



ARTICLE

Received 30 Nov 2016 | Accepted 9 May 2017 | Published 20 Jun 2017

DOI: 10.1057/palcomms.2017.51

OPEN

A reputation economy: how individual reward considerations trump systemic arguments for open access to data

Benedikt Fecher^{1,2}, Sascha Friesike^{2,3}, Marcel Hebing^{2,4} and Stephanie Linek⁵

ABSTRACT Open access to research data has been described as a driver of innovation and a potential cure for the reproducibility crisis in many academic fields. Against this backdrop, policy makers are increasingly advocating for making research data and supporting material openly available online. Despite its potential to further scientific progress, widespread data sharing in small science is still an ideal practised in moderation. In this article, we explore the question of what drives open access to research data using a survey among 1564 mainly German researchers across all disciplines. We show that, regardless of their disciplinary background, researchers recognize the benefits of open access to research data for both their own research and scientific progress as a whole. Nonetheless, most researchers share their data only selectively. We show that individual reward considerations conflict with widespread data sharing. Based on our results, we present policy implications that are in line with both individual reward considerations and scientific progress.

¹ Department "Research Infrastructure", German Institute for Economic Research (DIW Berlin), Berlin, Germany ² Department "Knowledge Dimension", Alexander von Humboldt Institute for Internet and Society, Berlin, Germany ³ KIN research group, VU Amsterdam (VUA), Amsterdam, Netherlands ⁴ German Socio-Economic Panel Study, German Institute for Economic Research (DIW Berlin), Berlin, Germany ⁵ Department "Science 2.0", German National Library of Economics (ZBW), Kiel, Germany

Introduction

In 1942, when Robert K. Merton formulated his four norms that comprise the ethos of ethical science (universalism, communism, disinterestedness and organized skepticism), he probably did not think of researchers archiving their data in a public repository. However, at least two of his norms relate directly to open access to research data, which means that data and supporting materials are made publicly available online (Berliner Erklärung, 2003). These are *communism*, the idea that there is a common ownership of scientific goods (here data), and *organized skepticism*, the idea that every scientist has the duty to let other researchers scrutinize his or her work (Merton, 1973). Well-documented and openly available datasets allow organized skepticism by enabling the replicability of research (Leonhart and Maurischat, 2004; Evans, 2010; Klein *et al.*, 2013; McNutt, 2014a; Fecher *et al.*, 2016). In this regard, open access to research data is a translation of the Mertonian norms for an ethical and democratic science to the digital age and a potential cure for the replication crisis we currently see in many scientific disciplines (McNutt, 2014a, 2016; Maxwell *et al.*, 2015). And it is for these reasons that open access to research data is currently mandated by prominent funding agencies and science policy makers (Organisation for Economic Co-operation and Development, 2007; Deutsche Forschungsgemeinschaft, 2012).

Numerous prominent German research organizations—we surveyed researchers in Germany—have advocated open access to research data in small science since the Berlin Declaration on Open Access (Max-Planck-Gesellschaft, n.d.). The largest German research funder, the German Research Association, has published a much-read guideline for making data available (DFG, 2009). In addition, the European commission is testing open access to research data as a funding condition in Horizon 2020, a large funding program that aims to foster research in Europe (European Commission, 2013).

Despite the broad consensus that open access to research data benefits the scientific enterprise, it is practiced only in moderation by academic researchers (Campbell *et al.*, 2002; Alsheikh-Ali *et al.*, 2011; Tenopir *et al.*, 2011; Enke *et al.*, 2012). The low willingness of researchers to provide access to the data underlying their research has been identified as a major problem for the scientific enterprise (Tenopir *et al.*, 2011; Enke *et al.*, 2012; Andreoli-Versbach and Mueller-Langer, 2014). Alsheikh-Ali *et al.* (2011) reviewed 500 research articles published in the 50 journals with the highest impact factor. They found that of the 50 journals, 44 (88%) had a data availability policy. Yet, of the 500 assessed papers, only 47 papers (9%) included a full dataset which was deposited online. Vlaeminck and Herrmann (2015) examined 346 journals in business and economics and found that only 49 journals mentioned a data availability policy (14.2%). Only few research funders actually require data management plans (Borgman, 2012). And in cases where open access to research data is successfully practiced, it is often a top-down community effort (Sawicki *et al.*, 1993), a cooperative project (Abazajian *et al.*, 2009), or an institutional service (for example, panel data). Which means that Merton's norms are a world away from today's academic practice.

So far, the discrepancy between the ideal of open access to research data and the actual behaviour of research professionals has only been assessed for single research fields and had not been subjected to any kind of cross field comparison. Based on a survey among 1564 mainly German researchers across disciplines, we examine if this problem holds true for multiple disciplines and what explains it. We find that data withholding can best be explained by strategic reputation considerations and misaligned incentives in the academic reward system. We conclude that a rethinking of the academic reward structure to offer more formal

recognition for intermediate products, such as data, code and consultation/transfer services, would have a positive impact on research collaboration and ultimately the integrity of the scientific enterprise as a whole.

Open access to research data: an untapped driver of scientific progress

In her speech “Open infrastructures for Open Science” at the European Federation of Academies of Science and Humanities Annual Meeting, Neelie Kroes, then Vice President of the European Commission responsible for the Digital Agenda, stated: “[...] sharing data, and having the forum to openly use and build on what is shared, are essential to science. They fuel the progress and practice of scientific discovery” (Kroes, 2012). The proclaimed benefit of open access to research data can essentially be explained by a qualitative and a quantitative perspective on scientific progress. It enables data-driven replication studies and thereby contributes to the integrity of scientific research (quality) and it allows asking new research questions based on openly available data and thereby utilizing synergies and preventing unnecessary duplication in the collection of data (quantity).

Enabling data-driven replication studies. Open access to research data can be a safeguard for scientific fraud and the dissemination of erroneous results by enabling data-driven replication studies. They can be regarded as a translation of Merton's organized skepticism to the digital age and, therefore, a driver for a sustainable scientific progress.

