

Fecher, Benedikt; Fräßdorf, Mathis; Wagner, Gert G.

Working Paper

Perceptions and practices of replication by social and behavioral scientists: Making replications a mandatory element of curricula would be useful

DIW Discussion Papers, No. 1572

Provided in Cooperation with:

German Institute for Economic Research (DIW Berlin)

Suggested Citation: Fecher, Benedikt; Fräßdorf, Mathis; Wagner, Gert G. (2016) : Perceptions and practices of replication by social and behavioral scientists: Making replications a mandatory element of curricula would be useful, DIW Discussion Papers, No. 1572, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin

This Version is available at:

<https://hdl.handle.net/10419/130610>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

1572

Discussion Papers

Deutsches Institut für Wirtschaftsforschung

2016

Perceptions and Practices of Replication by Social and Behavioral Scientists

Making Replications a Mandatory Element of Curricula Would Be Useful

Benedikt Fecher, Mathis Fräßdorf and Gert G. Wagner

Opinions expressed in this paper are those of the author(s) and do not necessarily reflect views of the institute.

IMPRESSUM

© DIW Berlin, 2016

DIW Berlin
German Institute for Economic Research
Mohrenstr. 58
10117 Berlin

Tel. +49 (30) 897 89-0
Fax +49 (30) 897 89-200
<http://www.diw.de>

ISSN electronic edition 1619-4535

Papers can be downloaded free of charge from the DIW Berlin website:
<http://www.diw.de/discussionpapers>

Discussion Papers of DIW Berlin are indexed in RePEc and SSRN:
<http://ideas.repec.org/s/diw/diwwpp.html>
<http://www.ssrn.com/link/DIW-Berlin-German-Inst-Econ-Res.html>

Perceptions and Practices of Replication by Social and Behavioral Scientists

Making Replications a Mandatory Element of Curricula Would Be Useful

Benedikt Fecher^{1,2,3}, Mathis Fräßdorf¹ and Gert G. Wagner^{1,4,5,6}

1 German Institute for Economic Research (DIW Berlin), Berlin

2 Alexander von Humboldt Institut für Internet und Gesellschaft, Berlin

3 Leibniz-Association, Berlin

4 Max Planck Institute for Human Development (MPIB), Berlin

5 Technische Universität Berlin (TU Berlin), Berlin

6 Institute for the Study of Labor (IZA), Bonn

Berlin, April 2016

Abstract

We live in a time of increasing publication rates and specialization of scientific disciplines. More and more, the research community is facing the challenge of assuring the quality of research and maintaining trust in the scientific enterprise.

Replication studies are necessary to detect erroneous research. Thus, the replicability of research is considered a hallmark of good scientific practice and it has lately become a key concern for research communities and science policy makers alike. In this case study we analyze perceptions and practices regarding replication studies in the social and behavioral sciences. Our analyses are based on a survey of almost 300 researchers that use data from the German Socio-Economic Panel Study (SOEP), a multidisciplinary longitudinal multi-cohort study. We find that more than two thirds of respondents disagree with the statement that replications are not worthwhile, because major mistakes will be found at some point anyway. Nevertheless, most respondents are not willing to spend their time to conduct replication studies. This situation can be characterized as a “tragedy of the commons”: everybody knows that replications are useful, but almost everybody counts on others to conduct them. Our most important finding concerning practical consequences is that among the few replications that are reported, a large majority is conducted in the context of teaching. In our view, this is a promising detail: in order to foster replicability, one avenue may be to make replication studies a mandatory part of curricula as well as of doctoral theses. Furthermore, we argue that replication studies need to be more attractive for researchers. For example, successful replications could be listed in the publication lists of replicated authors. Vice versa, data sharing needs to receive more recognition, for example by considering data production and subsequent data sharing as scientific output.

Introduction

The Importance of Replication Studies

In recent years, the scientific community has repeatedly experienced prominent instances of fraudulent and erroneous research. An example of the latter in the fields of social sciences and economics is Reinhart and Rogoff's study "Growth in a Time of Debt" [1] on the effectiveness of austerity-based fiscal policies for highly indebted economies. The results of the study clearly translated into politics. It was influential on the United States Republican Party's budget proposal "The Path to Prosperity" as well as the EU Commissioner for Economic Affairs Olli Rehn's address to the International Labour Organisation in 2013. The questionable weighting methods and coding errors were only discovered after economists from the University of Massachusetts Amherst conducted a replication study [2].

The Reinhart-Rogoff-case is of relevance for this article for three reasons. Firstly, it shows that erroneous research can have an impact on political and economic decision-making. As Lacetera and Zirulia [3] state: "Even a handful of fraudulently produced results, if not detected promptly, can [...] endanger whole scientific fields as well as society at large". Secondly, the Reinhart-Rogoff-case shows the importance of data availability for the replicability of scientific research. Herndon and his colleagues [2] could only conduct a replication study and discover statistical errors because the authors of the study provided Herndon with the original dataset. And thirdly, the Reinhart-Rogoff-case shows the great potential of replication studies in teaching. Herndon, a PhD student, conducted the replication study [2] as part of a semester project.

Replication studies have an important internal value because they contribute to the self-correction abilities of the self-referential scientific ecosystem [4–6].

Additionally, replication studies enrich any curriculum in the empirical social sciences by showing how time-consuming and difficult research can actually be [7,8]. Moreover, replication studies have an external impact because they build and ensure civil society's trust in science.

The Replication Crisis

In a massive study published in *Science* that aimed at replicating the effects from 100 psychological studies, only 39% of the main effects in the original articles could be replicated [9]. The state of replicability in psychology even leads some to speak of a “replication crisis” [10].

It seems, however, that issues with the replicability of scientific research are not necessarily limited to a single discipline. In fact, replicability of research is an issue across disciplines [11–15]. A recent paper in *Nature* reported a failure to replicate significant experiments in the domain of cancer research in 47 out of 53 cases [16]. Empirical economics also faces problems with replication. In a study that aimed to replicate 18 studies published in two top journals (*American Economic Review* and the *Quarterly Journal of Economics*) between 2011 and 2014, the researchers were able to find a significant effect in the same direction as the original study for 11 of their replications (61%) [17]. In an attempt to replicate 67 papers published in 13 well-regarded economics journals, Chang and Li [18] were only able to replicate 22 (33%) of the results using data and material the authors had provided to the respective journals. Excluding six papers that used confidential data and two papers that used proprietary software, Chang and Li then explicitly asked the authors of the original articles for assistance and were able to successfully replicate 29 of 59 papers (49%).

It is important to bear in mind that all the above-mentioned studies targeted results published in top journals in their domain. They show quite plainly that there is a quality challenge in science today, namely how to ensure the integrity of scientific research while fields and methods become more and more specialized. Furthermore the studies show that the traditional peer review may be insufficient to ensure the integrity of empirical scientific research. Because replicability of research is a core of the scientific paradigm, the scientific community has to increase the replicability of published research as well as the number of replications of published results.

Barriers to replication

On the one hand, practical attempts to replicate results often fail. On the other hand, replication studies are either rarely conducted or are not feasible in the first place. Duvendack et al. [19] 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 [20] and initiatives towards openness and transparency, replication studies using the same data as the primary investigator (types a and c) should be growing in importance because of the lower costs for the replicator. Theoretically, a researcher does not have to repeat a whole study but can use the underlying data to verify or falsify published results (as a “minimal standard”).

There are a number of reasons why replication studies are currently not being conducted. Often results cannot be replicated because data from published research is not made available [21–28] or has not been sufficiently documented [26,27][26].

