

## Factors influencing the data sharing behavior of researchers in sociology and political science

Zenk-Möltgen, Wolfgang; Akdeniz, Esra; Katsanidou, Alexia; Naßhoven, Verena; Balaban, Ebru

Postprint / Postprint

Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

### Empfohlene Zitierung / Suggested Citation:

Zenk-Möltgen, W., Akdeniz, E., Katsanidou, A., Naßhoven, V., & Balaban, E. (2018). Factors influencing the data sharing behavior of researchers in sociology and political science. *Journal of Documentation*, 74(5), 1053-1073. <https://doi.org/10.1108/JD-09-2017-0126>

### Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

**gesis**  
Leibniz-Institut  
für Sozialwissenschaften

### Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Mitglied der  
  
Leibniz-Gemeinschaft

*Zenk-Möltgen, Wolfgang; Akdeniz, Esra; Katsanidou, Alexia; Naßhoven, Verena; and Balaban, Ebru (2018), Journal of Documentation:*

## **Factors influencing the data sharing behavior of researchers in sociology and political science**

### **Abstract**

#### **Purpose**

Open data and data sharing should improve transparency of research. This article investigates how different institutional and individual factors affect the data sharing behavior of authors of research articles in sociology and political science.

#### **Design/methodology/approach**

Desktop research analyzed attributes of sociology and political science journals (n=262) from their websites. A second dataset of articles (n=1011, published 2012-2014) was derived from ten of the main journals (five from each discipline) and stated data sharing was examined. A survey of the authors used the Theory of Planned Behavior to examine motivations, behavioral control and perceived norms for sharing data. Statistical tests (Spearman's rho, Chi-square) examined correlations and associations.

#### **Results**

Although many journals have a data policy for their authors (78% in sociology, 44% in political science), only around half of the empirical articles stated that the data was available, and for only 37% of the articles could the data be accessed. Journals with higher impact factors, those with a stated data policy, and younger journals were more likely to offer data availability. Of the authors surveyed, 446 responded (44%). Statistical analysis indicated that authors' attitudes, reported past behavior, social norms and perceived behavioral control affected their intentions to share data.

#### **Research limitations/implications**

Less than 50% of the authors contacted provided responses to the survey. Results indicate that data sharing would improve if journals had explicit data sharing policies but authors also need support from other institutions (their universities, funding councils, professional associations) to improve data management skills and infrastructures.

#### **Originality/value**

This article builds on previous similar research in sociology and political science and explains some of the barriers to data sharing in social sciences by combining journal policies, published articles, and authors' responses to a survey.

## Introduction

Research data are a new currency in science. In recent years, more and more recommendations have been published on the citing of research datasets and the principal investigators' responsibility when creating data. Examples of discipline-independent recommendations on data sharing have been published by DataCite, the Research Data Alliance, the FORCE11 initiative, and the OECD (Brase *et al.*, 2015; Rauber *et al.*, 2015; Berman *et al.*, 2014; Data Citation Synthesis Group, 2014; OECD, 2013), and more specifically for the social sciences by the International Association for Social Science Information Services and Technology and the German Data Forum (IASSIST SIGDC, 2012; Rat für Sozial- und Wirtschaftsdaten, 2016). Open science strategies are being adopted by policy makers and research institutions. They are increasingly supporting the idea of open access not only to published materials but also to open research data and open code. The scientific potential of open data is enormous in terms of the replication of research results (King, 2006, p. 120; Agosti *et al.*, 2017), the re-purposing of old data (Moss *et al.*, 2015), and the increase of "effectiveness, productivity, and reproducibility" (Gregory *et al.*, 2018, p. 1) for science. In addition, the promise of open data is to allow better cooperation across academia, government, and the private sector (Groves, 2017).

Academic journals are one of the driving forces of science. They act as gatekeepers for the quality and form of scientific publications. In recent years academic journals in various disciplines have started adopting data sharing policies to promote transparency and replication. For that, access to empirical data and their documentation is necessary to enable progress based on existing knowledge. In reality, replication often suffers from a lack of data availability and poor or even non-existent documentation (Dewald *et al.*, 1986; Abrams *et al.*, 2014; Tenopir *et al.*, 2011).

While data and data sharing play an increasingly important role in empirical research, this is still not fully appreciated by researchers across all disciplines. The study by Kim & Zhang (2015) investigated data sharing decisions in relation to motivation, attitudes, perceived norms and controls in the STEM disciplines (science, technology, engineering, and mathematics). It reveals the significant effects of attitudinal beliefs (perceived career benefit/risk, perceived effort) on STEM researchers' data sharing behavior. The most important factors determining data sharing by researchers have also been found to lie in academic rewards (Kling and Spector, 2002) and recognition (Altman, 2016; Kankanhalli *et al.*, 2005). This leads to inter-disciplinary efforts to create incentives for data sharing like Altmetrics (Piwowar, 2013) or the DataCite efforts to link dataset authors with publications (Mongeon *et al.*, 2017). Further factors influencing data sharing behavior of researchers are potential risks including concerns about losing publication opportunities, necessary effort (Kim and Stanton, 2012; Tenopir *et al.*, 2011; Savage and Vickers, 2009) as well as demographic factors (Enke *et al.*, 2012).

Tenopir *et al.* (2011) showed in their inter-disciplinary study that even though scientists agree that the lack of data sharing is a major impediment to progress in science (60%), they do not make their data electronically available to others (46%). Specifically for disciplines like health research or social sciences, privacy requirements are an additional impediment for data sharing. A major dilemma is that while scientific disciplines as a whole benefit if data are made available as a public good, sharing data entails significant disadvantages for individual researchers. The hypothesized reason for researchers not sharing their data is that "their efforts and perceived risks outweigh the

potential individual benefits they expect from data sharing” (Fecher *et al.*, 2015, p. 19). Other publications investigating reasons for not sharing hint at the perceived ownership of data by the primary investigators, a need for control but also a significant time and resources problem when it comes to data management (Fecher *et al.*, 2015; Gherghina and Katsanidou, 2013; Zenk-Möltgen and Lepthien, 2014b).

Data sharing seems to be based on altruism, incentives and the availability of supportive structures (Horton and Katsanidou, 2011). High-quality data management and documentation have also been shown to result in better quality research and more transparent research sections (Katsanidou *et al.*, 2016). Based on these findings the current investigation seeks to gain a better understanding of what drives data sharing, examining institutional factors (analysis of journal data policies and published articles) and individual factors (survey among authors of articles) in the fields of sociology and political science.

### **Previous research**

This current article on data sharing is a continuation of two previous studies which investigated the impact of journal policies on authors’ data sharing behavior for sociology and political science separately (Zenk-Möltgen and Lepthien, 2014b; Gherghina and Katsanidou, 2013). The data of these previous studies (Zenk-Möltgen and Lepthien, 2014a; Gherghina and Katsanidou, 2014) were combined and extended, and a survey about data sharing attitudes and behaviors among the authors of the selected articles was added (see Fig. 1). The resulting data for the current study makes it possible to analyze three levels equally for sociology and political science: journal policies, authors’ behavior in their published articles, and authors’ behavior and attitudes measured by the survey.

