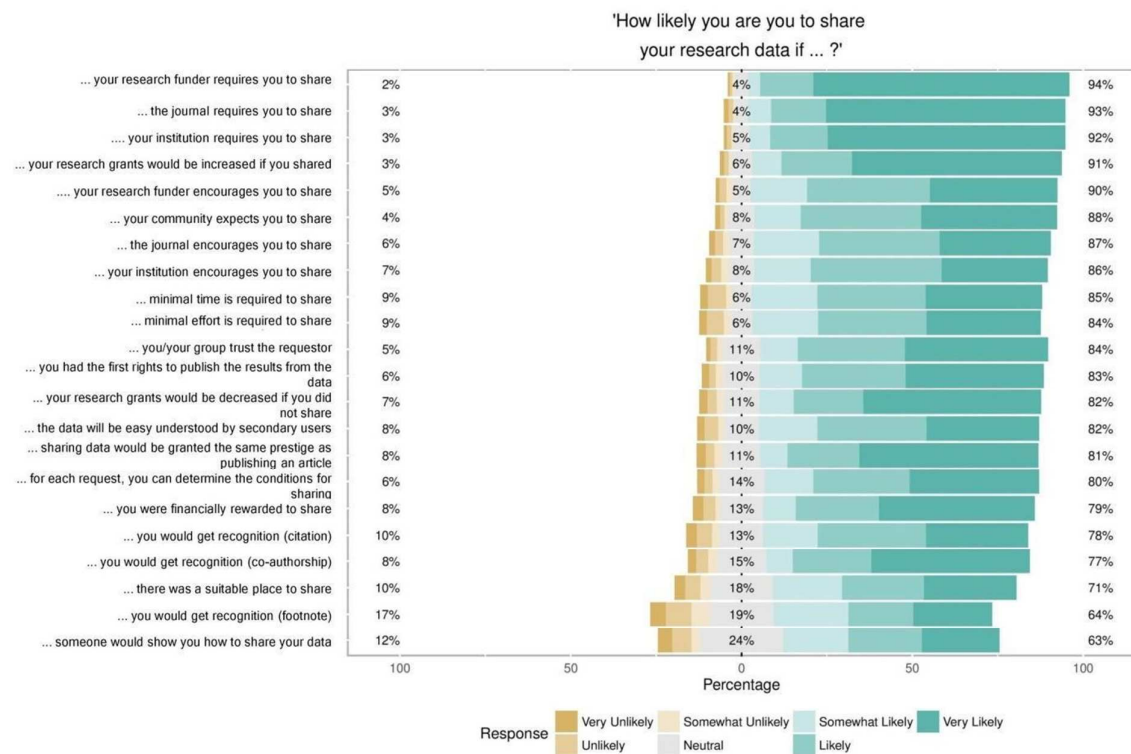


Figure 4. Results to the question on the likeliness to share under several conditions for sharing research data. For each statement, the percentages on the left represent the proportion of researchers who responded with “extremely unlikely”, “moderately unlikely”, and “slightly unlikely”. The percentages in the center show the proportion of researchers who responded with “neither likely nor unlikely”. The percentages on the right represent the proportion of researchers who responded with “extremely likely”, “moderately likely”, and “slightly likely”. Figure created using the “likert” package in R (Bryer & Speerschneider, 2015).

Conclusions

This study aimed to provide more insight into the barriers that psychologists perceive concerning the sharing of primary research data. In addition, the study also sought to determine the specific conditions under which psychologists would be more willing to share such data. The main findings of our study are:

1. Public data sharing is not a common practice.
2. The most frequently endorsed barriers for data sharing are: (1) sharing is not a common practice in the researchers' field; (2) researchers prefer to share their data only upon request; (3) researchers consider preparing data for sharing as excessively time consuming; and (4) researchers have never learned to share data.
3. Evaluated for others, respondents believe that the largest obstacles for data sharing are fears related to the quality of the data sets.
4. Evaluated for themselves, respondents believe that the largest obstacles for data sharing are fears of misinterpretations of their data, as well as the fear of being scooped.
5. In general, fear-related barriers were generally considered to affect other researchers more strongly than themselves.
6. To promote data sharing, mandatory data sharing (enforced by the relevant institution,

journal, or funder) and financial encouragement (e.g., increasing the research grant) appear to be highly effective measures to increase researchers' willingness to share data.

7. Researchers consider data sharing as both desirable and profitable for their particular research field, but this consideration does not extend to their own current research project.