Even archived data is rarely actually used [30] (Peters et al. 2016). The most important reason why a researcher might be reluctant to conduct a replication study is *“because it is difficult to successfully accomplish and it carries more risk than potential reward for both the replicator and the originator of the research”* [31]. Hamermesh [32] consequently considers the replicability of (economic) research as an “ideal professed but not necessarily practiced”. The underdeveloped replication culture can thus be attributed to low data sharing rates, issues with data documentation and the limited attractiveness of conducting replication studies. Furthermore, in more subjective disciplines, for example the arts and humanities and to a certain degree social sciences and economics, the objective measure of replicability might be difficult to apply.

This article focuses on the researchers’ stance towards data sharing (as a prerequisite for replication studies), their perspective on the replicability of research, and their own replication practices. The results are based on a survey among 300 social and behavioral scientists, who use an easily obtained, well-documented and frequently analyzed data set. We show that researchers value data sharing and replication highly, but at best, they engage in both practices modestly. Based on our results, we furthermore conclude science policy measures to strengthen an academic replication culture.

Materials and Methods

Our empirical study is based on a standardized survey among researchers who analyze data from the German Socio-Economic Panel Study (SOEP). The SOEP is a widely analyzed multi-cohort study of the German population (Wagner et al 2007). In total, there are more than 7000 documented publications based on SOEP data in a

wide range of journals. The SOEP survey is administered under the umbrella of the Leibniz Society at the German Institute for Economic Research (DIW Berlin). The data of the SOEP is available at no direct costs to researchers via a research data center. This is particularly interesting in the context of data-driven replication studies, since the SOEP data is easily available; this can thus be considered a lower cost situation for the replicator compared to replicate results from the scratch.

Since 2004, user surveys have been conducted in order to identify the practices and needs of the secondary data users [33]. The survey that we analyze in the following sections was open for responses from November 12th, 2015 to January 4th, 2016. It contained 18 questions on data sharing and replication studies.

Of a total of 5,149 addresses that were registered in the “user data base” of SOEP, 4519 (88%) addresses were valid and were reached. Out of these 936 (21%) respondents used the link to the online questionnaire. During the course of the first part of the questionnaire, which was not related to data sharing or replication, a total number of 303 respondents took the opportunity to leave the questionnaire. Of the remaining 633 participants (14% of reached addresses), 321 answered the second part related to data sharing and replication. 300 respondents (32% of initial respondents, 7% of contacted addresses) had at least one valid answer in the second part.

Results

This findings section begins by providing an overview of respondent characteristics. It then provides a detailed look at researchers’ perceptions on data sharing and replication and several key replication practices. The final section of the findings looks at influences on conducting replication studies.

Demographics of respondents

The respondents to the additional module on data sharing and replications are mainly male (61%) and on average slightly over 38 years old (median 35). 37% are post docs or equivalent, 28% are professors, 26% are doctoral students, while a minority of 8% are students (most of them student who work in assistant roles). The two main research fields are sociology (47%) and economics (39%), while the remaining 14% come mostly from psychology, and demography, statistics or political science. The majority of respondents (78%) work in Germany, 18% is based within the EU; 4% of the responding SOEP data users come from North America, Australia, or Asia. These five demographic variables (sex, age, status, field, location) are control variables in the multivariate analyses.

Data Sharing

The questions on data sharing could be answered on a five-point scale from “does not apply at all” (value 1) to “fully applies” (value 5). We combine values “1” and “2” to create the category “does not apply”, and values “4” and “5” to create the category “does apply”.

Using this transformation, Figure 1 shows that 76% of respondents think that researchers should share their data for further analyses, and 89% even believe that data sharing furthers scientific progress. In addition, 73% disagree with the statement that they would rather not publish in journals with data policies that mandate data publication, and only 23% agree with the statement that they experience negative effects from sharing their data. Despite all these rather positive statements regarding data sharing, only 24% of respondents state that it is common in their discipline to share data. Figure 2 depicts how those who have already produced data –

these represent 64% (187) of our respondents – engage in data sharing: only 8% have shared their data publicly, while 26% have shared within the scientific community, 32% have shared with people they knew, and 34% have never shared their data.

Figure 1. Researcher's opinion about data sharing

Figure 2. Researcher's experience in sharing research data with others

Perceptions on Replication

Asked about replications in general (again using a 5 point scale adjusted as described in the previous section), 84% agree that replications are necessary for improving scientific output. 50% agree that the effort needed to produce a replication study is too high, and 43% agree with the statement, that the success of a replication study cannot be sufficiently measured. Lastly, 71% of respondents disagree with the statement that replications are not worthwhile, because major mistakes will be found at some point anyway (see Figure 3). When considering differences in this statement by field, we find—controlling for sex, status, and location of workplace—economists tend to be less likely to strongly disagree with this statement. This implies that economists are more likely than other researchers to believe that major mistakes will be found at some point anyway. When looking at age-field combinations—controlling for sex and location of workplace—economists between 30 and 45 are the one who believe most in the academic market place (see table 1).

Figure 3. Researcher's opinion about replication studies

Replications of research articles based on the SOEP do not happen often – Figure 4 shows the distribution. 58% of our respondents never attempted any replication study of an article based on SOEP data. Of those respondents who had conducted a replication study more than half of them are conducted during regular coursework – either while teaching a class (13% of all respondents) or while being taught as a student (9%). 20% of the respondents used a replication of a SOEP article for their own research. Of those who never conducted a replication study, 76% never saw a need to do so, while the rest thought it would be too time consuming (15%) or did not have enough information (9%)—either about the data, the software or the way results in the original article were produced, i.e., the scripts—were not available.

Figure 4. Researcher’s experience in conducting replication studies

As for those who did replicate a SOEP article, 84% were able to reproduce the results of the original article (although the results were not always exactly identical to those found by the original authors), while only 16% were not able to do so. When asked about the reason why the results could not be completely replicated, 69% of the respondents stated that the information in the original article was insufficient to allow for replication (of those, 85% needed more information on the data analyses in the original article, 15% did not have enough information about the data), while 27% thought that the effort would have been too much. Only 4% stated that the original article contained mistakes.

Regression Analyses

For exploratory reasons, we conducted a linear regression, considering the above-mentioned demographic variables, i.e. gender, age, status, field, and work location (see table 2). For this, we define three dependent variables, which are all based on the question of whether the respondent has already conducted replications of an article based on SOEP data. From the possible answers “no”, “yes, for my own research”, “yes, for my coursework (as a teacher)”, and “yes, for my coursework (as a student)” we were able to construct the following three dichotomous variables: (1) *any replication*, which is “1” if the respondent conducted a replication study for his or her research or as a part of the coursework, “0” otherwise; (2) *scientific replication*, which is “1” when replication was reported “for own research”, “0” otherwise; and (3) “educational replication”, which is “1” if replications were conducted for coursework, “0” otherwise. Since professional status highly correlates with age, we only include the status variable (i.e. doctoral student, post-doc or professor).

The estimated coefficients show the shift of the probability to replicate due to a right hand variable. In the regression analysis, none of the status and field variables are significantly related to any of the three types of replication. Female researchers are more likely to conduct educational replications, an effect which translates into the overall replication analysis. Additionally, in this sample the respondents from Germany turn out to be more likely to conduct replications for educational reasons, which also then is a significant influence regarding all replications.

Discussion

Although our sample is not representative of the German Social and Behavioral science community, the results provide a general idea on the perception of replication

studies as well as the barriers to conducting them. Data sharing as well as replication studies are practices that are generally perceived positively among researchers and in line with good scientific practice. As our results show, however, both are ideals professed but not practiced. For example: 71% of respondents disagree with the statement that replications are not worthwhile, because major mistakes will be found at some point anyway, but most respondents are not willing to spend their own time conducting replication studies. This can be regarded as a “tragedy of the commons”: everybody knows that replications are useful, but almost everybody counts on others to conduct them. A possible explanation for this is that conducting replication studies is not worthwhile in the context of the academic reward system since they are often time-consuming and rarely published [31].