[Insert Fig. 1 here]

### **Figure 1: Research design**

The study “Data availability in political science journals” (Gherghina and Katsanidou, 2013) investigated the data policies of international peer-reviewed political science journals and the extent to which journals adopt these guidelines. Of the 120 journals in the final unit of analysis, 18 journals were found to have a journal-specific data policy, and one followed a common publisher policy. An e-mail survey of the other 102 journals revealed that most had no data policy and that few editors thought that data policies were necessary and intended to implement them soon). The availability of a data policy was also highly correlated with the age of the journal, type of readership, and the impact factor.

The factors which influence data sharing in certain sociological journals were analyzed in the study “Data sharing in sociology journals” (Zenk-Möltgen and Lepthien, 2014b). Only seven of the 140 sociology journals investigated were found to have an explicit data policy. 94 journals referred to a common policy for data sharing by a publishing association. Statements about the availability of data sets made by the authors of a total of 581 articles in five of these German and international sociology journals were analyzed in detail between 2012 and 2013 and compared with the actual availability of the data. Stated and proven data availability was highest in the “American Sociological Review” (ASR). This was also the highest ranked journal by impact factor in this

selection. 41% of ASR the data sets were found, followed by the “American Journal of Sociology” (AJS) with 26%. The study shows that good progress was being made concerning the sharing of data used for papers published in sociology journals. However, this progress is tempered by low rates of data availability, information or data citations and lack of data accessibility. High professional standards, explicit data policies and the recommendation of data repositories and data archives were found to encourage data sharing.

### **Conceptual framework and research questions**

Our research aims to identify the factors which encourage or impede the sharing of data by empirical sociology and political science researchers. The behavior of data sharing in the context of this article means that researchers enable access to their primary research data underlying their articles, e.g. by using a public repository or an archive. We also include the sharing of syntax codes into our analysis since researchers use to work with data because the codes are closely related to the datasets itself. Between different specific methods of data curation, manipulation, or analysis the respective results may vary. Thus, the availability of syntax code together with the data is relevant to the possibility of replication.

By using the two previous studies as the basis of our research and extending the data to equally cover journals and articles from sociology and political science we are able to investigate the institutional factors that may be influential for data sharing behavior (see fig. 1). Thus, section 1 examines data sharing policies of selected journals and the articles published. This includes authors’ (stated and actual) data sharing and citation behavior using desktop research methods.

The following research questions regarding the institutional factors are asked:

**(Q1 – journal data policy)** How many sociology and political science journals have a data policy? Which factors are related to this?

**(Q2 – stated data availability in article)** In how many articles do authors say that the underlying data are available? Are there differences between disciplines, over time, by data policy, journal age, journal language, or impact factor?

**(Q3 – checked data availability in article)** In how many articles can data be accessed? Does the amount of confirmed data availability vary by discipline, over time, by data policy or journal age/language/impact factor?

Section 2 reports on a survey on researchers’ views on data sharing that is added as a completely new component to the two previous studies, aiming to explain the individual factors influencing data sharing behavior. The Theory of Planned Behavior (Ajzen, 1991; 2005; Fishbein and Ajzen, 2010; Ajzen and Fishbein, 1980) is used as a theoretical framework to collect information about self-reported behavior and attitudes towards data sharing. TPB has successfully been used to explain a large variety of behaviors, especially in the area of health-related behaviors (Godin and Kok, 1996), but also for other areas like economics, environmental behaviors, and others (Harding *et al.*, 2007; Peng *et al.*, 2014). The TPB is therefore well suited to be applied to investigate the individual motivations regarding data sharing behavior (Kim and Stanton, 2012).

The TPB states that behavior is the result of a volitional and conscious decision to behave in a specific way. The best predictor of human behavior, according to Ajzen (1991), is the behavioral intention, which influences a person’s readiness to perform a

specific behavior. The intention is determined by attitudes towards behavior, subjective norm and perceived behavioral control. The intention is also related to individuals' past behavior. Social norms consist of two sub-constructs: an injunctive social norm, reflecting perceptions of what others approve of or think one ought to do, and a descriptive social norm, reflecting perceptions of other people's behavior and what is typical or average. Perceived behavioral control is conceptualized as a latent construct consisting of perceived capacity (belief that one is capable of performing a behavior) and perceived autonomy (perceived control/having the relevant skills and abilities). Research on behavioral decisions can also include past behavior or demographic factors, such as education, age or gender.

The following research questions were formulated based on the TPB to cover the individual factors for data sharing:

**(Q4 – attitude towards data sharing)** Is there a relationship between researchers' attitude toward data sharing and their behavioral intention to do so?

**(Q5 – subjective norm)** Is there a relationship between perceived social pressure to share data with others and behavioral intention to do so?

**(Q6 – perceived behavioral control)** Is there a relationship between perceived behavioral autonomy and self-efficacy to be able to share data and researchers' intention to do so?

**(Q7 – past behavior)** Is researchers' past data sharing behavior related to an increased intention to share data in the future?

**(Q8 – demography)** Do researchers' demographic characteristics affect the intention to share data?

Together, the analysis of institutional factors in section 1 and the investigation of individual factors in section 2 result in a better understanding of data sharing behavior of researchers in sociology and political science.

### **Methodology and data collection**

To address the research questions in section 1, two datasets were built covering two levels: the journal level and the article level nested within journals (see fig. 1). At the higher level, information about journals was gathered by desktop research inspecting each journal's web page information. This dataset covers sociology and political science journals listed in the Web of Science (Thomson Reuters, 2013). Variables coded include the name and the age of the journal, whether it is an English-language publication, the frequency of issues per year, the Web of Science impact factor, and the existence of a data policy for authors. To identify differences in contrast to the previous studies, we included a variable for the previous presence of a data policy (2010/2012). The lower level data set focuses on articles nested in all the issues from 2012 to 2014 of ten of the journals selected above. These articles were selected from five journals for each discipline, covering high ( $\geq 2$ ) and low impact factor journals, and also German-language journals for sociology. Only empirical articles were considered as relevant. The variables coded for each article included whether the authors state that the underlying data is available and whether the data could be found using the given reference. Authors' e-mail addresses were also collected to conduct the survey later. The data can be accessed at the datorium repository (Zenk-Möltgen *et al.*, 2017).

The next step was to design a questionnaire to measure researchers' motivations to share research data and syntax codes using the Theory of Planned Behavior (TPB). Following the steps described by Ajzen (2002, 2005), questions were constructed for each element of the theory: attitude, perceived norm (injunctive social norm and descriptive social norm), perceived behavioral control (capacity and autonomy), data sharing intention and past behavior. Information about respondents' views on behavioral outcomes, normative referents, and control factors was taken from a broad qualitative study by Kim and Stanton (2012). Kim and Stanton conducted qualitative interviews with researchers in STEM disciplines on data sharing methods, types of generated and shared data, motivations for and barriers to data sharing, work environments and demographic information which are assumed to be similar for the social sciences.

Demographic and individual trait questions (including gender, age, religion, position in the academic career, country, and working sector) were used to obtain additional information about respondents. The measurement instrument incorporating all the key constructs was then constructed. Research items were defined applying the TACT principle (description of behavior in terms of target, action, context and time) (Ajzen, 2002). The questionnaire was then piloted on a sample of 11 researchers who were not part of the targeted sample, and a slightly modified version of the instrument was created for the actual survey (the used survey items can be found in appendix A).