Limitations

The main limitation of this study is that the sample was self-selected. Specifically, the response rate was about 5% -- given the daily pressure on researchers' time, we do not consider this surprisingly low. Similarly modest response rates were found in past studies that surveyed researchers on the topic of data sharing. Schmidt, Gemeinholzer, and Treloar (2016) report a response rate of 4.32% (i.e., 1253 complete responses out of 29,000 invitations). Tenopir et al. (2011) estimate that their survey reached 15,000 people; with 1329 respondents who answered at least one question, this yields a response rate of 8.86%. This rate is likely to be substantially lower when excluding respondents who failed to complete the entire questionnaire. In sum, our response rate is modest but not unusually low.

Importantly, the 5% response rate still translated to a sizeable number of 600 respondents. The key concern, however, is that the researchers who accepted our invitation may already have had a strong opinion about data sharing. This possibility is difficult to overcome for survey research in general. But even if we had been able to enforce a 100% response rate, there is no certainty that participants would have answered the items truthfully.

Moreover, even if all participants were to answer truthfully, there is no certainty that their later actions would be consistent with their earlier answers; the drawbacks of data sharing may loom larger in real life than when filling out a survey online.

Several demographics factors may have brought about the modest response rate. For example, people who are still active in academia may be more interested in the survey than people who have moved into teaching positions; people who are in the early stages of their career may have more time or be more comfortable with online surveys than people in the later stages of their career. It is possible that some of these demographics may correlate with people's attitude concerning data sharing. For these reasons, the present results, although suggestive, cannot be blindly generalized to the population.

A final limitation is the fact that both the authors and co-authors of articles were approached for participation. Consequently, there is some dependency in the responses to this questionnaire. Authors and their co-authors often work with each other extensively and for a long periods of time, meaning they may have adopted each other's opinions on open science, or started working together because of common interests related to open science in the first place. These shared opinions may have resulted in dependent responses to our questionnaire.

Recommendations

The findings in this study lead to several straightforward recommendations, concerning future research, the development of educational materials, and policy.

With respect to future research, one of the respondents suggested to adopt a different perspective and study what motivates the researchers who regularly share their data. Another

suggestion is to study the sharing of qualitative data. Not only are qualitative data more complex, but they may also be more difficult to anonymize. Lastly, future research could also devote more attention to ethical, moral, and legal issues related to data sharing (see Meyer, this issue).

Currently, the practice of data sharing is not given much attention in the curriculum of psychological researchers (Wicherts, 2016). To remedy the barriers that data sharing is not a common practice and that researchers never learned to share data, we recommend that institutions develop and provide educational materials on this topic. Such materials can indicate where data can be shared (e.g., on the OSF) and outline how they can be shared (e.g., as a csv file, together with a text document that describes its contents according to some universal coding system). This will be useful even when the materials affect only a select subset of researchers. In addition, funders who make data sharing mandatory would also do well to include pointers to materials that educate researchers on where and how to share their data effectively.

An increased focus on the educational aspect of data sharing practices may also remedy the fact that researchers view data sharing as excessively time consuming. Knowing exactly where and how to share data will make the data sharing process more efficient and less time consuming. For example, researchers may already describe the variables in their dataset while the research project is still ongoing. A related recommendation is to develop software solutions that help automate the process of data sharing.

With respect to policies that promote data sharing, our results showed that researchers consider mandatory data-sharing policies as the most effective. Enforced by funders, journals,

1
2
3 and/or institutions, such policies will almost certainly increase researchers' propensity to
4
5 share. Such policies may include the refusal to publish any articles of which the data was not
6
7 uploaded into an free accessible online repository (applicable for journals), the refusal to
8
9 provide grants to studies in which data will not be shared (applicable to funders), or the
10
11 refusal to hire researchers who have proven unwilling to share research data (applicable to
12
13 institutions). Importantly, such policies should safeguard the anonymity of respondents and
14
15 respect the ethics involved in data sharing. However, even in challenging scenarios,
16
17 opportunities exist to ensure the research process and the resulting data are as open as
18
19 possible. For example, the Medical Research Council (MRC) ensures that sensitive data from
20
21 population and patient studies can be shared by signing data sharing agreements with
22
23 researchers: "Data-sharing agreements must prohibit any attempt to (a) identify study
24
25 participants from the released data or otherwise breach confidentiality, (b) make unapproved
26
27 contact with study participants" (Medical Research Council, 2017). Although the context of
28
29 this particular policy is medical, it could perhaps benefit clinical psychologists as well.
30
31
32