Issues with replicability are present across many fields of empirical research. In a study that aimed at replicating the effects found in 100 psychological studies, only 39% of the main effects could be replicated (Open Science Collaboration, 2015). The state of replicability in psychology has led some to speak of a “crisis of reproducibility” and prompted new quality assurance policies such as the pre-registration of experiments (Maxwell *et al.*, 2015). A study published in *Nature* reported the failure to replicate significant experiments in cancer research in 47 out of 53 cases (Begley and Ellis, 2012). In a study that aimed to replicate 18 effects published in two top economics journals (*American Economic Review* and the *Quarterly Journal of Economics*) between 2011 and 2014, the researchers were not able to find a significant effect in the same direction as the original study in seven of their replications (Camerer *et al.*, 2016).

It very-well be the case that this “crisis of replicability” is not a new phenomenon and that it only became apparent due to the increasing digitization of research. Furthermore, not being able to replicate a result does not necessarily mean that the results itself is incorrect (Fecher *et al.*, 2016). Nonetheless, the issues regarding replicability uncovered in these studies prompt academia to rethink its quality assurance procedures.

Data-driven replications are considered a potential solution (McNutt, 2014b; Fecher and Wagner, 2016). Duvendack *et al.*, (2015) differentiate between four types of replication studies: (a) *narrow replications* using the same data and methods as the replicated article, (b) *wide replications* using the same methods but different data, (c) *reproductions* using the same data but different methods, and (d) *replications that use new data and new methods*. In times of increasingly data-intensive research (Hey *et al.*, 2009) and ever increasing publication rates, replication studies using the same data as the primary investigator (types a and c) are gaining in importance. They allow other researchers to detect erroneous analyses and data manipulation. Moreover, they involve lower transaction costs for the replicator since the datasets are ideally already available (Fecher *et al.*, 2016). These

kinds of replications greatly benefit from open access to data from published results.

Besides the issue of data availability, replication efforts fail due to inadequate data documentation. In a study from 2009, Ioannidis *et al.* (2009) investigated 18 published and peer-reviewed research articles on microarray studies. Even though all articles featured some form of data, the authors were only able to reproduce two studies. In all other cases, crucial information was missing. In addition, replication studies are not very attractive for the replicator. As Park (Park, 2004) states, replication studies are rarely conducted “because [they are] difficult to successfully accomplish and [carry] more risk than potential reward for both the replicator and the originator of the research.” The little reward refers to the fact that replication studies are rarely published and that falsifications could infuriate influential members of the community in question.

Allowing new research questions based on openly available data. Open access to research data contributes to a more efficient resource allocation. It has the potential to minimize the overall collection effort and allows researchers to ask new research questions based on old data.

A prominent example of a project in which data sharing increased the speed of discovery in a whole research field is the Human Genome Project. The project was established in 1990 as an international research alliance whose aim it was to sequence the entire human genome by 2005 (Sawicki *et al.*, 1993). The alliance succeeded in its quest as early as 2003. A core reason for the early completion were the so-called Bermuda Principles, which state that data collected within the research alliance had to be annotated according to agreed standards and publicly archived within 24 hours after collection (Collins *et al.*, 2003). By defining documentation standards and obligating researchers to archive data, the Human Genome Project succeeded in sequencing the genome earlier than expected. The sequencing is considered as one of the greatest scientific achievements of all times. It has provided the basis for thousands of medical studies of inherited disorders.

Not only does open access to research data reduce the overall collection effort and spark new research, it also enables (new) methods such as semantic text analyses or meta analyses that aggregate many different datasets. In evidence-based medicine, for instance, meta analyses with individual patient data are considered the gold standard (Thomas *et al.*, 2014). In cancer research, a field that is increasingly using individual therapeutic methods, meta analyses enable more comprehensive factor analyses.

The discrepancy between the ideal and real life practice. Combined the examples of replication studies, the avoidance of data duplications, and novel research approaches show that there lies great potential in the reuse of openly available data from small science, which remains almost untouched (Cragin *et al.*, 2010). A study among 1329 environmental scientists from the Data Observation Network for Earth (DataOne) identified the lack of access to data from other researchers as the major obstacle to progress in the field (Tenopir *et al.*, 2011). Half of the respondents state that their own research has suffered because they could not access data from others. In the same study, 46% of the researchers stated that they do not archive their data electronically. Only 6% said that they have stored data publicly at least once in the past. Campbell *et al.* (2002) surveyed 1240 genetic researchers from the 100 US universities that receive the most funding from the National Institutes of Health (NIH). Forty-seven per cent of the surveyed researchers reported that they have

been denied data by colleagues despite contacting them personally. Andreoli-Versbach and Mueller-Langer (2014) studied the data sharing behaviour of 488 randomly selected economists. They found that only 12 (2.46%) provided open access to data that could be directly reused.

Research questions. There is a broad consensus that open access to research data benefits the scientific enterprise as a whole. Still, the potential of open access to data is utilized only superficially. This raises the question of what explains data withholding and how science policy could mandate open access data without restricting scientific freedom. To contribute to this discussion, we developed five research questions that we address with this article:

- *RQ1: What is the researchers' opinion on open access to data?* This question explores the view academic researchers have towards open access to research data.
- *RQ2: How and with whom have they shared data in the past?* Open access to research data can be understood as a particular form of data sharing. This research question explores with whom researchers share their data, and hence, which forms of data sharing are common in small science.
- *RQ3: What factors influence open access to research data?* This question explores whether there are any parameters that have an influence on researchers' behaviour in regards to open access to data.
- *RQ4: What are barriers and enablers for open access to data?* This question aims to unravel the specific aspects that either stop researchers from providing open access to their data or that promote this form of data sharing.
- *RQ5: For what purpose do researchers use secondary data?* The last question explores use cases for secondary data. Here, we want to understand what researchers would use openly available data for and if these use cases are in line with the literature.

Methods

Our standardized online survey covers questions on data collection, data management and data sharing/withholding practices of academic researchers from a broad variety of disciplines. The survey was conducted in October and November 2014.