The online questionnaire was launched on May 11, 2015, and was closed on July 7, 2015. The pretest was conducted on April 23, 2015. It was important for the survey that it took place not too late after the previous research, so that the authors would still remember their motivations for sharing data or not. The online survey system was provided by SoSci Survey (Leiner, 2014). For each article, one author was asked to take part in the online survey, usually the corresponding or first author. If an author was not reachable (email undeliverable or an absence notice was returned) one of the potentially available co-authors was contacted instead. A total of four reminders were sent to respondents on May 19, May 28, June 24 and June 30, 2015. The number of respondents increased with each reminder that was sent. The survey data can also be accessed at the datorium repository (Zenk-Möltgen *et al.*, 2017).

Descriptive statistics and significance tests were used to analyze the data. Cramer's V based on Pearson's Chi-square test was used for the evaluation of strength and statistical significance of associations between variables at nominal scale, e.g. availability of data and the presence of a data policy. Spearman's rho was used to determine the level of correlation between ordinal variables. For this type of data the cases are being ranked in a range of categories, e.g. the journal age or impact factor. Pearson's r was used to analyze linear correlations between variables of interval or ratio scale, e.g. between the numbers of articles of different type. It was also used to analyze correlations between the constructs of the Theory of Planned Behavior, assuming that measurement with Likert-scales can be treated as interval scales.

Cramér's V can have values between zero and plus one, and higher values indicate a stronger association between the variables. Spearman's rho and Pearson's r can have any value between minus one and plus one whereas a value of zero indicates no relationship, a value of less than zero indicates a negative association and a value of higher than zero indicates a positive association.

## Data analysis and results

### *Analysis of journal policies*

The focus of interest regarding the journals was the availability of data policies (addressing Q1). If a policy was found, different codes were used for each journal-specific policy, policy produced by a journal publisher or by any other association. The desktop research on the existence of a data policy for the sociology and political science journals reveals that only 16 of 142 sociology journals (11.3%) had an explicit data policy in 2016. This number is higher if an account is taken of the many journals which follow the data policies of their publishers or an academic association: in sociology 110 of 142 journals (77.5%) and political science 53 of the 120 selected journals (44.2%) had some sort of data policy for their authors. There are big differences in the details regarding the content of the policies: Some are recommendations, others are strong advice, and only a few of the policies have mandatory data sharing rules for authors. The way those rules are enforced differ very much between the journals, but it can be seen that only a small number of journals state on their web pages how adherence by authors is being checked. Many journals only refer to a general recommendation on data sharing, or to ethical standards in science provided by scientific associations or publishing companies (American Sociological Association, 1999; Elsevier, 2017).

We investigated which journal characteristics correlate with the existence of a data policy. The following correlations were found to be significant: English language journals are more likely to have a data policy (Spearman's  $\rho = .68$ ,  $p < .001$  in sociology,  $\rho = .37$ ,  $p < .001$  in political science). Journals with a higher impact factor are more likely to have a data policy ( $\rho = .53$ ,  $p < .001$  for sociology,  $\rho = .27$ ,  $p < .01$  for political science). The more issues a journal publishes each year, the more likely it is to have a data policy ( $\rho = .20$ ,  $p < .05$  in sociology,  $\rho = .30$ ,  $p < .001$  in political science). Journal age was not related to data policy. These results confirm the analysis of Gherghina and Katsanidou (Gherghina and Katsanidou, 2013) and Zenk-Möltgen and Lepthien (Zenk-Möltgen and Lepthien, 2014b) with the data for 2010/2012.

A change over time can be observed when comparing the results from 2010/2012 with the results from 2016. By 2016, the number of sociology journals with an explicit journal, publisher or association data sharing policy increased from 102 (71.3%) to 110 (77.2%). The number of sociology journals with no data sharing policy at all decreased from 39 (27.3%) to 32 (22.5%). This trend is even more pronounced in political science. Between 2010 and 2016, the number of journals with a journal or publisher policy rose from 19 (15.8%) to 53 (44.2%). The number of political science journals without a data policy fell from 101 (84.2%) to 67 (55.8%). There is a clear and growing trend for political science and sociology journals to adopt data sharing policies.

To answer our first research question about data policies of journals, we can conclude:

**(Q1)** A high number of journals do have a data policy of some kind in place. This is more common for journals in sociology than in political science. As many sociology journals refer to a policy created by their association of publishers, the rate is much lower for numbers of explicit journal policies in sociology. Thus we can state that awareness of data sharing in political science journals is higher than in sociology journals.

### *Analysis of published articles*



Table 1 shows the number of analyzed articles and the availability of a data policy for each journal and discipline. Altogether, 1011 empirical articles from the ten journals in the years 2012 to 2014 were analyzed: 676 for political science and 335 for sociology. When comparing the scientific fields to analyze data availability, it has to be taken into account that 75.6% of political science articles appeared in journals with a data policy (AJPS, ES, PA); this is only the case for 43% of sociology articles (ASR, SM). This results in a strong relationship between the factors of data policy availability and discipline for the analyzed articles ( $X^2 = 104.39$ ,  $p < .01$ , Cramer's  $V = .32$ ), making it difficult to differentiate between those factors when analyzing data availability. This also means that absolute numbers across disciplines must be interpreted with caution because more political science than sociology articles was selected from journals with a data policy. A similar relationship can be seen for language and discipline, as no articles in political science were selected from German language journals, but 32.2% of the articles in sociology.

[Insert Table 1 here]

From the authors of empirical articles in the study, only 56.5% stated that their data are available for other researchers, but only for 36.6% of the articles, the data could be found (see table 2). The rates of stated data availability (addressing Q2) and checked data availability (addressing Q3) were compared to identify any significant differences in journal attributes (using Spearman's rho/Cramér's  $V$ ). The stated availability of data correlated positively with the age of the journal ( $\rho = .10$ ,  $p < .01$ ), its impact factor ( $\rho = .29$ ,  $p < .001$ ), and was associated with the existence of a data policy in the journal, as expected ( $X^2 = 53.48$ ,  $V = .23$ ,  $p < .001$ ). Table 2 shows that the percentages of stated data availability are nearly identical for sociology (58.2%) and political science (55.6%), so that no significant association is found ( $X^2 = .61$ ,  $V = .02$ ,  $p = .435$ ). Data were actually more likely to be available the younger the journal was ( $\rho = -.11$ ,  $p < .001$ ) and the higher the impact factor was ( $\rho = .22$ ,  $p < .001$ ). Actual availability of data was associated with the presence of a data policy in the journal ( $X^2 = 79.08$ ,  $V = .28$ ,  $p < .001$ ), language ( $X^2 = 44.19$ ,  $V = .21$ ,  $p < .001$ ), and discipline ( $X^2 = 48.80$ ,  $V = .22$ ,  $p < .001$ ). Data are more frequently found in English-language journals than in German journals, and data from political science journals were more often found than in sociology journals (44.0% vs. 21.5%).

[Insert table 2 here]

Over time the number of articles which state that data are available and those actually providing such data had increased. Table 2 shows this development equally for sociology and political science. A difference found between the disciplines is that the rate of actually available data for sociology is much lower than for political science, but the increase over time can be seen in both disciplines (political science: 36.2%, 44.1%, 51.5%, sociology: 11.8%, 20.7%, 31.3%). These findings correspond with those of Gherghina and Katsanidou (2013) for the field of political science. Providing data with empirical articles has become more common in recent years.