We show that in the case of the well documented and openly available SOEP data, replication studies find few results to be erroneous. This could mean that researchers are more careful when using openly available data, because their results can easily be replicated. What is surprising, however, is the result that few replication studies based on the SOEP data are conducted despite the fact that the data is available for every researcher and well documented and thus easy to replicate. One reason might be that currently careful documentation and sharing of the code that is used for analyzing data is not common. The availability of syntax files could increase the replicability of research results and hence the number of data-driven replications. Thus for repositories it is worthy of consideration not only to implement the citation of data as well as the citation of code and syntax files.

Our results show that most of the replication studies are done in the context of teaching. In our view, this is a promising detail: in order to increase the number of replication studies, it may be feasible to make replications a mandatory part of cur-

ricula and an obligatory chapter of (cumulative) doctoral theses. In that way, students could ‘learn from the best’ while at the same time contributing to the overall integrity of scientific research.

On a general note, we propose that the research community should strive towards establishing a market and a culture of data sharing and re-use. Besides the aforementioned implementation in teaching, instruments seem suitable that take the academic reward system into account [4,34,35]. For instance, the attractiveness for replication studies would increase if more replication studies were published; especially in times of mega journals there is no limited space argument. An additional option could be increased funding explicitly for replication studies and meta analyses. Furthermore, positive replications could serve as a proof of research and therefore successful replications could be listed in the publication lists of replicated authors. Vice versa, data sharing needs to receive more recognition, for example by considering data production and subsequent data sharing as scientific output. In other words: the scientific community must treat the scientific paradigm more seriously and give credit in all cases where credit is due [37,38].

Acknowledgments

We thank Jan Goebel, Florian Griesse, Carolin Stolpe and Stefan Zimmermann for allowing us to pose our questions in the SOEP user survey, as well as providing us with the microdata. Marcel Hebing and Rainer Siegers helped in designing the questions. Last but not least we thank the respondents to the SOEP user survey for their answers.

References

1. Reinhart C, Rogoff K. Growth in a Time of Debt [Internet]. Cambridge, MA: National Bureau of Economic Research; 2010 Jan. Report No.: w15639. Available: <http://www.nber.org/papers/w15639.pdf>
2. Herndon T, Ash M, Pollin R. Does high public debt consistently stifle economic growth? A critique of Reinhart and Rogoff. *Camb J Econ*. 2013;38: 257–279. doi:10.1093/cje/bet075
3. Lacetera N, Zirulia L. The Economics of Scientific Misconduct. *J Law Econ Organ*. 2011;27: 568–603. doi:10.1093/jleo/ewp031
4. Fecher B, Wagner GG. A research symbiont. *Science*. 2016;351: 1405–1406. doi:10.1126/science.351.6280.1405-b
5. McNutt M. Reproducibility. *Science*. 2014;343: 229–229. doi:10.1126/science.1250475
6. Brüderl J. Sind die Sozialwissenschaften wissenschaftlich? Ergebnisse eines Replikationsexperiments [Internet]. Tagung “Rational Choice Soziologie” presented at; 2013 Nov 18; Venedig. Available: http://www.ls4.soziologie.uni-muenchen.de/studium_lehre/lehrveranstaltungen/wise1314/venice13/v_dienstag/bruederl.pdf
7. Hamermesh DS. Viewpoint: Replication in economics: Replication in economics. *Can J Econ Can Déconomique*. 2007;40: 715–733. doi:10.1111/j.1365-2966.2007.00428.x
8. Burman LE, Reed WR, Alm J. A Call for Replication Studies. *Public Finance Rev*. 2010;38: 787–793. doi:10.1177/1091142110385210
9. Open Science Collaboration. Estimating the reproducibility of psychological

science. *Science*. 2015;349: aac4716–aac4716. doi:10.1126/science.aac4716

10. Maxwell SE, Lau MY, Howard GS. Is psychology suffering from a replication crisis? What does “failure to replicate” really mean? *Am Psychol*. 2015;70: 487–498. doi:10.1037/a0039400

11. Scannell JW, Bosley J. When Quality Beats Quantity: Decision Theory, Drug Discovery, and the Reproducibility Crisis. Gasparini M, editor. *PLOS ONE*. 2016;11: e0147215. doi:10.1371/journal.pone.0147215

12. Bertoia ML, Bertrand KA, Sawyer SJ, Rimm EB, Mukamal KJ. Reproducibility of Circulating MicroRNAs in Stored Plasma Samples. Jeyaseelan K, editor. *PLOS ONE*. 2015;10: e0136665. doi:10.1371/journal.pone.0136665

13. Wan Y-W, Mach CM, Allen GI, Anderson ML, Liu Z. On the Reproducibility of TCGA Ovarian Cancer MicroRNA Profiles. Toland AE, editor. *PLoS ONE*. 2014;9: e87782. doi:10.1371/journal.pone.0087782

14. Garijo D, Kinnings S, Xie L, Xie L, Zhang Y, Bourne PE, et al. Quantifying Reproducibility in Computational Biology: The Case of the Tuberculosis Drugome. Ouzounis CA, editor. *PLoS ONE*. 2013;8: e80278. doi:10.1371/journal.pone.0080278

15. Diekmann A. Empirische Sozialforschung Grundlagen, Methoden, Anwendungen. Reinbek bei Hamburg: Rowohlt-Taschenbuch-Verl.; 2010.

16. Begley CG, Ellis LM. Drug development: Raise standards for preclinical cancer research. *Nature*. 2012;483: 531–533. doi:10.1038/483531a

17. Camerer CF, Dreber A, Forsell E, Ho T-H, Huber J, Johannesson M, et al. Evaluating replicability of laboratory experiments in economics. *Science*. 2016; doi:10.1126/science.aaf0918

18. Chang AC, Board of Governors of the Federal Reserve System, Li P. Is Eco-

nomics Research Replicable? Sixty Published Papers from Thirteen Journals Say “Usually Not.” *Finance Econ Discuss Ser.* 2015;2015: 1–26.

doi:10.17016/FEDS.2015.083

19. Duvendack M, Palmer-Jones RW, Weed RW. Replications in Economics: A Progress Report. *Econ J Watch.* 2015;12: 164–191.

20. Hey AJG, Tansley S, Tolle KM. The fourth paradigm: data-intensive scientific discovery. Redmond, Wash.: Microsoft Research; 2009.

21. Alsheikh-Ali AA, Qureshi W, Al-Mallah MH, Ioannidis JPA. Public Availability of Published Research Data in High-Impact Journals. Boutron I, editor. *PLoS ONE.* 2011;6: e24357. doi:10.1371/journal.pone.0024357

22. Tenopir C, Allard S, Douglass K, Aydinoglu AU, Wu L, Read E, et al. Data Sharing by Scientists: Practices and Perceptions. Neylon C, editor. *PLoS ONE.* 2011;6: e21101. doi:10.1371/journal.pone.0021101

23. Andreoli-Versbach P, Mueller-Langer F. Open access to data: An ideal professed but not practised. *Res Policy.* 2014;43: 1621–1633. doi:10.1016/j.respol.2014.04.008

24. Fecher B, Friesike S, Hebing M. What Drives Academic Data Sharing? Phillips RS, editor. *PLOS ONE.* 2015;10: e0118053. doi:10.1371/journal.pone.0118053

25. Piwowar HA. Who Shares? Who Doesn't? Factors Associated with Openly Archiving Raw Research Data. Neylon C, editor. *PLoS ONE.* 2011;6: e18657. doi:10.1371/journal.pone.0018657

26. Tudur Smith C, Dwan K, Altman DG, Clarke M, Riley R, Williamson PR. Sharing Individual Participant Data from Clinical Trials: An Opinion Survey Regarding the Establishment of a Central Repository. Hills RK, editor. *PLoS ONE.* 2014;9: e97886. doi:10.1371/journal.pone.0097886