Survey instrument. To diminish the perceived intrusiveness, fear of disclosure and social desirability effects we use an online survey (Tourangeau and Yan, 2007). The survey contains closed multiple-choice questions as rating scales. It covers questions on sociodemographics, the individual working context of the researcher, publication preferences, common impediments and incentives for sharing data, and expectations for using secondary data (Fecher *et al.*, 2014).

The full survey can be found on the project's GitHub page (<https://github.com/data-sharing/persistent/tree/master/dsa-03/>). The survey instrument is based on a previous study, consisting of a systematic review of data sharing studies and a secondary data user survey (Fecher *et al.*, 2015b).

We conducted two pretest rounds, the first with researchers on the usability and comprehensibility of the survey, and the second with experts on data archiving and data reuse. The pretests led to minor changes in the wordings of the questions and to a shorter survey. We contacted every faculty head of 60 German universities and asked them to distribute the survey to researchers in their faculty. We selected the universities based on the number of students and chose the 20 largest, the 20 smallest and 20 medium-sized ones. Additionally, we contacted researchers from the four biggest German research organizations, the Max Planck Society, the Leibniz Association, the Helmholtz Association and the Fraunhofer Gesellschaft, and uploaded a link to the survey on our project website and on the website of the German Data Forum. In our mails to the faculty heads and in the introductory text of the survey, we specifically addressed researchers that work with data. That being said, our sample is a convenience sample and not representative of the entire population of academic researchers in Germany or worldwide.

Dataset. Overall, 2661 people started the questionnaire, but not all respondents finished it. We excluded respondents who did not answer any questions about their status, employer and discipline and those who had answered <20% of the questions. We were left with 1564 valid entries—which represents about 59% of all

respondents who started the survey (Fecher *et al.*, 2016). Eighty-eight per cent of the respondents work in German institutions, while 12% of the respondents work in other countries. The relatively high number of researchers outside Germany in the sample can be explained by respondents reached via mailing lists and website postings. The average age of the respondents is 38 years. Figure 1 shows the composition of our sample by academic status and disciplinary background.

We derived the categories for the scientific disciplines from the “Statistische Jahrbuch” (Germany’s statistical yearbook) (Statistisches Bundesamt and Statistisches Bundesamt, 2016). In Table 1 we compare the numbers for the six disciplines (natural science, social science, humancscience, engineering, humanities and agricultural science) in total and separately for professors and researchers (with or without a PhD). The comparison highlights a disciplinary bias in the data. We over-sampled in particular the social sciences (31% in our sample versus 15% in the statistical yearbook) and the natural sciences (34% versus 25%). *Vice versa*, human sciences (12% versus 27%) and engineering (8% versus 18%) are underrepresented.

Variables and binomial regression. The discussion in the results section is mostly based on univariate frequency tables. We used a 5-point Likert scale that was depicted with equal spaces between the response options (Bentler and Chou, 1987).

In addition, the appendix contains further tables including two binominal logistic regression models (Supplementary Appendix Table A.2). We prepared the binominal regressions out of an exploratory interest, which means that we did not optimize the models. Even if the models are not optimized, they provide more information than the sole frequency tables because they enable us to control for multiple external factors.

The choice for the dependent and independent variables derives from our systematic review on academic data sharing that we conducted in preparation of the survey (Fecher *et al.*, 2015b). Based on the systematic review, we designed a framework for the discussion of data sharing in academia including aspects like the perspective of the data producer, the secondary data user, or the influence of the community and legal systems. This framework also identifies enablers and barriers for data sharing. In the survey, we focused on those factors that concern the data producer.

The first model uses the dependent variable “willingness to share data with a broad audience” and the second model uses “has shared data with a broad audience in the past.” We differentiate between these two since the actual willingness to disclose a dataset has been shown to differ from the actual researcher behaviour. Regarding the independent variables, we include various factors, identified in the systematic review. These factors comprise the status of the researcher (for example, student, researcher or professor), the discipline, the experience in using secondary data, the research aims, structural knowledge regarding data sharing, opinions on data sharing and the kind of data (qualitative, quantitative, sensitive) that is mostly used. Furthermore, we control for sex and age. For those independent variables, which concern opinions, we assessed approval or rejection of a statement by grouping the responses into two categories.

Results

To make the results of our study more accessible we structured them by reference to the five research questions we introduced earlier.

What is the researchers’ general opinion on open access to data? Despite the fact that open access to research data is widely considered a way of fostering scientific progress and although it is promoted by science policy makers, few researchers make their data publicly available. To see how researchers view open access to data, one question battery targeted the researchers’ general opinion. Table 2 summarizes the results to this set of questions.

The results show a general consensus in the research community that open access to data benefits academic research and that researchers should make their data publicly available. Eighty-three per cent of the respondents agree that openly available research data is a major contribution to scientific progress. Seventy-six per cent of the respondents say that other researchers should make their data publicly available. The majority of the researchers in our survey sees no disadvantage in making data publicly available. Seventy-four per cent disagree that they are deterred from publishing articles if a journal requires the publication of data. Seventeen per cent of the researchers agree with the statement that sharing data brings more disadvantages than advantages. While the results may be influenced by social desirability, the clarity of the responses shows that open access to research data is considered beneficial for research in general and is not considered detrimental to the individual researcher who shares data.

How and with whom have researchers shared data in the past?

Open access to research data can be understood as one form of data sharing. Therefore, this research question explores with whom researchers share their data, and hence, which particular forms of data sharing are common in small science. For this purpose, we asked the researchers if they themselves have already shared data with others and if so, with whom. Table 3 summarizes the responses to that question.