For the answers to the research questions on data sharing behavior of authors of published journal articles this means:

**(Q2)** In more than half of the empirical articles, authors state that the data is available for other researchers. There was no significant difference between sociology and political science in this respect.

**(Q3)** Data could actually be accessed in one third of the empirical research articles, less than the proportion of articles (over a half) that stated data was available for other researchers. In sociology, the proportion of articles in which data was actually available was much less than for political science.

#### *Analysis of the survey among researchers*

After looking into the institutional factors of journal attributes for data sharing behavior of authors, we investigated the individual factors by analyzing the survey data. After excluding missing data, 446 authors participated in the survey, covering 44.1% of all articles. Table 1 shows the response rates by journal in more detail. An overview of the variables and the full text of questions and answers in the questionnaire can be found in Appendix A.

Demographic information for the respondents was collected, including gender, age, position, religion, working sector, and country (see table 3). Among the 446 participants, there were 260 males (58.3%) and 103 females (23.1%), mostly aged 30-39 (42.4%). The distribution of gender in political science is more similar to the overall distribution, whereas in sociology there are more female respondents (30.5%). Furthermore, researchers are somewhat younger in sociology than in political science.

The majority of the respondents are employed by a university, a college or a technical university and live in the United States or Germany. Disciplinary differences can be found according to the stated country: the number of sociology researchers who live in Germany is more than three times higher than the number of political scientists, which is due to the selection of two German-language journals in the sample for sociology.

[Insert Table 3 here]

#### *How often do researchers share data or code?*

Before analyzing the relationships between the TPB constructs, we look at the self-reported data sharing behavior. Respondents were asked to indicate the number of empirical journal articles they had published and the number of data and syntax codes they had shared and cited during the last three years. The answers show (see table 4) that respondents typically published five or six articles (average of 5.68 for sociology and 6.04 for political science) and all respondents together published 2474 empirical articles in total. Data was shared according to the respondents for 1102 of these articles; code was shared for only 775 articles.

Table 4 shows the results of the total number of articles published by each group of a particular position in the academic career and the number of the shared data and shared syntax code with each article. We computed the “data sharing rate” as the number of articles with shared data divided by all articles of the respondents, and there are obvious differences between the groups of academic career positions. The highest data-sharing

rate shows the group of Associate Professors/Readers/Senior Lecturers (56.0%), followed by the group of doctoral students (52.5%). The lowest rate was discovered for researchers without a Ph.D. (22.2%), but they show low rates of publishing activity as well. Syntax code sharing is even less common in social and political science. The highest rate shows the group of Assistant Professors / Lecturers (41.1%).

[Insert Table 4 here]

Table 4 also shows the distribution of the scientific field and the number of articles, which were published in sociology and political science, and for which data or syntax codes were shared. For both the data sharing and syntax sharing rate the number of participants in political science is about twice as high as in the field of sociology (53.9% vs. 25.2% for data sharing, 38.2% vs. 17.2% for syntax sharing).

#### *Relationship between past behavior constructs*

To check for relationships between the different answers regarding behavior, we calculated correlation coefficients (Pearson's  $r$ ) amongst all past behavior variables, including past sharing behavior in regards to research data and syntax code. The following table 5 presents correlations between the stated number of shared and cited research data or syntax codes and the frequency with which respondents said they had shared data and syntax codes over the last three years.

[Insert Table 5 here]

At first glance, there is a high correlation ( $r = .81$ ,  $p < .001$ ) between the number of published articles which share research data and the number of published articles which share syntax code. There is also a correlation between the number of times syntax is shared and the amount of research based on self-collected data ( $r = .38$ ,  $p < .001$ ). This means that data and syntax codes are most likely to be shared when self-collected data is analyzed and if the respondent states that research data ( $r = .41$ ,  $p < .001$ ) and syntax codes ( $r = .56$ ,  $p < .001$ ) have been shared over the last three years. The findings further indicate that there is a strong positive correlation between the number of published articles and the number of published articles based on secondary data ( $r = .70$ ,  $p < .001$ ) as well as the number of published articles based on self-collected data ( $r = .65$ ,  $p < .001$ ). Furthermore, the analysis shows a strong positive correlation between past data sharing and past syntax sharing behavior ( $r = .60$ ,  $p < .001$ ). The more often researchers shared data, the more often they also shared their syntax codes when they published empirical journal articles.

We can conclude that authors are more willing to share data the more articles they publish and the greater the proportion of data they collect themselves. Whereas authors' syntax code sharing is positively correlated with data sharing, the number of articles published, and the proportion of data authors collect themselves.

#### *Attitudes, perceived behavioral control, social norm, past behavior, and intention*

To investigate the influence of researchers' motivations on data sharing, we included survey items based on constructs of the Theory of Planned Behavior (TPB). Table 6 gives the descriptive statistics and correlations between all TPB components. Respondents

were asked about their level of agreement on a seven-point scale. The constructs comprise the researchers' attitudes towards data sharing, the perceived social norm (injunctive and descriptive aspects), the perceived behavioral control (capacity and autonomy aspects), and intention to share data. Past data sharing behavior was also included into the analysis since Ajzen and Fishbein (1980) state, that behavior which is shown in the past will most likely influence the intention of a person to show that specific behavior again in the future.

[Insert Table 6 here]

The analysis reveals that there is a strong correlation between reported past behavior and the intention to share data ( $r = .70$ ,  $p < .001$ ), and a medium strong correlation with social norm (injunctive,  $r = .44$ ,  $p < .001$  and descriptive,  $r = .54$ ,  $p < .001$ ) and attitude towards data sharing ( $r = .51$ ,  $p < .001$ ). The intention to share data is strongly correlated with attitude ( $r = .67$ ,  $p < .001$ ) and capacity aspects of perceived behavioral control ( $r = .69$ ,  $p < .001$ ), and moderately correlated with both aspects of social norm (descriptive and injunctive both  $r = .48$ ,  $p < .001$ ).

Moderate correlations exist between attitude and social norm (injunctive,  $r = .49$ ,  $p < .001$  and descriptive,  $r = .35$ ,  $p < .001$ ) and perceived behavioral control capacity,  $r = .57$ ,  $p < .001$ ). Capacity aspects and social norm ( $r = .48$ ,  $p < .001$  and  $r = .46$ ,  $p < .001$ ) are also significantly correlated.

#### *Differences in sociology and political science for individual motivations on data sharing*

To evaluate researchers' motivations, we analyzed past behavior and the differences according to the disciplinary field of the journal (sociology/political science). The test of the difference between the TPB variables by discipline helps in understanding the differences between the group characteristics.

Past data sharing behavior was significantly associated with type of discipline ( $X^2 = 93.26$ ,  $p < .001$ ,  $V = .47$ ). Most of the respondents (38.7%) who published in sociological journals never shared their research data. In contrast, almost the same percentage (39.5%) of the respondents who published in political journals always shared their research data. Intention to share data was significantly associated with type of discipline ( $X^2 = 88.66$ ,  $p < .001$ ,  $V = .45$ ). More than half of the respondents with a background in political science stated that they always intend to share their research data (54.7%); this is only the case for some of the respondents who published in sociological journals (17.7%). There was a strong association between perceived behavioral capacity and discipline ( $X^2 = 70.17$ ,  $p < .001$ ,  $V = .41$ ). The sociology respondents were about evenly distributed in their confidence in their ability (self-efficacy) to share data (13.1% disagree, 23.4% agree) whereas 56.8% of the political science respondents agreed that they were confident that they could share data (2.5% disagree).