33
34 In addition to mandating data sharing, our survey suggests that what may also be
35
36 effective is positive reinforcement, such as financial encouragement by funders. From the
37
38 perspective of the funder this could be seen as cost coverage for the sharing and curation of
39
40 data. However, from the researcher's perspective it may be perceived as an incentive for data
41
42 sharing.
43
44

45 To conclude, the findings of this study suggest that although researchers perceive
46
47 barriers to data sharing, at least some important barriers can be overcome relatively easily.
48
49 Specifically, we recommend that guidelines for mandatory data sharing are used in
50
51

combination with the development of educational materials that demonstrate where and how data can be shared in an effective and systematic manner. We hope that this study will contribute to both the acceleration of scientific progress, and the extensive employment of transparent research practices.

References

- Alsheikh-Ali, A. A., Qureshi, W., Al-Mallah, M. H., & Ioannidis, J. P. (2011). Public availability of published research data in high-impact journals. *PloS One*, 6(9), e24357.
- Anderson, M. S., Martinson, B. C., & De Vries, R. (2007). Normative dissonance in science: Results from a national survey of US scientists. *Journal of Empirical Research on Human Research Ethics*, 2(4), 3-14.
- Borgman, C. L. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*, 63, 1059-1078.
- Bryer, J., & Speerschneider, K. (2015). likert: Functions to Analyze and Visualize Likert Type Items. R package version 1.3.3. <https://CRAN.R-project.org/package=likert>.
- Ceci, S. J. (1988). Scientists' attitudes toward data sharing. *Science, Technology, & Human Values*, 13(1/2), 45-52.
- Ceci, S. J., & Walker, E. (1983). Private archives and public needs. *American Psychologist*, 38(4), 414.

- Cheah, P., Tangseefa, D., Somsaman, A., Chunsuttiwat, T., Nosten, F., Day, N., Parker, M. (2015). Perceived benefits, harms and views about how to share data responsibly: A qualitative study of experiences with and attitudes towards data-sharing among research staff and community representatives in Thailand. *Journal of Empirical Research on Human Research Ethics*, 10, 278-289
- Fecher, B., Friesike, S., & Hebing, M. (2015). What drives academic data sharing?. *PloS One*, 10, e0118053.
- Giofrè, D., Cumming, G., Fresc, L., Boedker, I., & Tressoldi, P. (2017). The influence of journal submission guidelines on authors' reporting of statistics and use of open research practices. *PloS One*, 12(4), e0175583.
- John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological science*, 23(5), 524-532.
- Johnson, J. E. (2016). Concordat on Open Research Data. Retrieved from <http://www.rcuk.ac.uk/documents/documents/concordatonopenresearchdata-pdf/>
- Kidwell, M. C., Lazarević, L. B., Baranski, E., Hardwicke, T. E., Piechowski, S., Falkenberg, L. S., ... & Errington, T. M. (2016). Badges to acknowledge open practices: a simple, low-cost, effective method for increasing transparency. *PLoS Biology*, 14(5), e1002456.
- Medical Research Council. (2017). *MRC Policy and Guidance on Sharing of Research Data from Population and Patient Studies* (v01-01 30). Retrieved from

<https://www.mrc.ac.uk/publications/browse/mrc-policy-and-guidance-on-sharing-of-research-data-from-population-and-patient-studies/>

Morey, R. D., Chambers, C. D., Etchells, P. J., Harris, C. R., Hoekstra, R., Lakens, D., ... & Vanpaemel, W. (2016). The Peer Reviewers' Openness Initiative: incentivizing open research practices through peer review. *Royal Society Open Science*, 3, 150547.

Meyer, M. (this issue). Practical tips for ethical data sharing. *Advances in Methods and Practices in Psychological Science*.

National Institutes of Health (2015). Plan for increasing access to scientific publications and digital scientific data from NIH funded scientific research.

National Science Foundation (2014, December 26. Proposal Preparation Instructions. Retrieved from http://www.nsf.gov/pubs/policydocs/pappguide/nsf15001/gpg_2.jsp#dmp.

Nature. Availability of data, material and methods. Retrieved March 17, 2017, from <http://www.nature.com/authors/policies/availability.html> (n.d.).

Netherlands Organisation for Scientific Research (2016). Data management protocol. Retrieved March 17, 2017, from <http://www.nwo.nl/en/policies/open+science/data+management>.