27. Milia N, Congiu A, Anagnostou P, Montinaro F, Capocasa M, Sanna E, et al. Mine, Yours, Ours? Sharing Data on Human Genetic Variation. Stanyon R, editor. PLoS ONE. 2012;7: e37552. doi:10.1371/journal.pone.0037552
28. Tenopir C, Dalton ED, Allard S, Frame M, Pjesivac I, Birch B, et al. Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide. van den Besselaar P, editor. PLOS ONE. 2015;10: e0134826. doi:10.1371/journal.pone.0134826
29. Ioannidis JPA. Why Most Published Research Findings Are False. PLoS Med. 2005;2: e124. doi:10.1371/journal.pmed.0020124
30. Peters I, Kraker P, Lex E, Gumpenberger C, Gorraiz J. Research data explored: an extended analysis of citations and altmetrics. Scientometrics. 2016; doi:10.1007/s11192-016-1887-4
31. Park CL. What is the value of replicating other studies? Res Eval. 2004;13: 189–195. doi:10.3152/147154404781776400
32. Hamermesh DS. Viewpoint: Replication in economics: Replication in economics. Can J Econ Can Déconomique. 2007;40: 715–733. doi:10.1111/j.1365-2966.2007.00428.x
33. Hebing M, Griesse F, Napieraj J, Pahl M, Stolpe C, Wagner GG. Zur Struktur von empirischen Sozial-, Verhaltens- und Wirtschaftsforschern: ein Überblick über die Ereignisse der SOEP-Nutzerbefragungen. SOEPpapers. 2014;708. Available: http://www.diw.de/sixcms/detail.php?id=diw_01.c.492119.de
34. McNutt M. #IAmAResearchParasite. Science. 2016;351: 1005–1005. doi:10.1126/science.aaf4701
35. McNutt M. Reproducibility. Science. 2014;343: 229–229. doi:10.1126/science.1250475