The results show that most researchers have shared data in the past, although most have done so selectively. Across disciplines, 13% of all respondents stated that they have shared research data publicly at least once in the past. In contrast, 16% of our respondents stated that they have never shared data with other researchers. This result is surprising, since previous research reported numbers below 10% for never having shared research data (Campbell *et al.*, 2002; Tenopir *et al.*, 2011). The numbers regarding the scope of sharing vary considerably. For example, 58% of the respondents state that they have shared data with

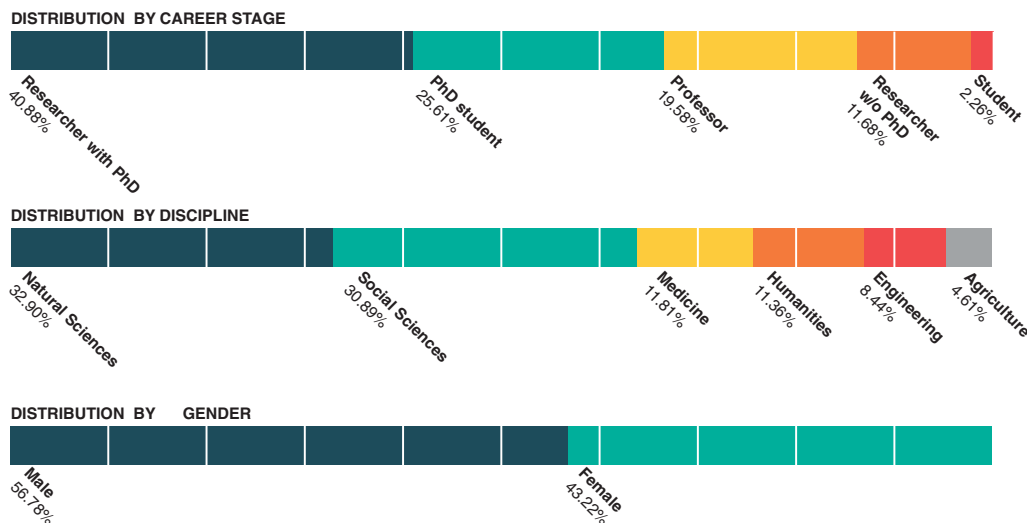


Figure 1 | Sample composition by career stage, discipline and gender.

colleagues whom they know personally, but only 6% of all respondents stated that they have shared data with commercial researchers. Data sharing is already common practice among researchers. However, it mainly occurs among colleagues that know each other, for example, in joint research projects. In contrast, open access to data is a practice that is far less common.

There are also interesting results regarding the discipline and gender. For example: 28% of the social scientists and economists say that they have never shared data, whereas only 9% of natural scientists never shared data. 25% of the female researchers have never shared data, compared with 13% of the male researchers (see Supplementary Appendix Table A.1).

What factors influence open access to research data? This question explores whether there are any parameters that have an influence on the researchers' behaviour in regards to open access to data. Out of an exploratory interest, we tested factors that influence the sharing behaviour of a researcher. The sharing behaviour is defined here as the willingness to make data publicly available and the actual experience of making data publicly available.

The results of the regression analysis show that there is no significant influence of a researchers' age on his/hers willingness to make data publicly available and on whether the respondent has done so in the past. The same holds true for the status a researcher has reached and his/her experience in academic research. This result is surprising, since previous research suggested that senior researchers are more willing to share data than their younger colleagues (Tenopir *et al.*, 2011).

Knowledge regarding data management has a positive effect on the data sharing behaviour. Researchers who know how to make data publicly available are significantly more willing to do so ($P < 0.008$) and they are more likely to have done so in the past ($P < 0.001$). Researchers who have used secondary data before are significantly more willing to make it available publicly ($P < 0.052$). The results show that knowledge on data management and previous experience are good indicators for data sharing.

Publishing preferences have an effect on the data sharing behaviour. Researchers that value open access highly are significantly more likely to make their data publicly available ($P < 0.001$; see Supplementary Appendix Table A.2). The fact that a journal has a high reputation and a fast publishing process have no significant influence ($P = 0.627$, respectively, $P = 0.110$).

There are also interesting descriptive results. Overall "only" 56% of the respondents say that they know where to find secondary data and 50% know where and how to publish research data. Natural scientists know more about where and how they can publish data than respondents from other disciplines. Researchers from engineering and agriculture know far less about where they can find secondary data for their research than respondents from other disciplines (see Supplementary Appendix Table A.3). Seventy-nine per cent of the respondents in our survey agree that reputation/impact is important when publishing research results. It is by far the most important criteria when publishing, followed by a fast publishing process with 52% agreement, and lastly open access, with an agreement level of 39% (see Supplementary Appendix Table A.4).

What are barriers and enablers for open access to data? With this question we want to unravel the specific aspects that either stop researchers from providing open access to their data or that promote this form of data sharing.

By a wide margin, the most prominent concern researchers have is that "other researchers could publish before me" (80%; see Table 4); the test variable "to publish before sharing" (85%) is the second most important enabler for making data publicly available (see Supplementary Appendix Table A.5). Hence, most researchers would keep a dataset to themselves until they are sure they have published every aspect of it. Furthermore, 46% say the concern that their data could be misinterpreted prevents them from making it publicly available. Only a few researchers (12%) are concerned about being "criticized or falsified." The latter is interesting since potential falsifications have been widely hypothesized as a reason not to make data available (Longo and Drazen, 2016a).

The majority of the respondents (73%) disagree that criticism or falsification would prevent them from making data publicly available (Table 4). The effort that went into the data collection is considered a barrier to making data available for 59% of the researchers. They agreed with the statement "I would not share my data if the data collection required considerable effort." However, researchers are more likely to withhold their data if the sharing itself is a major effort (Table 4). Overall the effort to make data available is the second biggest impediment we could find (only behind "other researchers could publish before me").