Nosek, B. A., Alter G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler S. J., ..., Yarkoni, T. (2015). Promoting an open research culture: Author guidelines for journals could help to promote transparency, openness, and reproducibility. *Science*, 348(6242), 1422-1425.

- Open Science Framework (n.d.). Badges to Acknowledge Open Practices. Retrieved from <https://osf.io/tvyxz/wiki/home/>.
- Piwowar, H. A., & Chapman, W. W. (2010). Public sharing of research datasets: a pilot study of associations. *Journal of Informetrics*, 4, 148-156.
- PLOS. Data Availability (n.d.). Retrieved March 17, 2017, from <http://journals.plos.org/plosone/s/data-availability>.
- Registered Reports (n.d.). Comparison of Registered Reports. Retrieved April 04, 2017, from https://docs.google.com/spreadsheets/d/1D4_k-8C_UENTRtbPzXfhjEyu3BfLxdOsn9j-otrO870/edit#gid=0
- Reidpath, D. D., & Allotey, P. A. (2001). Data sharing in medical research: an empirical investigation. *Bioethics*, 15(2), 125-134.
- Savage, C. J., & Vickers, A. J. (2009). Empirical study of data sharing by authors publishing in PLoS journals. *PloS One*, 4, e7078.
- 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.
- Schmidt, B., Gemeinholzer, B., & Treloar, A. (2016). Open Data in Global Environmental Research: The Belmont Forum's Open Data Survey. *PloS one*, 11(1), e0146695
- Schönbrodt, F. (n.d.). German Psychological Society fully embraces open data, gives detailed recommendations. Retrieved March 17, 2017, from <http://www.nicebread.de/open-data-recommendations/>.

- Science. Science: editorial policies. Retrieved March 17, 2017, from
<http://www.sciencemag.org/authors/science-editorial-policies> (n.d.).
- Sturges, P., Bamkin, M., Anders, J. H., Hubbard, B., Hussain, A., & Heeley, M. (2015).
 Research data sharing: Developing a stakeholder-driven model for journal policies.
Journal of the Association for Information Science and Technology, 66, 2445-2455.
- Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., et al. (2011). Data Sharing by
 Scientists: Practices and Perceptions. *PLoS One*, 6, e21101.
- The Royal Society (n.d.). Data sharing and mining. Retrieved March 17, 2017, from
<https://royalsociety.org/journals/ethics-policies/data-sharing-mining/>.
- Vanpaemel, W., Vermorgen, M., Deriemaeker, L., & Storms, G. (2015). Are we wasting a
 good crisis? The availability of psychological research data after the storm. *Collabra:
 Psychology*, 1(1).
- Vines, T. H., Andrew, R. L., Bock, D. G., Franklin, M. T., Gilbert, K. J., Kane, N. C., ... &
 Veen, T. (2013). Mandated data archiving greatly improves access to research data.
The FASEB Journal, 27, 1304-1308.
- Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, A. J., Selker, R.,
 Gronau, Q. F., Dropmann, D., Boutin, B., Meerhoff, F., Knight, P., Raj, A., van
 Kesteren, E.-J., van Doorn, J., Smira, M., Epskamp, S., Etz, A., Matzke, D., de Jong,
 T., van den Bergh, D., Sarafoglou, A., Steingroever, H., Derks, K., Rouder, J. N., &
 Morey, R. D. (in press). Bayesian inference for psychology. Part II: Example
 applications with JASP. *Psychonomic Bulletin & Review*.

- 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.
- Wellcome. (2016, July 28). Wellcome signs open data concordat. Retrieved from <https://wellcome.ac.uk/news/wellcome-signs-open-data-concordat>
- Wicherts, J. M. (2016). Data re-analysis and open data. In: Plucker, J. & Makel, M. (Eds.) (in press). *Doing Good Social Science: Trust, Accuracy, Transparency*. Washington: American Psychological Association.
- 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.
- Wicherts, J. M., Borsboom, D., Kats, J., & Molenaar, D. (2006). The poor availability of psychological research data for reanalysis. *American Psychologist*, 61, 726.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Bouwman, J. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, 3, 160018.
- Wolins, L. (1962). Responsibility for raw data. *American Psychologist*, 17, 657–658.