Another important factor regarding data sharing is represented by the perceived social norms which focus on injunctive and descriptive aspects. Both variables were significantly associated with the disciplinary background of the respondents. Respondents who published in political science journals were those who more often agreed fully that others would approve of data sharing (58.7%,  $X^2 = 41.95$ ,  $p < 0.001$ ,  $V = .31$ ) and that others actually share their research data (10.5%,  $X^2 = 75.30$ ,  $p < 0.001$ ,  $V = .42$ ).

Considering all these results, it is not surprising that almost all respondents publishing their articles in political science journals have a positive attitude towards data sharing (72.1%). Although most respondents with sociological journal articles do have a positive attitude towards data sharing as well (47.1%), the frequency of answers with medium-strong agreement is still high. This shows that those respondents are less supportive of data sharing than respondents with political science background ( $X^2 = 45.23$ ,  $p < .001$ ,  $V = .32$ ).

Regarding the research questions about the individual factors influencing data sharing behavior we can say:

**(Q4)** There is a positive correlation between a researcher's attitude toward data sharing and his/her behavioral intention to do so.

**(Q5)** There is a positive correlation between perceived social pressure to share data with others and behavioral intention to do so.

**(Q6)** There is a positive correlation between perceived behavioral autonomy and self-efficacy to be able to share data and a researcher's intention to do so

**(Q7)** Researchers' past data sharing behavior increases with the intention to share data.

**(Q8)** A researcher's demography affects the intention to share data: discipline, country, and to some extent also working sector, position in career and age all play a role in this respect.

## **Discussion**

This study adds its results to the thus far pessimistic literature showing that data sharing practices are still not as wide-spread as they should (Dewald *et al.*, 1986; Abrams *et al.*, 2014; Tenopir *et al.*, 2011). Despite the increasing number of sociology and political science journals adopting data policies, data sharing for researchers is still not a self-evident practice. And when it does happen, it is far from formalized or complete. It should not go unnoticed that only about one-third of the articles where authors stated data availability could stand to the fact check; for the others, the data simply could not be found. Similar to the analysis of Kim/Stanton (2012) and Kim/Zhang (2015) in the area of STEM research, who found that disciplines with well-defined standards and available data repositories enable researchers to practice data sharing, this study shows for sociology and political science that institutional incentives for data sharing need to be increased. However, there is not only a dark side to the findings of this paper. There is indeed a wind of change to be felt, as the results clearly show a slight but steady increase of data sharing practices over time, especially in journals with higher impact factors, with data policies for their authors, and published in English. That comes due to increased awareness of the importance of data availability, especially among political scientists – a new finding that adds to the study of Zenk-Möltgen/Lepthien (2014).

The results in this study took research on the data sharing topic a step further by shedding light on the individual motivations of researchers: Data sharing and syntax sharing practices go hand in hand, as the one increases so does the other. Authors are also more motivated to share if it is their own data they collected. The data sharing behavior reinforces itself; authors that have done it once tend to repeat the practice.

This paper adds to the literature by testing relationships of constructs based on the Theory of Planned Behavior about data sharing behavior in sociology and political science. Indeed we found that data sharing adheres to this theory: The intention to share data is strongly affected by the attitude of the researchers, and they are motivated by

social norms, perceived behavioral autonomy and self-efficacy, which is in line with previous findings (Kim/Stanton 2011). The strongest indicators for data sharing based on this theory showed to be reported past behavior of data sharing of respondents and their intention to do so in the future, and that is an additional insight of this study. Other relevant influences for the intention to share data are the respondent's country, working sector, position in the academic career, and age.

Finally, the study revealed significant differences between political scientists and sociologists, showing that political scientists engage more in data sharing behavior. Even in their intention to engage in data sharing, their perceived capacity, and their attitude towards data sharing sociologists are more reserved than political scientists. Examples of data sharing activities in sociology and political science point in the direction that this might be because sociology more often deals with qualitative interviews or in-depth surveys than political science. These data are more difficult to share due to anonymization problems and the involved ethics of research and at the same time are more difficult to be re-used. Standardized quantitative surveys conducted at a large scale national level (e.g. like national election studies) can much easier be anonymized than surveys of smaller, more special populations. Qualitative studies also often do not have computer code for the analysis, but conduct non-automatic procedures that need to be documented in a different way. For both disciplines the legal ownership of research data needs to be clearly defined for making data sharing happen.

### **Limitations**

The case selection of journals poses a clear limitation to this study. The list of journals used to analyze the existence of data policy was retrieved from the Thomson Reuters Web of Science and their classification of journals was used. Even knowing that the Web of Science covers most of the internationally visible publications in sociology and political science, a selection bias cannot be excluded entirely. The classification also has additional classes, which are more or less related to sociology and political science, as well as sub-classes containing even more journals that could have been selected for analysis. However, the number of analyzed journals has been limited to create a manageable project within our available resources and to focus on the most significant publications in political science and sociology.

The survey data pose an additional set of limitations: First, the data were collected from authors of a limited selection of only ten journals, even when these can be considered to be relevant and broadly acknowledged. Second, the response rate of 44.1% of the survey participants is quite good but is not sufficient to exclude a self-selection of authors. This could happen in a way that researchers responded more often when they were in favor of data sharing. Third, the analyses are based on bivariate correlation or association coefficients, meaning that the analyses only reflect pairwise relationships between the constructs. However, this investigation shows that most of the analyzed concepts are related to each other, and this builds a strong recommendation to conduct further research into individual characteristics related to data sharing of researchers in sociology and political science.

### **Conclusions**

The findings show that there is a gap between the recommendations of policymakers and research institutions and the reality of data sharing in sociology and political science,

even though authors seem to be engaging in the practice more over the last three years. Academic journals can play a major role in closing this gap as these institutions are one of the driving forces of science. Establishing a solution-oriented data policy with clear criteria and implications is a sure step towards a better data sharing practice. Journals have already done this for referencing literature, so they have previous experience in setting the standards.

Journals do not function in a void. There are other institutions that can support this practice and enhance a data sharing culture. Research funding agencies, research councils, academic libraries, and employers (e.g., universities, government bodies, or commercial companies) with their policies can have an effect on data sharing practices and this should be subject to future investigations.

However, all kinds of policies will only be followed by a majority of the authors if additional individual factors are taken into account. The analysis in this paper clearly stated two obstacles: the knowledge about existing infrastructures and the know-how for data sharing. Data sharing infrastructures should offer low-threshold technologies to enable easy data sharing for researchers. In sociology, secure data sharing infrastructures for research data with sensitive information regarding privacy are needed. In addition, the creation of a well-established knowledge-base on how to access and use data infrastructure is not a trivial business. This should be part of the toolkit that young scientists acquire with their overall data skills education. In the same way how universities established courses on quantitative and qualitative methods to train students in data analysis they should also teach data management and data sharing practices.

Data sharing needs to acquire a critical mass to become a common practice. The findings show that researchers are influenced by peers and that their peers need to expect data sharing behavior and practice it themselves. If more formalized ways are offered on how to do data sharing, the community in its clear pursuit of more transparency will comply.