Appendix

Figure 1

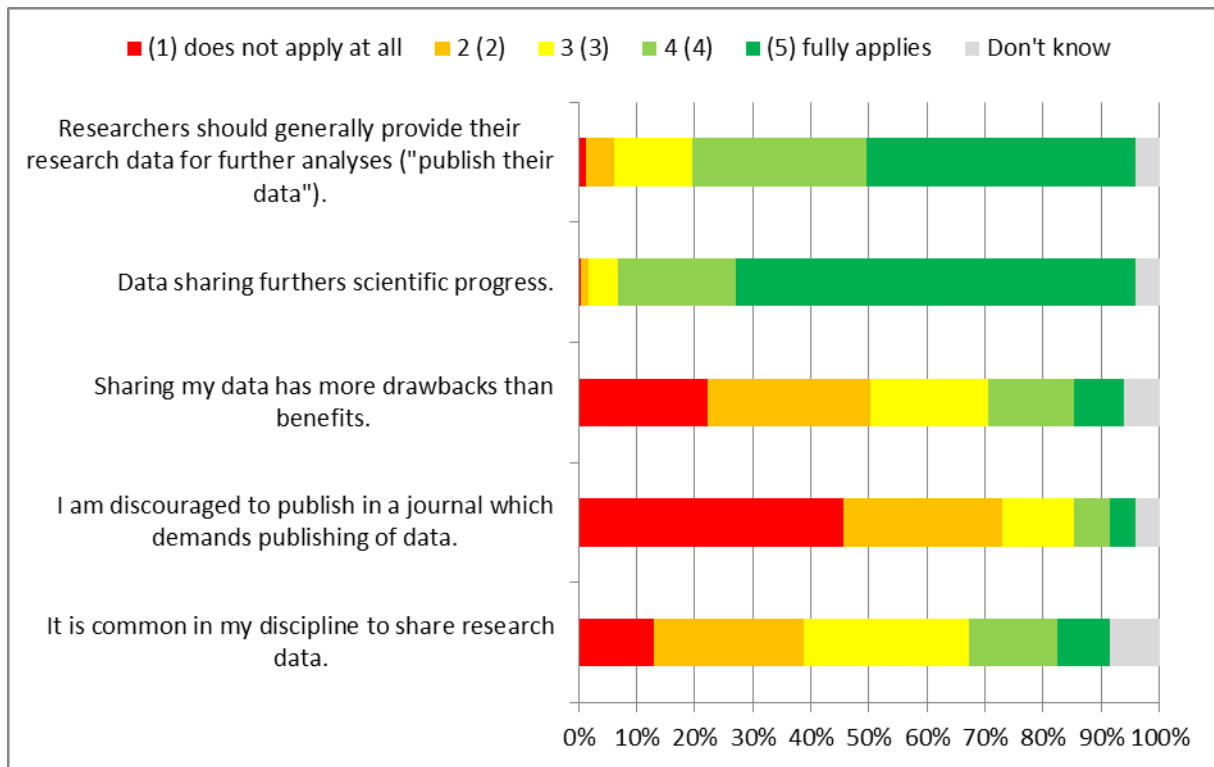


Figure 2

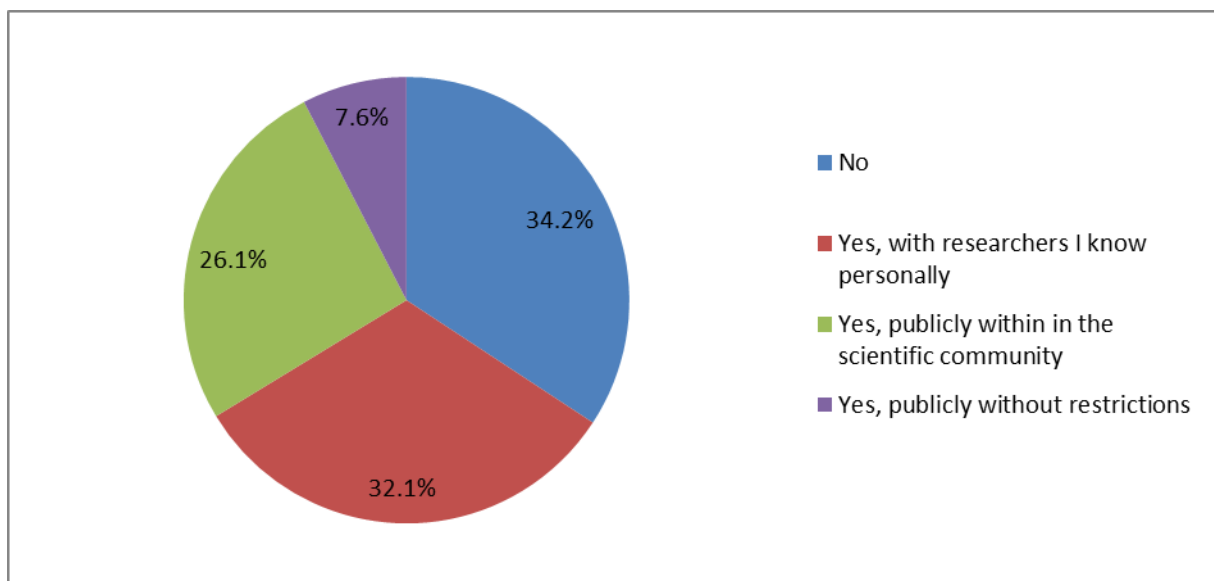


Figure 3

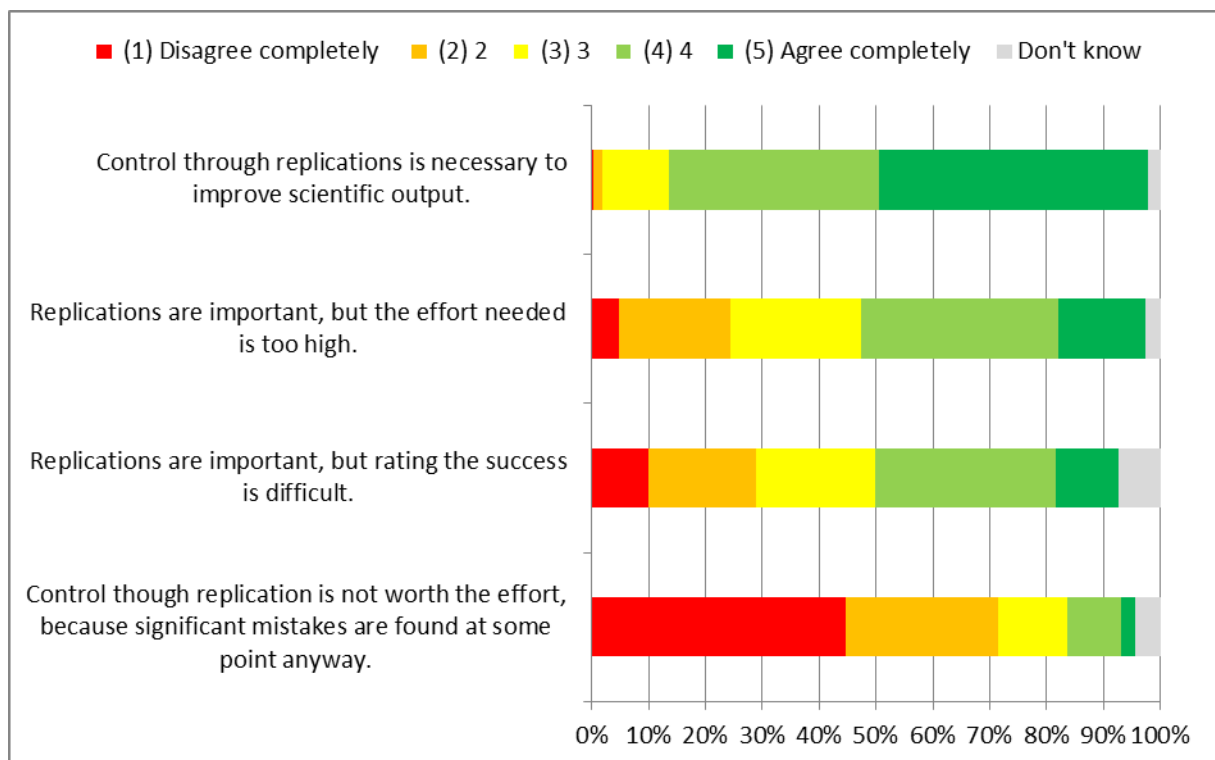


Figure 4

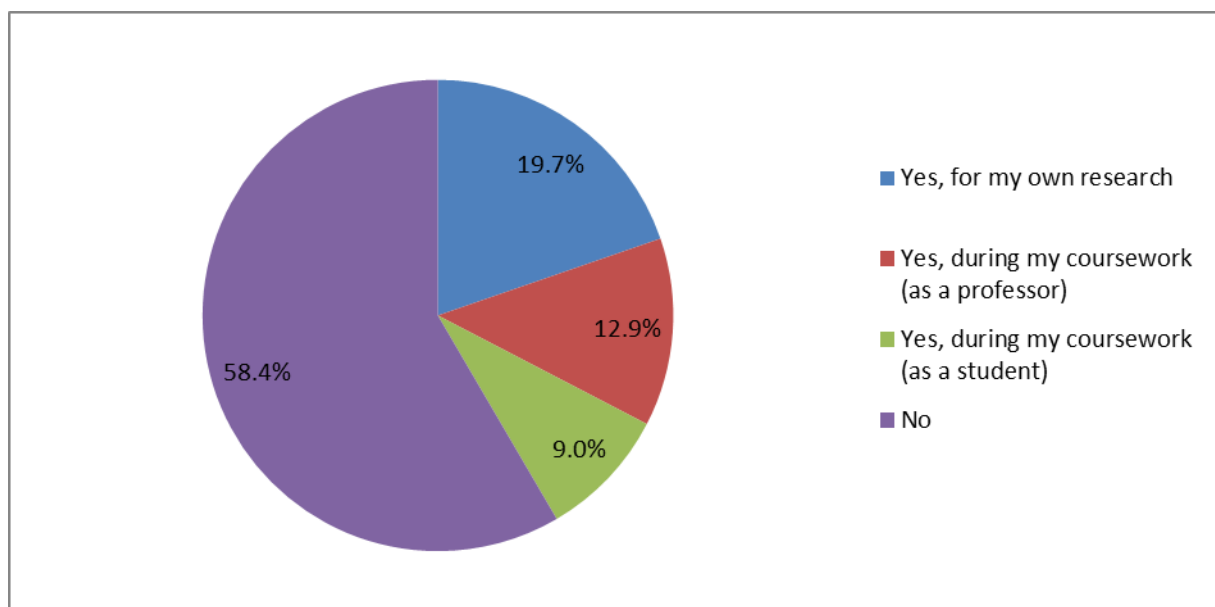


Table 1. Regression results, linear probability model (0,1 coding of the left hand variable).

Statement: "Using replication to check findings is not worth the effort since important mistakes will eventually come to light anyway."		
	Disagrees strongly with statment	
Sex		
(Male)		
Female	-0.006	
	(0.072)	
Status		
(Undergraduate)		
Doctoral Student	0.067	
	(0.125)	
Post-Doc	-0.055	
	(0.131)	
Professor	0.121	
	(0.129)	
Field		
(Economics)		
Sociology	0.125	
	(0.073)	
Other	0.184	
	(0.101)	
Location of Work		
Germany	0.265	
	(0.079)	
(Other countries)		
Constant	0.143	
	(0.139)	
Observations	227	
R-squared	0.100	
Notes:		
Standard errors in parentheses. Coefficients significant on a 10% level are in bold. Reference categories are marked in parentheses in the left hand column. Analyses using probit and logit models did not yield different results.		
Source: SOEP User Survey 2015		

Table 2. Regression results, linear probability model (0,1 coding of the left hand variables “any replication done”, “replication for scientific purposes”, “replication for educational purposes”).

	Type of Replication		
	Any	Scientific	Educational
Sex			
(Male)			
Female	0.155	0.040	0.115
	(0.067)	(0.054)	(0.057)
Status			
(Undergraduate)			
Doctoral Student	-0.151	-0.018	-0.133
	(0.129)	(0.104)	(0.109)
Post-Doc	-0.074	-0.033	-0.040
	(0.134)	(0.108)	(0.114)
Professor	-0.088	-0.104	0.015
	(0.132)	(0.107)	(0.113)
Field			
(Economics)			
Sociology	-0.069	-0.068	-0.001
	(0.071)	(0.058)	(0.061)
Other	-0.075	0.020	-0.095
	(0.098)	(0.080)	(0.084)
Location of Work			
(Other countries)			
Germany	0.197	0.035	0.161
	(0.077)	(0.062)	(0.066)
Constant	0.319	0.211	0.108
	(0.142)	(0.115)	(0.121)
Observations	241	241	241
R-squared	0.0678	0.021	0.0783
Notes:			
Standard errors in parentheses. Coefficients significant on a 10% level are in bold. Reference categories are marked in parentheses in the left hand column. Analyses using probit and logit models did not yield different results.			
Source: SOEP User Survey 2015			



Thank you for interest in this year's SOEP User Survey!

Participation in our user survey is entirely voluntary and the results will be stored in anonymous form. No names or addresses will be saved. The data will be evaluated by DIW Berlin and employees of DIW Berlin only. The results of the survey will be published at the beginning of 2016 on the website above and in the SOEPnewsletter.

Note for users who have completed past SOEP user surveys:

We encourage respondents to past SOEP user surveys to take part again this year! The content is constantly being updated and the technology is being improved. This year, we have also included new topics (data sharing and open access publications), which are covered in the second part of the survey.