Online Appendix

Table A1. An Overview of the Findings of Two Empirical Studies on Barriers for Data

Sharing

<i>Article</i>		<i>Data</i>		
Tenopir et al. (2011)		<i>Reasons not to share data</i>	<i>Frequency</i>	<i>Percentage</i>
(DataONe Survey) <i>N</i> = 1,329	Insufficient time	603	53.6%	
	Lack of funding	445	39.6%	
	No rights to publish data	271	24.1%	
	No place to put data	264	23.5%	
	Lack of standards	222	19.8%	
	Sponsor does not require	196	17.4%	
	Do not need data	169	15.0%	
	Other reasons	164	14.6%	
	Should not be available	162	14.4%	
Schmidt, Gemeinholzer & Treloar (2016)		<i>Barriers to share data</i>	<i>Minor</i>	<i>Barrier</i>
<i>(The Belmont Forum)</i> <i>N</i> = 1,330	Desire to publish results before releasing data	14%	32%	54%
	Legal constraints	19%	34%	47%
	Loss of credit or recognition	21%	38%	41%
	Misinterpretation or misuse	26%	37%	37%
	Loss of control over intellectual property	28%	38%	34%
	Organizational constraints	24%	44%	32%
	Commercial use and exploitation	30%	38%	32%
	Difficulty of clarifying rights	26%	49%	26%
	Concerns about impact of data release	37%	41%	22%
	Concerns about legal liability or release	34%	46%	21%

Table A2. Bayes Factor, Median of Effect Size, and 95% Credible Interval of the Median for the Six Fear-Related Barriers, Evaluated for Others and Researchers Themselves

Items	BF ₊ 0	Median Effect Size	95% Credible Interval	
			Lower bound	Upper bound
Fear of losing control over intellectual property (Others vs. Self)	7.263e+57	0.755	0.661	0.848
Fear of alternative analyses, which may expose invalid conclusions (Others vs. Self)	1.152e+78	0.914	0.818	1.011
Fear of the discovery of errors in the data (Others vs. Self)	6.203e+72	0.872	0.780	0.970
Fear of being scooped (Others vs. Self)	9.345e+59	0.772	0.683	0.862
Fear of misinterpretations of the data (Others vs. Self)	1.351e+35	0.566	0.477	0.651
Fear of secondary use of data for unintended purposes (Others vs. Self)	1.565e+42	0.626	0.538	0.716

Table A3. Bayes Factor, Median of Effect Size, and 95% Credible Interval of the Median for

the Difference between the Desirability and Profitability of Sharing for the Research Field versus Researchers' Current Projects.

Items	BF ₊₀	Median	95% Credible Interval	
		Effect Size	Median	
			Lower bound	Upper bound
Desirability research field vs. desirability current research project	9.059e +19	0.416	0.335	0.500
Profitability research field. profitability current research project	9.656e +49	0.692	0.604	0.781

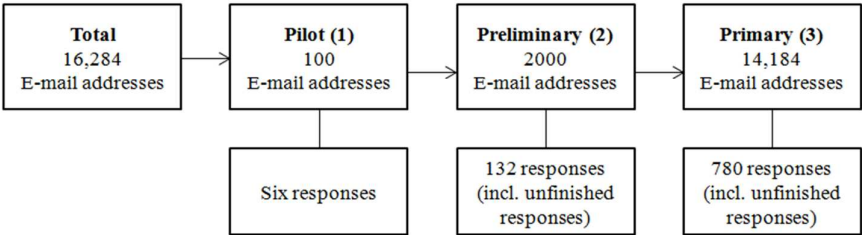


Figure A1. Overview of the three phases in this study: (1) pilot, (2) preliminary, and (3) primary, and the number of E-mail invites per phase. In the boxes below are the response rates for every phase.

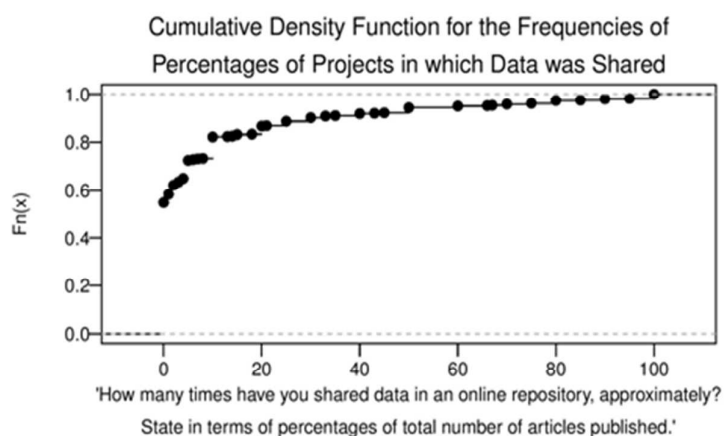


Figure A2. The cumulative density function for the frequencies of percentages of projects in which researchers shared their data.