Future work needs to focus on a more in-depth analysis of factors influencing data sharing behavior. The concepts of the TPB have been successfully used in the paper to show individual factors regarding data sharing behavior; they should be analyzed with a full structural equation model in future work to get a deeper understanding of the causal relationships between the different concepts.

[Insert Appendix A here]

## References

- Abrams, S., Cruse, P., Strasser, C., Willet, P., Boushey, G., Kochi, J., Laurance, M. and Rizk-Jackson, A. (2014), “DataShare. Empowering Researcher Data Curation”, *International Journal of Digital Curation*, Vol. 9 No. 1.
- Agosti, M., Ferro, N. and Silvello, G. (2017), “Digital Libraries: From Digital Resources to Challenges in Scientific Data Sharing and Re-Use”, in Flesca, S., Greco, S., Masciari, E. and Domenico Saccà (Eds.), *Comprehensive Guide Through the Italian Database Research Over the Last 25 Years*, pp. 27–41.
- Ajzen, I. (1991), “The theory of planned behavior”, *Organizational Behavior and Human Decision Processes*, Vol. 50 No. 2, pp. 179–211.
- Ajzen, I. (2002), “Constructing a TpB Questionnaire. Conceptual and Methodological Considerations”.
- Ajzen, I. (2005), “Attitudes, Personality and Behaviour”, 2nd Edition, *McRaw Hill International*.
- Ajzen, I. and Fishbein, M. (1980), *Understanding attitudes and predicting social behaviour*.
- Altman, M. (2016), “The Role of Research Funding and Policy Community in Data Citation — Rewards, Incentives, and Infrastructure”, available at: <http://informatics.mit.edu/blog/role-research-funding-and-policy-community-data-citation-%E2%80%94-rewards-incentives-and>.
- American Sociological Association (1999), “Code of Ethics of the ASA Committee on Professional Ethics”, available at: [http://www.asanet.org/sites/default/files/code\\_of\\_ethics\\_aug\\_2017\\_2\\_1.pdf](http://www.asanet.org/sites/default/files/code_of_ethics_aug_2017_2_1.pdf) (accessed 26 January 2018).
- Berman, F., Wilkinson, R. and Wood, J. (2014), “Building Global Infrastructure for Data Sharing and Exchange Through the Research Data Alliance. Guest Editorial”, *D-Lib Magazine*, No. 20(1/2).
- Brase, J., Sens, I. and Lautenschlager, M. (2015), “The tenth anniversary of assigning DOI names to scientific data and a five year history of DataCite”, *D-Lib Magazine*, No. 21(1/2).
- Data Citation Synthesis Group (2014), “Joint Declaration of Data Citation Principles - FINAL”, available at: <https://www.force11.org/group/joint-declaration-data-citation-principles-final> (accessed 13 March 2017).
- Dewald, W.G., Thursby, J.G. and Anderson, R.G. (1986), “Replication in Empirical Economics: The Journal of Money, Credit and Banking Project”, *American Economic Review*, Vol. 76 No. 4, pp. 587–603.
- Elsevier (2017), “Sharing Research Data”, available at: <https://www.elsevier.com/authors/author-services/research-data> (accessed 26 January 2018).
- 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*, Vol. 11, pp. 25–33.
- Fecher, B., Friesike, S. and Hebing, M. (2015), “What Drives Academic Data Sharing”, *PLoS One*, Vol. 10 No. 2, pp. 1–25.



- Fishbein, M. and Ajzen, I. (2010), *Predicting and changing behavior: The reasoned action approach*, Taylor & Francis.
- Gherghina, S. and Katsanidou, A. (2013), “Data availability in political science journals”, *European Political Science*, Vol. 12, pp. 333–349.
- Gherghina, S. and Katsanidou, A. (2014), *Data availability policies in political science journals*, GESIS Data Archive, Data file, available at: <https://doi.org/10.7802/61>.
- Godin, G. and Kok, G. (1996), “The theory of planned behavior: a review of its applications to health-related behaviors”, *American journal of health promotion*, No. 11.2, pp. 87–98.
- Gregory, K., Cousijn, H., Groth, P., Scharnhorst, A. and Wyatt, S. (2018), *Understanding Data Retrieval Practices: A Social Informatics Perspective*, arXiv:1801.04971.
- Groves, R.M. (2017), “The Promise of Collaborative Data Sharing Across Research Sectors”, in David L. Vannette and Jon A. Krosnick (Eds.), *The Palgrave Handbook of Survey Research*, pp. 297–300.
- Harding, T.S., Mayhew, M.J., Finelli, C.J. and Carpenter, D.D. (2007), “The Theory of Planned Behavior as a Model of Academic Dishonesty in Engineering and Humanities Undergraduates”, *Ethics & Behavior*, Vol. 17 No. 3.
- Horton, L. and Katsanidou, A. (2011), “Purposing your survey: archives as a market regulator, or how can archives connect supply and demand? IASSIST Quarterly Winter 2011, pp. 18-23.”, *IASSIST Quarterly Winter 2011*, pp. 18–23.
- IASSIST SIGDC (2012), “Quick Guide to Data Citation”.
- Kankanhalli, A., Tan, B.C.Y. and Wei, K.-K. (2005), “Contributing Knowledge to Electronic Knowledge Repositories. An Empirical Investigation”, *Management Information Systems Research*, Vol. 29 No. 1.
- Katsanidou, A., Horton, L. and Jensen, U. (2016), “Data Policies, Data Management and the Quality of Academic Writing”, *International Studies Perspectives*, Vol. 17 No. 4, pp. 379–391.
- Kim, Y. and Stanton, J.M. (2012), “Institutional and Individual Influences on Scientists' Data Sharing Practices”, *Journal of Computational Science Education*, Vol. 3 No. 1, pp. 47–56.
- Kim, Y. and Zhang, P. (2015), “Understanding data sharing behaviors of STEM researchers. The roles of attitudes, norms, and data repositories”, *Library & Information Science Research*, Vol. 37 No. 3, pp. 189–200.
- King, G. (2006), “Publication, Publication”, *Political Science & Politics*, Vol. 39 No. 1, pp. 119–125.
- Kling, R. and Spector, L. (2002), “Academic Rewards for Scholarly Research Communication via Electronic Publishing”, *CSI Working Paper*.
- Leiner, D.J. (2014), “SoSci Survey. Computer Software”, available at: <http://www.soscisurvey.com> (accessed 18 May 2017).
- Mongeon, P., Robinson-Garcia, N., Jeng, W. and Costas, R. (2017), “Incorporating data sharing to the reward system of science: Linking DataCite records to authors in the Web of Science”, *Aslib Journal of Information Management*, Vol. 69 No. 5, pp. 545–556.
- Moss, E., Cave, C. and Lyle, J. (2015), *Sharing and Citing Research Data: A Repository's Perspective*.