Section A: Data usage

A1. Do you use the SOEP data for empirical or methodological analysis?

Empirical analysis ☐

Methodological analysis ☐

Both ☐

I don't work with the data myself ☐

A2. Please state the three most important topic areas in SOEP for your research in order of importance.

Please choose max. three of the following

Demography, population ☐

Labor market and employment ☐

Income, public transfers, assets, and social security ☐

Family, relationships, and social networks ☐

Health and health care ☐

Housing, home features, and home production activities of private households ☐

Education and qualifications ☐



A7. Which of the SOEP studies do you already know and which of them do you use?

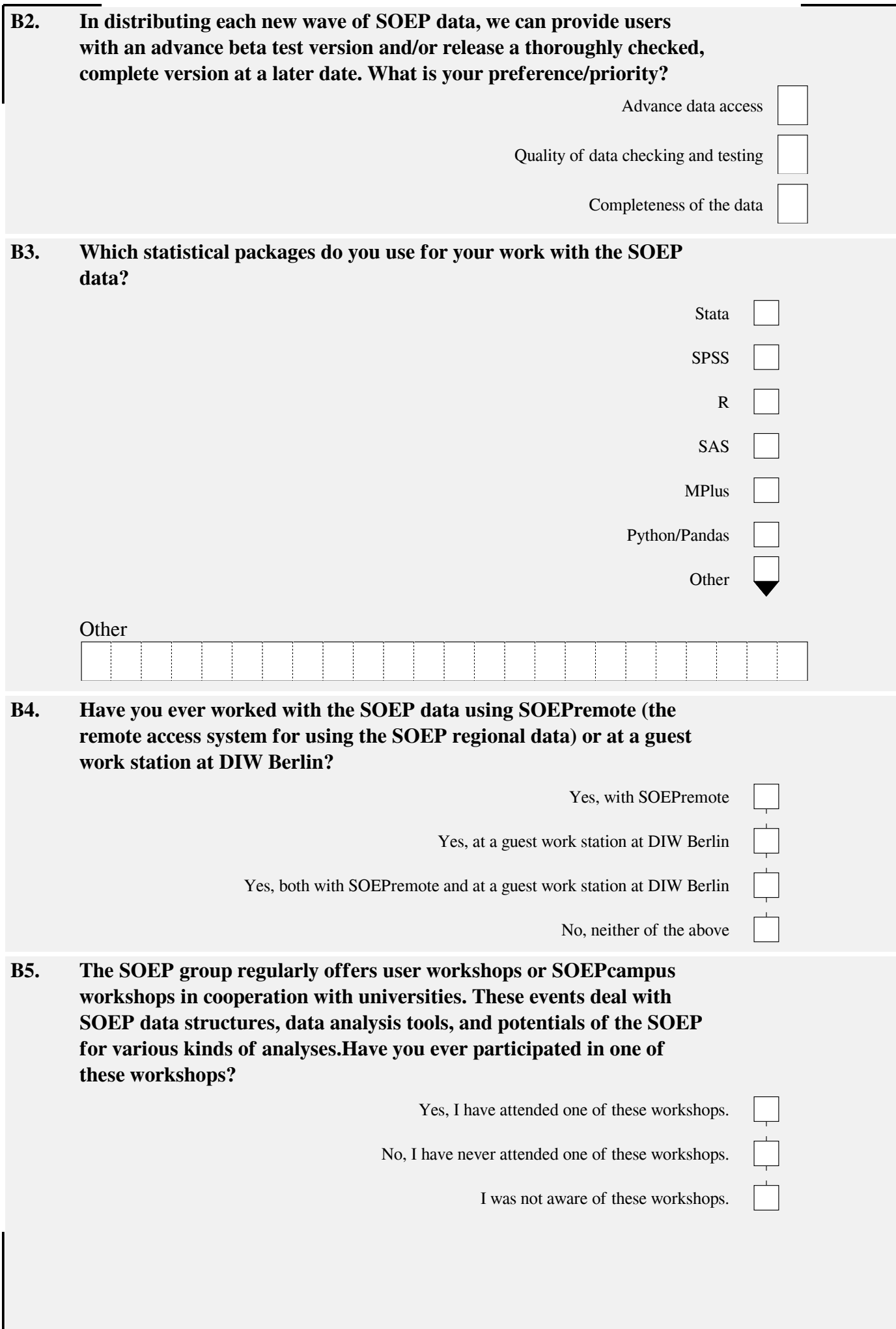
A8. Why don't you personally work with the SOEP data?

[illegible]

Section B: How do you use the SOEP data?

B1. What year did you start working with the SOEP data?

[illegible]





Section C: Data documentation

The SOEP is currently revising its data documentation system and is currently providing data on two metadata platforms:

1. The classic SOEPinfo (panel.gsoep.de/soepinfo), providing SOEP since 1997
2. The new Paneldata.org (paneldata.org), which will replace the previous SOEPinfo. Here, along with documentation on the core SOEP study (SOEP-Core), you will also find the SOEP Innovation Sample (SOEP-IS) and other studies.

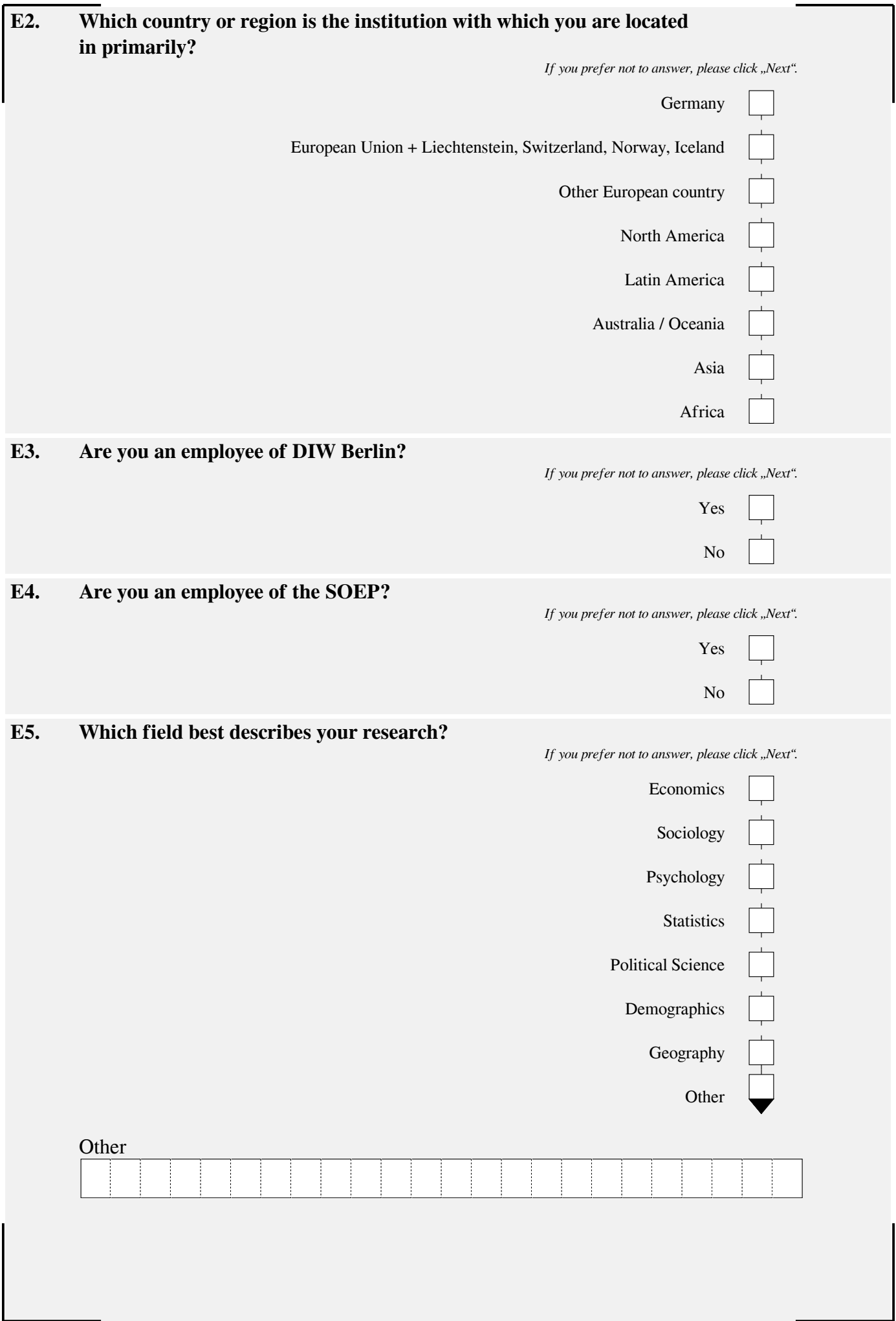
In the next block of questions, we ask for your feedback on how to make the transition from the old to the new system as smooth and easy as possible for you.