Interestingly, despite the demand for more formal recognition, only a minority regards "Co- authorship" as a motivator. Across

Table 1 | Comparison of the german sub-sample from our survey and the basic population of researchers using numbers from Germany's statistical yearbook

	Our data			Statistical yearbook		
	Researcher	Professor	Total	Researcher	Professor	Total
Natural science	215	59	385	42,683	9026	53,334
Social science	161	71	344	16,850	11,185	30,791
Humanscience	81	25	132	52,283	3789	56,810
Engineering	35	17	87	28,030	9463	38,418
Humanities	60	25	117	17,925	6449	28,104
Agricultural science	29	19	58	3213	1008	4345
Total	481	216	1123	160,984	40,920	211,802
Natural science	44.70%	27.31%	34.28%	26.51%	22.06%	25.18%
Social science	33.47%	32.87%	30.63%	10.47%	27.33%	14.54%
Humanscience	16.84%	11.57%	11.75%	32.48%	9.26%	26.82%
Engineering	7.28%	7.87%	7.75%	17.41%	23.13%	18.14%
Humanities	12.47%	11.57%	10.42%	11.13%	15.76%	13.27%
Agricultural science	6.03%	8.80%	5.16%	2.00%	2.46%	2.05%

Table 2 | General perception towards open access to data

	Observation	IRR* (%)	Mean	Standard deviation	[95% Conf. interval]	Responses in per cent:					
						Strongly disagree	2	3	4	Agree completely	
On a scale from 1 to 5; 1 = "Strongly disagree"—5 = "Agree completely"											
"Researchers should generally publish their data"	1491	95	4.10	1.00	4.05	4.15	1.95	5.90	16.57	31.32	44.27
"I have more disadvantages than advantages when I share my data with others"	1419	91	2.32	1.16	2.26	2.38	28.61	33.19	21.35	11.49	5.36
"It deters me from publishing when a journal requires the publication of my data"	1412	90	1.94	1.14	1.88	2.00	48.30	25.71	13.95	7.86	4.18
"Openly available research data is a great contribution to scientific progress"	1449	93	4.34	0.94	4.29	4.39	1.73	3.80	11.32	25.05	58.11

*IRR = item-response-rate

disciplines, 79% of the respondents say that data citation would motivate them to make data available to others; only 10% say it would not. Financial support for sharing data is considered a motivator by only 17% of the respondents and rejected by 65% of the researchers (see Supplementary Appendix Table A.5). The results indicate that researchers seek more formal recognition.

Despite the fact that the overall pattern in how the respondents evaluate statements are similar in all disciplines, there are some interesting differences. For example: 58% of the medical researchers see a co-authorship as an enabler for sharing but only 23% of the social scientists and economists and 21% of the humanities scholars. This result indicates that co-authorship for sharing data is accepted among medical researchers in contrast to all other disciplines (see Supplementary Appendix Table A.6).

For what purpose do researchers use secondary data? Our last question explores use cases for secondary data. Here, we want to understand what researchers would use openly available data for. If there was no demand for secondary data, the argument for open research data in small science would be invalid. We, therefore, asked researchers if they have used secondary data before and for which purpose they used it. Across all disciplines, 69% of researchers have worked with secondary data before.

Social science has the highest rate of secondary data users, with 78% of social scientists using this kind of data followed by natural science at 69%. The lowest rate of secondary data users can be found in agricultural science with 58% (see Supplementary Appendix Table A.7). The majority of the researchers across disciplines prefer to use secondary data to address novel research questions than to verify results. This assessment is especially true for the social sciences, where 69% of the respondents would use secondary data for novel research questions and only 17% to replicate results. In comparison, in medicine, 48% of the respondents would use secondary data for novel research questions and 38% for replicating results (see Supplementary Appendix Table A.8).

Discussion and policy implications

In the remainder of this article, we will highlight and interpret the findings of our study and reflect on potential policy measures. In all, 5.1 and 5.2 focus on theoretical derivations and 5.3 on practical implications of our results.

The social dilemma of academic data sharing. In many ways, open access to research data complies with good scientific practice in an increasingly digitized research environment. It increases *collaboration* and fosters *discovery* by treating data as a communitarian asset. This is in line with Merton's ideal of communalism as a characteristic of a democratic science (Merton, 1973). By enabling data-driven replication studies, it, furthermore, corresponds to Merton's organized skepticism, and Popper's critical rationalism and the notion that researchers should not accept anything as a final truth (Popper, 2002). It is, therefore, understandable that science policy makers are increasingly mandating open access to research data (for example most recently the EU competitiveness council (Enserink, 2016).

However, the results of the survey confirm a mismatch between the expected societal benefits of open access to research data and the individual researcher's behaviour. Despite the fact that academia would be collectively better off if everyone shared, researchers rarely make their primary data openly available. This can be described as a social dilemma, a situation in which the individual's behaviour is at odds with public interest and social benefit. To this extent, the survey explains the individual attitudes and behaviours of academic researchers towards data sharing but

Table 3 | “Have you ever shared your research data with others?”

Binary, multiple choice; 1 = selected; 0 = not selected	Observation	Mean	Standard error	[95% Conf. interval]		Responses in per cent:	
						Selected	Not selected
“Yes, with researchers who I know personally”	1564	0.58	0.12	0.56	0.61	58.06	41.94
“Yes, with researchers within my institute or my organization”	1564	0.49	0.13	0.47	0.52	49.36	50.64
“Yes, with researchers who work on similar topics”	1564	0.40	0.12	0.38	0.43	40.35	59.65
“Yes, with all non-commercial researchers”	1564	0.14	0.08	0.13	0.16	14.45	85.55
“With commercial researchers”	1564	0.06	0.01	0.05	0.07	5.88	94.12
“Yes, with the public”	1564	0.13	0.01	0.11	0.15	13.11	86.89
“No”	1564	0.16	0.01	0.14	0.18	16.30	83.70

also reveals systemic shortcomings of the existing scientific reward structure.

What explains this social dilemma? To a large degree, data withholding and selective sharing can be explained by strategic publication considerations. Our results suggest that—unlike article publications—research data does not count as a stand-alone research output, but is rather seen as a raw material for article publications. More specifically: the main impediment to open access to research data is the concern that other researchers could publish with it first. This is particularly troublesome, since the value of open access to data is recognized, not only by prominent science funders, but also by the researchers in our survey.

Furthermore, a concern about transparency (for example, criticism or falsifications) is not seen as an impediment by most researchers. This is surprising since fear of falsification has widely been discussed as a major impediment for sharing data (Tenopir *et al.*, 2011; Acord and Harley, 2012; Longo and Drazen, 2016b). Our result could to some degree be related to social desirability issues that are always present in survey studies. It is also worth pointing out that previous studies focussed on other research areas (mainly the United States) and single fields (for example, environmental studies). Differences in academic cultures and the competitive nature of a research environment might also hold explanations for these differences. It would, therefore, be worthwhile to compare practices in different research areas, for example, regarding the effects of national policies, cultural differences or the competitiveness of the research environment.