- OECD (2013), *New Data for Understanding the Human Condition: International Perspectives: OECD Global Science Forum Report on Data and Research Infrastructure for the Social Science*.
- Peng, J., Zhi-cai, J. and Lin-jie, G. (2014), “Application of the Expanded Theory of Planned Behavior in Intercity Travel Behavior”, *Discrete Dynamics in Nature and Society*, Vol. 2014 No. 2, pp. 1–10.
- Piwowar, H. (2013), “Altmetrics: Value all research products”, *Nature*, Vol. 493, p. 159.
- Rat für Sozial- und Wirtschaftsdaten (RatSWD) (2016), *Forschungsdatenmanagement in den Sozial-, Verhaltens- und Wirtschaftswissenschaften: Orientierungshilfen für die Beantragung und Begutachtung datengenerierender und datennutzender Forschungsprojekte, RatSWD Output Series*, Berlin.
- Rauber, A., Asmi, A., van Uytvanck, D. and Pröll, S. (2015), “Data Citation of Evolving Data. Recommendations of the Working Group on Data Citation (WGDC)”.
- Savage, C.J. and Vickers, A.J. (2009), “Empirical Study of Data Sharing by Authors Publishing in PLoS Journals”, *PLoS One*, Vol. 4 No. 9.
- Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A.U., Wu, L., Read, E., Manoff, M. and Frame, M. (2011), “Data sharing by scientists: practices and perceptions”, *PLoS One*, Vol. 6 No. 6, e21101.
- Thomson Reuters (2013), “Social sciences citation index. Journal list”, available at: <http://ip-science.thomsonreuters.com/cgi-bin/jrnlst/jlresults.cgi?PC=SS>.
- Zenk-Möltgen, W., Katsanidou, A., Akdeniz, E., Naßhoven, V. and Balaban, E. (2017), *Replication data for: Factors influencing the data sharing behavior of researchers in sociology and political sciences*, available at: <http://doi.org/10.7802/1487>.
- Zenk-Möltgen, W. and Lepthien, G. (2014a), *Data from: Data sharing in sociology journals*, GESIS Data Archive, available at: <https://doi.org/10.7802/65>.
- Zenk-Möltgen, W. and Lepthien, G. (2014b), “Data sharing in sociology journals”, *Online Information Review*, Vol. 38 No. 6, pp. 709–722.

**Data availability in political science journals**  
(Gherghina and Katsanidou, 2013)

- Journal policies**
- Political science
  - 120 journals
  - Data collection: 2011
  - Website analysis and e-mails to the editors

**Data sharing in sociology journals**  
(Zenk-Möltgen and Lepthien, 2014)

- Journal policies**
- Sociology
  - 140 journals
  - Data collection: 2013
  - Website analysis
- Articles of the journals**
- Sociology
  - 5 journals selected with 581 articles
  - published 2012 and 2013
  - 222 empirical articles
  - Article content analysis

**Factors influencing the data sharing behavior of researchers in sociology and political science**

- Journal policies**
- Combination and update of the previous studies
  - Political science and sociology
  - 262 journals
- Articles of the journals**
- Addition of 5 political science journals
  - Addition of articles in 2014
  - 1011 articles of political science and sociology (2012-2014)
- Authors of the articles**
- Online-Survey
  - 446 respondents (political science: 292, sociology: 154)



	Desktop research				Survey	
	262 Journals		1011 Articles		446 Authors responded	
	Political Science	Sociology	Political Science	Sociology	Political Science	Sociology
Journals	120	142	5	5	5	5
Articles			676	335	676	335
Respondents					292	154

Scientific Field	Journal	Year of publication of Journal	5 year Impact-Factor	Data policy (1=yes)	English (1=yes)	N articles (2012-2014)	N articles per field	Response Survey	Response Rate
Sociology	American Sociological Review (ASR)	1936	6.0970	1	1	129		52	40.3%
	American Journal of Sociology (AJS)	1895	4.9120	0	1	83		29	34.9%
	Sociological Methodology (SM)	1969	3.3580	1	1	15		5	33.3%
	Zeitschrift für Soziologie (ZfS)	1972	.7020	0	0	54		36	66.7%
	Kölner Zeitschrift für Soziologie und Sozialpsychologie (KZfSS)	1921	.7010	0	0	54		32	59.3%
<b>Sum</b>							<b>335</b>	<b>154</b>	
Political science	American Journal of Political Science (AJPS)	1973	4.3240	1	1	181		78	43.1%
	Comparative European Politics (CEP)	2003	.8490	0	1	60		22	36.7%
	European Journal of Political Research (EJPR)	1973	2.0650	0	1	105		59	56.2%
	Electoral Studies (ES)	1982	1.2950	1	1	250		103	41.2%
	Political Analysis (PA)	1987	3.4030	1	1	80		30	37.5%
<b>Sum</b>							<b>676</b>	<b>292</b>	
<b>Total sum</b>							<b>1011</b>	<b>446</b>	<b>44.1%</b>

Table 1: Selected journals in sociology and political science

Scientific field		2012	2013	2014	All articles	
Political Science	Stated availability	116	122	135	373	
		54.7%	53.0%	59.0%	55.6%	
		212	230	229	671	
	Total	100.0%	100.0%	100.0%	100.0%	
		Checked availability	76	101	118	295
			36.2%	44.1%	51.5%	44.2%
210	229		229	668		
Total	100.0%	100.0%	100.0%	100.0%		
	Sociology	Stated availability	53	69	73	195
			52.0%	57.0%	65.2%	58.2%
102			121	112	335	
Total	100.0%	100.0%	100.0%	100.0%		
	Checked availability	12	25	35	72	
		11.8%	20.7%	31.3%	21.5%	
102		121	112	335		
Total	100.0%	100.0%	100.0%	100.0%		
	Total	169	191	208	568	
		53.8%	54.4%	61.0%	56.5%	
314		351	341	1006		
Total	100.0%	100.0%	100.0%	100.0%		
	Checked availability	88	126	153	367	
		28.2%	36.0%	44.9%	36.6%	
312		350	341	1003		
Total	100.0%	100.0%	100.0%	100.0%		

Table 2: Data availability by scientific field and publication year of article

Demographic variables		TOTAL	Frequency		TOTAL (%)	Frequency (%)	
			AL	Soc		Pol	Soc
Gender	Male	260	80	180	58.3	52.0	61.6
	Female	103	47	56	23.1	30.5	19.2
	Other	1	0	1	0.2	0.0	0.3
	<i>Missing</i>	82	27	55	18.4	17.5	18.8
Age	Under 29	17	9	8	3.8	5.8	2.7
	30-39	189	70	119	42.4	45.5	40.8
	40-49	91	26	65	20.4	16.9	22.3
	50-59	28	7	21	6.3	4.6	7.2
	60+	20	6	14	4.5	3.9	4.8
	<i>Missing</i>	101	36	65	22.7	23.4	22.3
	Position	Professor	91	33	58	20.4	21.4
Associate Professor		82	19	63	18.4	12.3	21.6
Assistant Professor		100	27	73	22.4	17.5	25.0
Researcher with PhD		53	24	29	11.9	15.6	9.9
Researcher without PhD		10	8	2	2.2	5.2	0.7
Doctoral Student		21	13	8	4.7	8.4	2.7
Graduate Student		2	1	1	0.5	0.7	0.3
Undergraduate Student		1	0	1	0.2	0.0	0.3
<i>Missing</i>		86	29	57	19.3	18.8	19.5
Religion		Christian	106	33	73	23.8	21.4
	Muslim	1	0	1	0.2	0.0	0.3
	Jewish	9	3	6	2.0	2.0	2.1
	Atheist	101	45	56	22.7	29.2	19.2
	Agnostic	98	29	69	22.0	18.8	23.6
	Other	19	7	12	4.3	4.6	4.1
	<i>Missing</i>	112	37	75	25.1	24.0	25.7
Working Sector	University, College, or TU	325	112	213	72.9	72.7	73.0
	Public Research Institute	24	10	14	5.4	6.5	4.8
	Private Research Institute	7	3	4	1.6	2.0	1.4
	Other	2	0	2	0.5	0.0	0.7
	<i>Missing</i>	88	29	59	19.7	18.8	20.2
	Country	United States	162	56	106	36.3	36.4
Germany		77	51	26	17.3	33.1	8.9
United Kingdom		19	0	19	4.3	0.0	6.5
Netherlands		13	3	10	2.9	2.0	3.4
Italy		12	1	11	2.7	0.7	3.8
Switzerland		9	2	7	2.0	1.3	2.4
Sweden		8	0	8	1.8	0.0	2.7
Canada		6	2	4	1.4	1.3	1.4
Belgium		6	1	5	1.4	0.7	1.7
Denmark		5	-	5	1.1	0.0	1.7
Austria		5	3	2	1.1	2.0	0.7
Spain		5	1	4	1.1	0.7	1.4
Other (Frequency $\leq 3$ )		29	5	24	6.5	3.3	8.2
<i>Missing</i>		90	29	61	20.2	18.8	20.9
<b>Total</b>			446	154	292		34.5