C1. We want to replace the previous SOEPinfo with our new metadata portal, paneldata.org. Which metadata portals are you already familiar with and what documentation do you use?

	I use it on a regular basis.	I have worked with it before.	I am familiar with it but have never worked with it.	I am not familiar with it.
Paneldata.org	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
SOEPinfo	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Possibility to download individual files on the SOEP website	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Zip file of the SOEP data v.30	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
The documentation on the data DVD	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

C2. Are there any reasons why you have not used the new SOEPinfo v.2 yet?

[illegible]





E6. Do you teach at the university level?

If you prefer not to answer, please click „Next“.

Yes ☐

No ☐

E7. Do you advise young researchers who are working with the SOEP data?

If you prefer not to answer, please click „Next“.

Yes ☐

No ☐

E8. Please state your sex.

If you prefer not to answer, please click „Next“.

Female ☐

Male ☐

E9. How old are you?

If you prefer not to answer, please click „Next“.

--	--	--	--	--	--	--	--	--	--

Section F: Data Sharing in Academia

The questions in this section deal with your work with research data in general. If you work with other datasets besides the SOEP, please consider your experience with research data in general and not just with the SOEP.

F1. Thank you for your participation up to this point. This brings the SOEP-related user survey to an end, but we would still like to hear your opinions on a few questions about secondary data use, open access, and re-analysis in the second part of this survey. Would you like to continue?

Yes ☐

No ☐



F2. Please rate your agreement with the following statements:

	Disagree completely 1	2	3	4	Agree completely 5	Don't know
Researchers should make their research data available ("publish" the data, so to speak) for further analysis as a general rule.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
It's normal in my research area / research community to share research data.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Sharing my data with others has more disadvantages than advantages for me.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I know where and how to find relevant secondary data for my research.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
When a journal requires that the data be published it deters me from submitting there.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Open access to research data contributes significantly to progress in scientific research.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I know where and how to make data I have collected available to others.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Researchers should make their data available to others at as early a stage as possible.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Researchers should make their research data available after publication of their first article.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I can imagine using data from other researchers for my own work.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

F3. Have you ever collected data yourself?

Yes ☐

No ☐

F4. Have you ever shared your own research data with the research community or the general public?

No ☐

Yes, but only with researchers I know personally ☐

Yes, publicly, but only for scientific purposes ☐

Yes, publicly, without any restrictions ☐

F5. Do you work with other secondary data (that is, data made available by others for use by the scientific community) besides the SOEP data?

Yes ☐

No ☐

[illegible]

F7. When using secondary data, it would be important to me...

[illegible]



F9. How relevant are the following data sources for your work with secondary data?

F10. Aside from the SOEP, what other micro data sets do you use?

Sonstige

F11. Which of the following statements apply to you?

The SOEP data are the main data I use in my work. ☐

I combine the SOEP data with other data sets. ☐

I use the SOEP data for international comparative analysis. ☐



F12. How important are open access publications to you—that is, scientific articles provided for free over the Internet?

	Not important at all 1	2	3	4	Extremely important 5	Don't know
Open access is ...	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

F13. Please rate your agreement with the following statements about increasing use of open access publication?

	Disagree completely 1	2	3	4	Agree completely 5	Don't know
Open access is not important to me since discussion papers and manuscripts are available through other channels.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I would only put publications on the Internet if I didn't have to pay for it.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Government and research funding organizations should buy licenses from publishers to provide open access.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
In order to provide open access across the board, public research funding organizations should have their own journals.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
The journals run by research funding organizations will never be as high in quality as journals run by academic publishers.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Section G: Re-analysis

G1. Replication of research findings can serve as preparation for one's own research or as an example in a learning context. Have you ever attempted to replicate an article based on SOEP data?

Yes, for my own research	<input type="checkbox"/>
Yes, in a learning context (as the instructor)	<input type="checkbox"/>
Yes, in a learning context (as a student)	<input type="checkbox"/>
No	<input type="checkbox"/>

G2. Why is it that you have never attempted to replicate a SOEP article?



G6. How would you rate the value of replications in scientific research in general?

	Disagree completely 1	2	3	4	Agree completely 5	Don't know
Replication to check findings is fundamental for improving scientific output	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Replications are important, but they are too time-consuming	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Using replication to check findings is not worth the effort since important mistakes will eventually come to light anyway	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Replications are important, but it is difficult to evaluate the success of a replication	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

G7. Are there other aspects that are important to you on the subject of replication? Please state them here!

Section H: Closure

We are grateful for your participation in this survey and have just one more request.

SOEP Director Jürgen Schupp will be celebrating his 60th birthday on 12th January 2016. He has been key part of the study since 1984. We would like to invite his friends and acquaintances, research colleagues, and SOEP data users to send him their greetings, anecdotes, and best wishes. Important: the following link will open a new page. Your message will not be linked to the user survey in any way. [If you don't have time right now, you can use the link later or send us an e-mail: usersurvey@diw.de]

<https://fragebogen.diw.de/limesurvey2/index.php/198575?lang=en>

H1. In conclusion, we welcome your feedback or suggestions if there is anything else you would like to tell us!

Your suggestions can also refer to this user survey.



Thank you for your participation!