Understanding academia as a reputation economy. Merton (1957) explained that a researcher’s objective is to establish priority of discovery by being the first to report a finding. In exchange the researcher receives a reward in the form of peer recognition. This notion of reputation as a driver for scientific exchange has been recognized in ethnographic and sociological works such as Bourdieu’s *Homo academicus* (Bourdieu, 1984) and Luhman’s conception of science as an autopoietic system. The latter considers reputation as the only science-specific reward (Luhmann, 1990). Similarly, the economics of science try to explain researchers’ behaviour with formal peer recognition (for example, citation).

In this tradition, academia can be described as a reputation economy, a knowledge exchange system that is driven by each individual’s desire to accumulate reputation and to achieve—following Bourdieu’s field theory (Bourdieu, 1993)—a desirable level of social status (Fecher *et al.*, 2015a). In such a system, researchers share information and knowledge first and foremost if it pays off in form of peer recognition. This stands in contrast to other economies, most notably those where individuals strive to accumulate money. The traditional scientific autonomy, as it is

common in Germany and large parts of the Western world, is conducive to the relative insularity of this self-reproducing social system (Luhmann, 1990). The researchers in our sample even reject financial rewards in return for making their research data publicly available.

Understanding academia as a reputation economy helps to explain data withholding and selective sharing. Reputation in the current academic system is for the most part tied to high impact (journal) publications (Leahey, 2008; Schlöpfer and Schneider, 2010), whereas intermediate products, such as data, have little to no exchange value. This explains why researchers strategically withhold datasets until they have published every aspect of them, as sharing too early would risk a competitive disadvantage. The reputation metaphor also offers an explanation for the reuse behaviour of the surveyed researchers: They favour pursuing new research questions with secondary data instead of conducting replication studies, as these are harder to publish (Hamermesh, 2007). Understanding the researchers’ behaviour in these terms also shines a light on why datasets of published articles are poorly documented (Alsheikh-Ali *et al.*, 2011): A researcher has no particular incentive to invest time and effort to make sure a dataset is easily reusable (Acord and Harley, 2012).

From the point of view of a somewhat rational researcher in a reputation economy, data can be described as a powerful resource that, if shared without exploitation, loses its exclusive value in a competitive and self-referential system for reputation. This is why open research data remains, an ideal professed but not practiced (Andreoli-Versbach and Mueller-Langer, 2014). Data is, in Bourdieu’s, terms an objectified cultural capital that has little direct exchange value in a competitive field for recognition (Bourdieu, 1983). It is reasonable to assume that more researchers would make their data publicly available if this behaviour generated a reputational payoff.

Policy implications: towards a market for research data.

Understanding academia as a reputation economy can help designing policies that are in line with the academic reward system, academic autonomy and Merton’s ideal of a democratic science. Making data management and data publication a worthwhile activity for researchers would, therefore, steer the academic system towards more scientific progress. A simple rule of three that is based on the idea of a reputation economy would be:

- (a) **Increasing reputational benefits** of open access to data for the individual researchers,
- (b) **Reducing transaction costs** and legal uncertainties of open access to data,
- (c) **Increasing market transparency** by making open access to research data more visible to members of the research community

Table 4 | Barriers: "I would not share my data ..."

	Observation	IRR* (%)	Mean	Standard deviation	[95% Conf. interval]	Responses in per cent.					
						Strongly disagree	2	3	4	Agree completely	
On a scale from 1 to 5; 1 = "Does not apply at all"—5 = "Applies completely"											
"... if other researchers could use my data to publish before me"	1409	90.09	4.25	1.17	4.19	4.31	5.61	5.75	8.73	18.03	61.89
"... if others could criticize or falsify my work"	1402	89.64	1.99	1.18	1.93	2.05	46.65	25.89	15.05	6.70	5.71
"... if the data could be misinterpreted"	1383	88.43	3.16	1.35	3.08	3.23	15.84	17.43	21.04	26.75	18.94
"... if the data collection required considerable effort"	1392	89.00	2.57	1.26	2.50	2.63	26.51	23.42	24.71	17.53	7.83
"... if a major effort were required to share the data"	1403	89.71	3.55	1.19	3.49	3.61	7.84	11.69	21.53	35.64	23.31

*IRR = item-response-rate

Increasing reputational benefits. In science today, reputational gains are mainly bound to text-based publications. It is likely that more researchers would invest time and effort in data curation and publication if data citations and appropriate impact scores would count stronger in funding applications recruitment decisions. A rarely discussed option for the promotion of data management and publications could be awards for good datasets. Best paper awards are common practice across disciplines. Awarding good datasets, for instance at key conference, would highlight the importance of data and provide researchers that have invested efforts with data publication with peer recognition (Friesike *et al.*, 2015).

Reducing transaction costs. Insufficient data documentation has often been discussed as a major obstacle for replication studies (Ioannidis *et al.*, 2009; Fecher and Wagner, 2016; Peters *et al.*, 2016; Munafò *et al.*, 2017). A way to increase the reproducibility of research could, therefore, be the implementation and adoption of data documentation and availability policies by scientific journals and funding agencies. Furthermore, scientific communities could set standards for data documentation. This would help individual researchers to clearly understand what a good dataset looks like and how a documentation should be carried out.

Increasing market transparency. This third rule applies mainly to the research infrastructure providers. Making it visible who shares data and what datasets are reused will increase the value that researchers attribute to shared datasets. Today, it is often difficult to figure out the status of a dataset. To increase the visibility would signal researchers that the community appreciates their efforts. In its simplest form this could be a label on each publication highlighting that all underlying data is publicly available.