Table 3: Demographics of the survey respondents

<b>Discipline</b>	Number of researchers	Average number of published articles	How many empirical articles published?	With how many did you share data?	With how many did you share syntax code?	Data sharing rate	Syntax sharing rate
Sociology	142	5.68	807	203	139	25.2%	17.2%
Political Science	276	6.04	1667	899	636	53.9%	38.2%
<b>Position in academic career</b>							
Professor	91	7.49	682	310	224	45.5%	32.8%
Associate Professor / Reader / Senior Lecturer	82	6.32	518	290	204	56.0%	39.4%
Assistant Professor / Lecturer	100	5.16	516	265	212	51.4%	41.1%
Researcher with PhD	53	4.64	246	58	47	23.6%	19.1%
Researcher without PhD	10	2.70	27	6	3	22.2%	11.1%
Doctoral Student	21	2.81	59	31	22	52.5%	37.3%
<b>Total</b>	418	5.92	2474	1102	775	44.5%	31.3%

Table 4: Data sharing rate by discipline and position in academic career

Construct	n	Mean	SD	Media n	Min, Max	Sum	1	2	3	4	5	6	7	8	9
1 data sharing	418	4.67	2.20	5	{1, 7}		1								
2 code sharing	415	4.27	2.31	5	{1, 7}		.60***	1							
3 articles	418	5.92	4.30	5	{0, 35}	2474	.08	.10	1						
4 articles, self-collected	399	3.20	3.11	3	{0, 25}	1276	.13**	.06	.65***	1					
5 articles, secondary	390	3.27	3.32	3	{0, 27}	1276	.00	.09	.70***	.11*	1				
6 articles, shared data	375	2.94	3.66	2	{0, 45}	1102	.45***	.37***	.50***	.52***	.37***	1			
7 articles, shared code	368	2.11	2.73	1	{0, 16}	775	.41***	.56***	.49***	.38***	.33***	.81***	1		
8 data citing	353	4.59	6.69	3	{0, 50}	1620	.14*	.15**	.38***	.10	.47***	.34***	.26***	1	
9 code citing	354	0.64	2.91	0	{0, 50}	227	.03	.05	.06	.04	.05	.09	.15**	.15**	1

\*Correlations are significant at the  $p < .05$  level. \*\*  $p < .01$  level. \*\*\*  $p < .001$  level.

Table 5: Correlation matrix for past behavior variables



Construct	Mean	SD	1	2	3	4	5	6	7
1 attitude	6.24	1.21	1						
2 norm, injunctive	5.94	1.43	.49***	1					
3 norm, descriptive	3.94	1.67	.35***	.42***	1				
4 control, capacity	5.54	1.81	.57***	.48***	.46***	1			
5 control, autonomy	4.29	2.13	.09	.19***	.08	.32***	1		
6 intention	5.62	1.59	.67***	.48***	.48***	.69***	.20***	1	
7 data sharing	4.67	2.20	.51***	.44***	.54***	.59***	.12*	.70***	1

\*Correlations are significant at the  $p < .05$  level. \*\*  $p < .01$  level. \*\*\*  $p < .001$  level.

Table 6: Correlations among TPB constructs and past behavior

## Appendix A: Variables of the survey dataset

Shortcut	Variable Name	Concept	Question Text
Gender	IF01	Demography	What is your gender? (1=male, 2=female, 3= other)
Age	IF02	Demography	What is your year of birth? (open-ended question)
Position	IF03	Demography	What level have you reached in your academic career? (1 = Professor 2 = Associate Professor/Reader/Senior Lecturer 3 = Assistant Professor/Lecturer 4 = Researcher with PhD 5 = Researcher without PhD 6 = Doctoral Student 7 = Graduate student 8 = Undergraduate student )
Religion	IF04	Demography	What is your religious preference? (1 = Christian, 2 = Muslim, 3 = Jewish, 5 = Atheist, 6 = Agnostic 7 = Other, please specify)
Working Sector	IF06	Demography	For which type of employer do you work? (1 = University, College, or Technical University, 2 = Public Research Institute, 3 = Private Research Institute, 4 = Data Collection Agency/Market Research, 5 = Software Company, 6 = Other, please specify)
Country	IF05	Demography	In which country do you live?
attitude	Q102_01	Attitude towards data sharing	When publishing empirical journal articles it would be good for researchers in general to share research data. (disagree 1- 7 agree)
norm, injunctive	Q201_01	Social Norm (injunctive)	Most researchers who are important to me in my work would approve that I share research data. (disagree 1- 7 agree)
norm, descriptive	Q203_01	Social Norm (descriptive)	Most researchers like me, who publish empirical journal articles, actually share research data. (disagree 1- 7 agree)
control, capacity	Q301_01	Perceived Behavioral Control (capacity)	I am confident that I can share research data. (disagree 1- 7 agree)
control, autonomy	Q401_01	Perceived Behavioral Control (autonomy)	It is my own decision if I share research data. (disagree 1- 7 agree)
intention	Q101_01	Intention to share	When publishing empirical journal articles within the next 3 years, I will share my research data.(never 1- 7 always)
data sharing	Q501_01	Past data sharing behavior	In the past three years, when I published empirical journal articles, I have shared research data. (never 1-7 always)
code sharing	Q501_02	Past syntax sharing behavior	In the past three years, when I published empirical journal articles, I have shared syntax code. (never 1-7 always)
articles	Q503	Number of published articles	How many empirical journal articles did you publish during the past 3 years?
articles, self-collected	Q504	Number of published articles based on self-collected data	How many of your published journal articles were based on self-collected data?
articles, secondary	Q505	Number of published articles based on secondary data	How many of your published journal articles were based on secondary use of data?
articles, shared data	Q506	Number of published articles by which research data had been shared	With how many of your published journal articles did you share your research data?
articles, shared code	Q507	Number of published articles by which syntax code had been shared	With how many of your published journal articles did you share your syntax code?
data citing	Q508	Frequency of citing data of other researchers	How often did you cite secondary data of other researchers with a footnote or a reference in an article?
code citing	Q509	Frequency of citing syntax code of other researchers	How often did you cite syntax code of other researchers with a footnote or a reference in an article?