**If you have any questions, please feel free to contact us at any time:
usersurvey@diw.de**

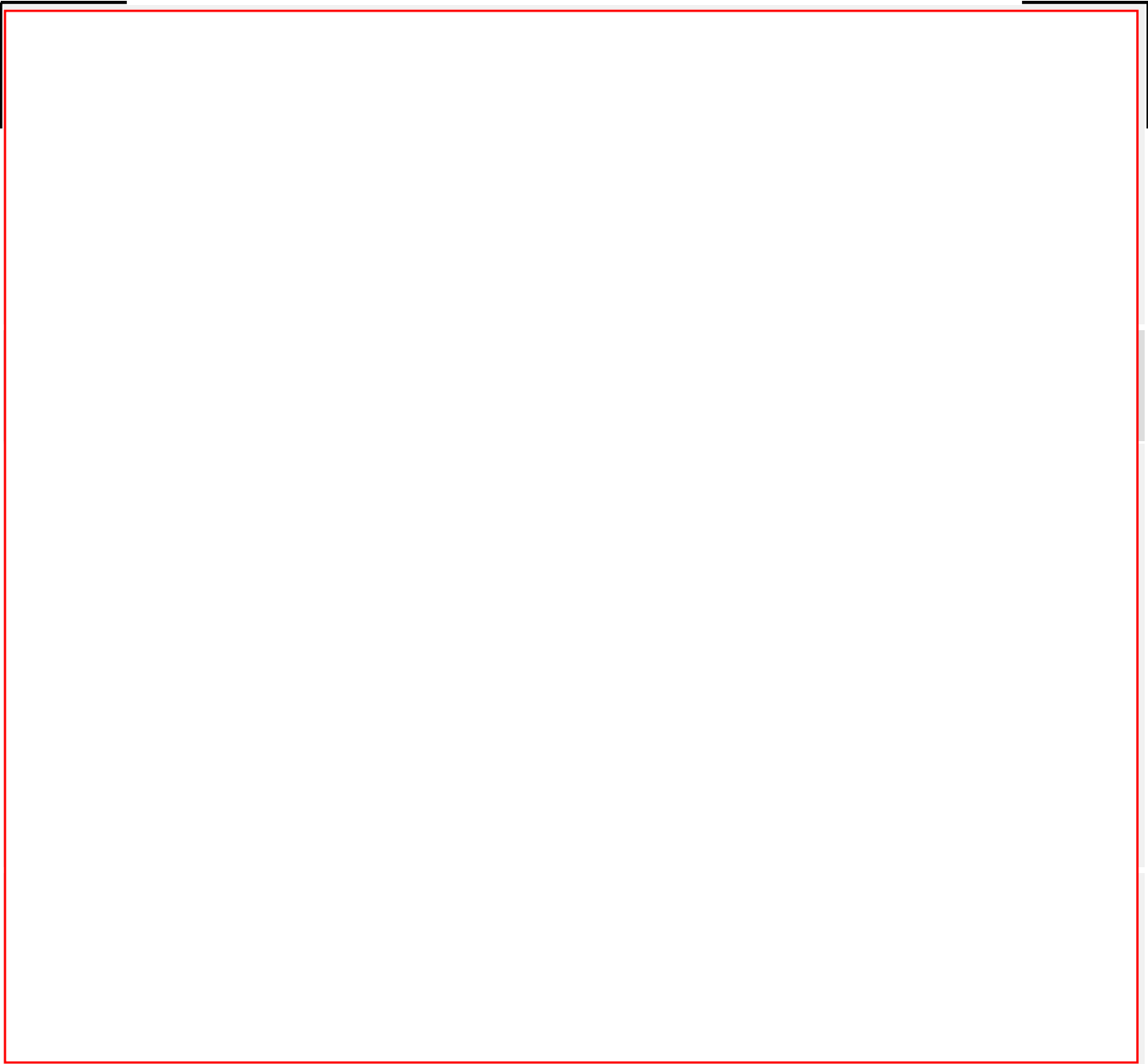


Thank you for interest in this year's SOEP User Survey!

Participation in our user survey is entirely voluntary and the results will be stored in anonymous form. No names or addresses will be saved. The data will be evaluated by DIW Berlin and employees of DIW Berlin only. The results of the survey will be published at the beginning of 2016 on the website above and in the SOEPnewsletter.

Note for users who have completed past SOEP user surveys:

We encourage respondents to past SOEP user surveys to take part again this year! The content is constantly being updated and the technology is being improved. This year, we have also included new topics (data sharing and open access publications), which are covered in the second part of the survey.



Section E: Particulars

[illegible]

If you prefer not to answer, please click „Next“.

Student

Research associate/Post-doc

Doctoral student	
------------------	--

Professor

Other

Other

[illegible]



E8. Please state your sex.

If you prefer not to answer, please click „Next“.

Female ☐

Male ☐

E9. How old are you?

If you prefer not to answer, please click „Next“.

--	--	--	--	--	--	--	--	--	--

Section F: Data Sharing in Academia

The questions in this section deal with your work with research data in general. If you work with other datasets besides the SOEP, please consider your experience with research data in general and not just with the SOEP.

F1. Thank you for your participation up to this point. This brings the SOEP-related user survey to an end, but we would still like to hear your opinions on a few questions about secondary data use, open access, and re-analysis in the second part of this survey. Would you like to continue?

Yes ☐

No ☐



F2. Please rate your agreement with the following statements:

	Disagree completely 1	2	3	4	Agree completely 5	Don't know
Researchers should make their research data available ("publish" the data, so to speak) for further analysis as a general rule.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
It's normal in my research area / research community to share research data.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Sharing my data with others has more disadvantages than advantages for me.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I know where and how to find relevant secondary data for my research.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
When a journal requires that the data be published it deters me from submitting there.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Open access to research data contributes significantly to progress in scientific research.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I know where and how to make data I have collected available to others.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Researchers should make their data available to others at as early a stage as possible.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Researchers should make their research data available after publication of their first article.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I can imagine using data from other researchers for my own work.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

F3. Have you ever collected data yourself?

Yes ☐

No ☐

F4. Have you ever shared your own research data with the research community or the general public?

No ☐

Yes, but only with researchers I know personally ☐

Yes, publicly, but only for scientific purposes ☐

Yes, publicly, without any restrictions ☐

F5. Do you work with other secondary data (that is, data made available by others for use by the scientific community) besides the SOEP data?

Yes ☐

No ☐



	Disagree completely	2	3	4	Agree completely	Don't know
	[]1				[]5	
that the data are being provided by an organization.	[]	[]	[]	[]	[]	[]
that I know and trust the researcher providing the data.	[]	[]	[]	[]	[]	[]
that the data are being provided for free or at a low service cost (e.g., to cover shipping).	[]	[]	[]	[]	[]	[]
that relevant articles have already been published with the data.	[]	[]	[]	[]	[]	[]
that I will receive access to the data quickly and easily.	[]	[]	[]	[]	[]	[]
that the data collection process is documented understandably.	[]	[]	[]	[]	[]	[]
that the data are easy for me to use.	[]	[]	[]	[]	[]	[]
that the data are already available in the format of preferred statistical software.	[]	[]	[]	[]	[]	[]
that there is a hotline or someone who I can contact with questions.	[]	[]	[]	[]	[]	[]
that data processing and analysis scripts are provided along with the data.	[]	[]	[]	[]	[]	[]

F7. When using secondary data, it would be important to me...

[illegible]



F9. How relevant are the following data sources for your work with secondary data?

F10. Aside from the SOEP, what other micro data sets do you use?

Sonstige

F11. Which of the following statements apply to you?

The SOEP data are the main data I use in my work. ☐

I combine the SOEP data with other data sets. ☐

I use the SOEP data for international comparative analysis. ☐



F12. How important are open access publications to you—that is, scientific articles provided for free over the Internet?

	Not important at all 1	2	3	4	Extremely important 5	Don't know
Open access is ...	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

F13. Please rate your agreement with the following statements about increasing use of open access publication?

	Disagree completely 1	2	3	4	Agree completely 5	Don't know
Open access is not important to me since discussion papers and manuscripts are available through other channels.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I would only put publications on the Internet if I didn't have to pay for it.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Government and research funding organizations should buy licenses from publishers to provide open access.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
In order to provide open access across the board, public research funding organizations should have their own journals.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
The journals run by research funding organizations will never be as high in quality as journals run by academic publishers.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Section G: Re-analysis

G1. Replication of research findings can serve as preparation for one's own research or as an example in a learning context. Have you ever attempted to replicate an article based on SOEP data?

Yes, for my own research	<input type="checkbox"/>
Yes, in a learning context (as the instructor)	<input type="checkbox"/>
Yes, in a learning context (as a student)	<input type="checkbox"/>
No	<input type="checkbox"/>

G2. Why is it that you have never attempted to replicate a SOEP article?



G6. How would you rate the value of replications in scientific research in general?

	Disagree completely 1	2	3	4	Agree completely 5	Don't know
Replication to check findings is fundamental for improving scientific output	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Replications are important, but they are too time-consuming	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Using replication to check findings is not worth the effort since important mistakes will eventually come to light anyway	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Replications are important, but it is difficult to evaluate the success of a replication	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

G7. Are there other aspects that are important to you on the subject of replication? Please state them here!



Thank you for your participation!

**If you have any questions, please feel free to contact us at any time:
usersurvey@diw.de**