With this study we looked at individual researchers. Yet, it needs to be pointed out that researchers are embedded in a cultural and normative setting: their research environment. Merton himself (Merton, 1985) argued that his norms for ethical and modern research are not just individualistic values but need to be reflected on an institutional and organizational level. And indeed if we look at the work that researchers perform many actions are not directly tied to journal publications. They review grant applications, they edit journals, they organize conferences, they share lecture notes among many other things. These forms of community service exist and researchers perform them because their research communities have taught them that this is an appropriate behaviour. And therefore, apart from creating incentives for the individual researcher to make data available, a "market for research data" inevitably implies changes in academic organizations. This concerns, for instance, the integration of data management in the curricula at universities, the recruitment policies of research organizations, and funding for data management. This, furthermore, entails top-down directives with respect to funding conditions and data management policies. Ideally, also research organizations perceive a value in terms of producing good datasets, for example because it enhances its reputation and the chances for funding.

As the study shows, the ideal of open science, as implied by Merton's norms of ethical research—in particular communism and organized skepticism—is not the most evident behavioural frame for academic researchers and academic institutions. Instead, arguments for open science are trumped by individual reward considerations that are often tailored to publications. To steer academia towards more openness, it needs to shift focus towards intermediate products, such as data.

References

- Abazajian KN *et al.* (2009) The seventh data release of the sloan digital sky survey. *The Astrophysical Journal Supplement Series*; **182** (2): 543–558.
- Acord SK and Harley D (2012) Credit, time, and personality: The human challenges to sharing scholarly work using Web 2.0. *New Media & Society*; **15** (3): 379–397.
- Alsheikh-Ali AA, Qureshi W, Al-Mallah MH and Ioannidis JPA (2011) Public availability of published research data in high-impact journals. *PLoS ONE*; **6** (9): e24357.
- Andreoli-Versbach P and Mueller-Langer F (2014) Open access to data: An ideal professed but not practised. *Research Policy*; **43** (9): 1621–1633.
- Begley CG and Ellis LM (2012) Drug development: Raise standards for preclinical cancer research. *Nature*; **483** (7391): 531–533.
- Bentler PM and Chou C-P (1987) Practical issues in structural modeling. *Sociological Methods & Research*; **16** (1): 78–117.
- Berliner Erklärung (2003) *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities*, http://openaccess.mpg.de/67605/berlin_declaration_engl.pdf, accessed 13 September 2016.
- Borgman CL (2012) The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*; **63** (6): 1059–1078.
- Bourdieu P (1993) *The Field of Cultural Production: Essays on Art and Literature*. Columbia University Press: New York, pp 29–73.
- Bourdieu P (1983) Ökonomisches Kapital, kulturelles Kapital, soziales Kapital. In: Kreckel R (ed). *Soziale Ungleichheiten. Soziale Welt Sonderband 2*. Göttingen, Germany, pp 183–198.
- Bourdieu P (1984) *Homo Academicus*. Les Éditions de Minuit: Paris, France.
- Camerer CF *et al.* (2016) Evaluating replicability of laboratory experiments in economics. *Science*; **351** (6280): 1433–1436.
- Campbell EG *et al.* (2002) Data withholding in academic genetics: Evidence from a national survey. *JAMA*; **287** (4): 473–480.
- Collins FS, Morgan M and Patrinos A (2003) The human genome project: Lessons from large-scale biology. *Science*; **300** (5617): 286–290.
- Cragin MH, Palmer CL, Carlson JR and Witt M (2010) Data sharing, small science and institutional repositories. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*; **368** (1926): 4023–4038.
- Deutsche Forschungsgemeinschaft. (2012) *Die Digitale Transformation Weiter Gestalten – Der Beitrag Der Deutschen Forschungsgemeinschaft Zu Einer Innovativen Informationsinfrastruktur für die Forschung*; Positionspapier der Deutschen Forschungsgemeinschaft Deutsche Forschungsgemeinschaft (DFG): Bonn, Germany.
- DFG WLI (LIS). (2009) *Empfehlungen Zur Gesicherten Aufbewahrung und Bereitstellung Digitaler Forschungsprimärdaten*. Richtlinie: Bonn, Germany.
- Duwendack M, Palmer-Jones RW and Reed W (2015) Replication in economics: A progress report. *Econ Journal Watch*; **12** (2): 164–191.
- Enke N, Thessen A, Bach K, Bendix J, Seeger B and Gemeinholzer B (2012) The user's view on biodiversity data sharing — Investigating facts of acceptance and requirements to realize a sustainable use of research data —. *Ecological Informatics*; **11**, 25–33.
- Enserink M (2016) In dramatic statement, European leaders call for ‘immediate’ open access to all scientific papers by 2020. *Science*, <http://www.sciencemag.org/news/2016/05/dramatic-statement-european-leaders-call-immediate-open-access-all-scientific-papers>, accessed 14 June 2016.
- European Commission. (2013) *Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020*, http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf, accessed 9 September 2016.
- Evans JA (2010) Industry collaboration, scientific sharing, and the dissemination of knowledge. *Social Studies of Science*; **40** (5): 757–791.
- Fecher B and Wagner GG (2016) A research symbiont. *Science*; **351** (6280): 1405–1406.
- Fecher B *et al.* (2014) Data Sharing in Academia Questionnaire, <https://doi.org/10.5684/dsa-02>, accessed 4 April 2017.
- Fecher B *et al.* (2015a) A Reputation Economy: Results from an Empirical Survey on Academic Data Sharing. *DIW Berlin Discussion Paper 1454*, <http://dx.doi.org/10.2139/ssrn.2568693>.
- Fecher B (2015b) What drives academic data sharing? *PLoS ONE*; **10** (2): e0118053.
- Fecher B *et al.* (2016) Data Sharing in Academia Dataset. SOEP Socio-Economic Panel Study, <https://doi.org/10.5684/dsa-03>, accessed 4 April 2017.