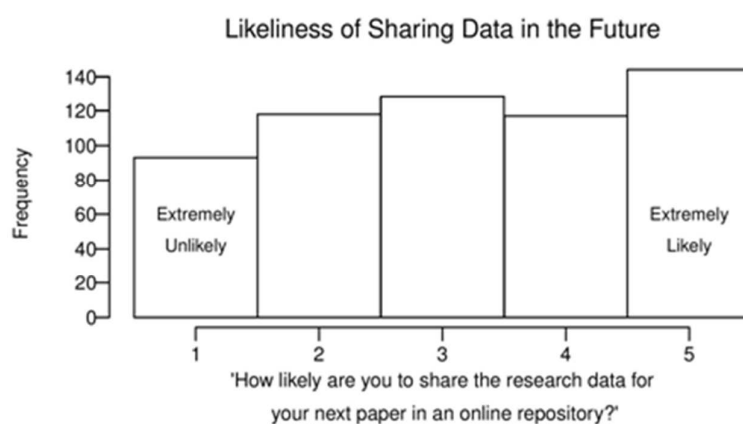


Figure A3. “How likely are you to share the research data for your next paper in an online repository?”. The values on the x axis refer to the five answering options: 1 = Extremely unlikely, 2 = Somewhat unlikely, 3 = Neither likely nor unlikely, 4 = Somewhat likely, 5 = Extremely likely.

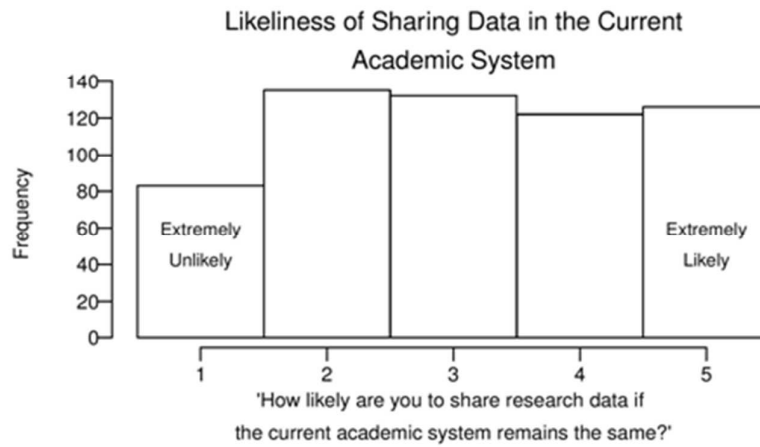


Figure A4. “How likely are you to share research data if the current academic system remains the same?”. The values on the x axis refer to the five answering options: 1 = Extremely unlikely, 2 = Somewhat unlikely, 3 = Neither likely nor unlikely, 4 = Somewhat likely, 5 = Extremely likely.

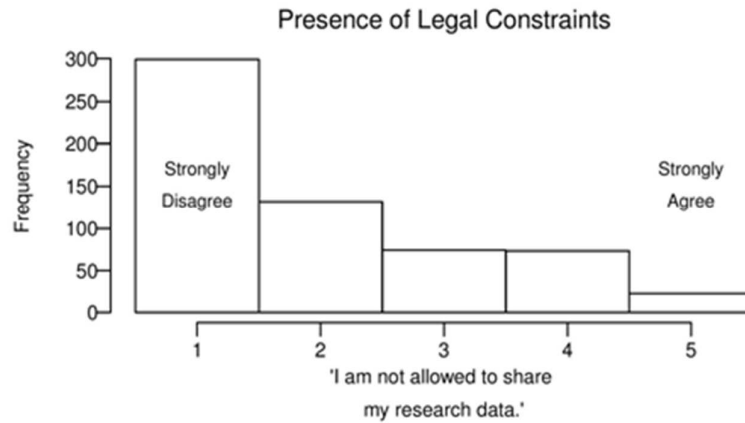


Figure A5. “I am not allowed to share my research data”. The values on the x axis refer to the five answering options: 1 = Strongly disagree, 2 = Somewhat disagree, 3 = Neither agree nor disagree, 4 = Somewhat agree, 5 = Strongly agree.