- Fecher B, Fräßdorf M and Wagner GG (2016) *Perceptions and Practices of Replication by Social and Behavioral Scientists: Making Replications a Mandatory Element of Curricula Would Be Useful*. DIW Discussion Papers, Berlin: DIW Berlin, http://www.diw.de/documents/publikationen/73/diw_01.c.531685.de/dp1572.pdf.
- Friesike S *et al.* (2015) Reputation instead of obligation: forging new policies to motivate academic data sharing. *LSE Impact Blog*, <http://blogs.lse.ac.uk/impactofsocialsciences/2015/06/02/reputation-instead-of-obligation-new-policies-to-motivate-academic-data-sharing/>, accessed 4 January 2016.
- Hamermesh D (2007) *Replication in Economics*. National Bureau of Economic Research: Cambridge, MA.
- Hey AJG, Tansley S and Tolle KM (2009) *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Microsoft Research: Redmond, WA.
- Ioannidis JPA *et al.* (2009) Repeatability of published microarray gene expression analyses. *Nature Genetics*; **41** (2): 149–155.
- Klein T, Kopp J and Rapp I (2013) Metaanalyse mit Originaldaten: Ein Vorschlag zur Forschungssynthese in der Soziologie. *Zeitschrift für Soziologie*; **42** (3): 222–238.
- Kroes N (2012) *Opening Science Through e-Infrastructures*. Rome, Italy, http://europa.eu/rapid/press-release_SPEECH-12-258_en.htm, accessed 19 March 2015.
- Leahy E (2008) Overseeing research practice: The case of data editing. *Science, Technology, & Human Values*; **33** (5): 605–630.
- Leonhart R and Maurischat C (2004) Meta-Analysen auf Primärdatenbasis? Probleme und Lösungsansätze. *Zeitschrift für Evaluation*; (1): 21–34.
- Longo DL and Drazen JM (2016a) Data sharing. *New England Journal of Medicine*; **374** (3): 276–277.
- Longo DL and Drazen JM (2016b) Data sharing. *New England Journal of Medicine*; **374** (3): 276–277.
- Luhmann N (1990) *Die Wissenschaft der Gesellschaft*; Suhrkamp-Taschenbuch Wissenschaft Suhrkamp: Frankfurt am Main, Germany.
- Max-Planck-Gesellschaft. (n.d.) Signatories Berlin Declaration. *Open Access Max-Planck-Gesellschaft*, <http://openaccess.mpg.de/319790/Signatories>, accessed 30 December 2015.
- Maxwell SE, Lau MY and Howard GS (2015) Is psychology suffering from a replication crisis? What does ‘failure to replicate’ really mean? *American Psychologist*; **70** (6): 487–498.
- McNutt M (2014a) Reproducibility. *Science*; **343** (6168): 229–229.
- McNutt M (2014b) Reproducibility. *Science*; **343** (6168): 229–229.
- McNutt M (2016) #IAmAResearchParasite. *Science*; **351** (6277): 1005–1005.
- Merton RK (1957) Priorities in scientific discovery: A chapter in the sociology of science. *American Sociological Review*; **22** (6): 635.
- Merton RK (1973) *The Sociology of Science: Theoretical and Empirical Investigations*. University of Chicago Press: Chicago, IL.
- Merton RK (1985) *Entwicklung und Wandel von Forschungsinteressen: Aufsätze zur Wissenschaftssoziologie*; 1. Aufl. Wissenschaftsforschung, Beratung Suhrkamp: Frankfurt am Main, Germany.
- Munafò MR *et al.* (2017) A manifesto for reproducible science. *Nature Human Behaviour*; **1** (1): 21.
- Open Science Collaboration. (2015) Estimating the reproducibility of psychological science. *Science*; **349** (6251): aac4716–aac4716.
- Organisation for Economic Co-operation and Development. (2007) *OECD Principles and Guidelines for Access to Research Data from Public Funding*. OECD: Paris, France.
- Park CL (2004) What is the value of replicating other studies? *Research Evaluation*. **13** (3): 189–195.
- Peters I, Kraker P, Lex E, Gumpenberger C and Gorraiz J. (2016) Research data explored: An extended analysis of Citations and Altmetrics. *Scientometrics*; **107** (2): 723–744.
- Popper KR (2002) *Logik der Forschung*. Julius Springer: Vienna, AS.
- Sawicki MP, Samara G, Hurwitz M and Passaro E (1993) Human genome project. *The American Journal of Surgery*; **165** (2): 258–264.
- Schläpfer F and Schneider F (2010) Messung der akademischen Forschungsleistung in den Wirtschaftswissenschaften: Reputation vs. Zitierhäufigkeiten: Messung der akademischen Forschungsleistung. *Perspektiven der Wirtschaftspolitik*; **11** (4): 325–339.
- Statistisches Bundesamt and Statistisches Bundesamt. (2016) *Statistisches Jahrbuch Deutschland 2016*.
- Tenopir C *et al.* (2011) Data sharing by scientists: Practices and perceptions. *PLoS ONE*; **6** (6): e21101.
- Thomas D, Radji S and Benedetti A (2014) Systematic review of methods for individual patient data meta-analysis with binary outcomes. *BMC Medical Research Methodology*; **14** (1): 79.
- Tourangeau R and Yan T (2007) Sensitive questions in surveys. *Psychological Bulletin*; **133** (5): 859–883.
- Vlaeminck S and Herrmann L-K (2015) Data policies and data archives: A new paradigm for academic publishing in economic sciences? *Stand Alone*; 145–155.

Data availability

The data and supplementary material used for this study are available through the research data center (RDC) of the German Socio-Economic Panel (SOEP), which also ensures the long-term preservation of the data (DOI: 10.5684/dsa-03); <https://github.com/data-sharing/persistent/tree/master/dsa-03/>. To protect the interests of the participants of the survey, the RDC requires researchers to sign a non-disclosure agreement before accessing the data. Researchers are then allowed to analyse the full dataset and to publish the data in aggregated form, again protecting the privacy of the participants.

Additional information

Supplementary Information: accompanies this paper at <http://www.palgrave-journals.com/palcomms>

Competing interests: The authors declare that they have no competing financial interests.

Reprints and permission information is available at http://www.palgrave-journals.com/pal/authors/rights_and_permissions.html

How to cite this article: Fecher B, Friesike S, Hebing M and Linek S (2017) A reputation economy: how individual reward considerations trump systemic arguments for open access to data. *Palgrave Communications*. 3:17051 doi: 10.1057/palcomms.2017.51.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017