

Forschungsdatenaustausch in den Wirtschaftswissenschaften

Benedikt Fecher, Deutsches Institut für Wirtschaftsforschung (DIW Berlin)
Augsburg, 4. September 2016

»Open science is also about making sure that science serves innovation and growth. It guarantees open access to publicly-funded research results and the possibility of knowledge sharing by providing infrastructures. Facilitating access to those data will encourage re-use of research output.«

Günther Oettiner

Agenda

1. Warum Forschungsdatenaustausch?
2. Zwischen Ideal und Wirklichkeit
 - a. Forscher
 - b. Forschungsförderer
 - c. Fachzeitschriften
3. Ein Markt für Forschungsdaten

Warum Forschungsdatenaustausch?

- Replizierbarkeit
- Synergien
- Neue Methoden

Warum Forschungsdatenaustausch?

- Replizierbarkeit
- Synergien
- Neue Methoden

Warum Forschungsdatenaustausch?



Sign In | Link your subscription | TNY Store

THE NEW YORKER



NEWS CULTURE BOOKS & FICTION SCIENCE & TECH BUSINESS HUMOR CARTOONS MAGAZINE VIDEO ARCHIVE SUBSCRIBE Q

APRIL 26, 2013

THE REINHART AND ROGOFF CONTROVERSY: A SUMMING UP

BY JOHN CASSIDY



JOHN CASSIDY



In one of life's little ironies, last Friday's disappointing G.D.P. figures, which reflected a sharp fall in government spending, appeared on the same day that the economists Carmen Reinhart and Kenneth Rogoff published an *Op-Ed* in the *Times* defending their famous (now infamous) research that conservative politicians around the world had seized upon to justify penny-pinching policies. Addressing a new paper by three lesser lights of their profession from the University of Massachusetts, Amherst, which uncovered data omissions, questionable methods of weighting, and elementary coding errors in Reinhart and Rogoff's original work, and which went around the world like a viral video, the Harvard duo dismissed the entire brouhaha as "academic kerfuffle" that hadn't vitiated their main points.



MOST POPULAR

1. Why You Hate Google's New Logo
BY SARAH LARSON
2. Six Weeks' Paid Leave Opposed By People With Thirty-Three Weeks' Paid Leave
BY ANDY BOROWITZ

Fehlerhafte Analysen und Rechenfehler

COMMENT

CHEMISTRY What John Dalton did for science **p.32**

REPRODUCTION Abstract art and experimental science compared **p.33**

REPRODUCIBILITY A call to hallmark labs that produce replicable work **p.34**

PEER REVIEW We've let the perfect become the enemy of the good **p.34**



Stop ignoring misconduct

Efforts to reduce irreproducibility in research must also tackle the temptation to cheat, argue Donald S. Kornfeld and Sandra L. Titus.

The history of science shows that irreproducibility is not a product of our times. Some 350 years ago, the chemist Robert Boyle penned essays on "the unsuccessfulness of experiments". He warned readers to be sceptical of reported

work. "You will meet with several Observations and Experiments, which ... may upon further trial disappoint your expectation." He attributed the problem to a "lack of skill in the scientist and the lack of purity of the ingredients", and what would today be

referred to as inadequate statistical power. By 1830, polymath Charles Babbage was writing in more cynical terms. In *Reflections on the Decline of Science in England*, he complains of "several species of impositions that have been practised in science", namely "hoaxing, forging, trimming and cooking". In other words, irreproducibility is the product of two factors: faulty research practices and fraud. Yet, in our view, current initiatives to improve science dismiss the second factor. For example, leaders at the US National Institutes of Health (NIH) stated in 2014: "With rare exceptions, we have no evidence to suggest that irreproducibility is caused by scientific misconduct". In 2015, a symposium of several UK science-funding agencies convened to address reproducibility, and decided to exclude discussion of deliberate fraud. To dismiss the role of research misconduct is mistaken and unfortunate. At best, ignoring deliberate misconduct in efforts to reduce irreproducibility is a wasted opportunity, like tilling a field without clearing it of rocks. At worst, it permits destructive behaviour to persist and flourish.

SCALE OF EVIDENCE
Only 10–12 individuals are found guilty by the US Office of Research Integrity (ORI) each year. That number, which the NIH used to dismiss the role of research misconduct, is misleadingly low, as numerous studies show. For instance, a review of 2,047 life-science papers retracted from 1973 to 2012 found that around 43% were attributed to fraud or suspected fraud. A compilation of anonymous surveys suggests that 2% of scientists and trainees admit that they have fabricated, falsified or modified data. And a 1996 study of more than 1,000 post-docs found that more than one-quarter would select or omit data to improve their chances of receiving grant funding. Admittedly, many causes of irreproducibility do not involve dishonesty. The NIH has promoted responsible research for 25 years by funding studies on research integrity, creating educational resources and backing the ORI. Nonetheless, we contend that when scientific leaders minimize "hoaxing, forging, trimming and cooking" as contributors to irreproducibility, they choose to ignore the problem rather than confront it. This ▶

Warum Forschungsdatenaustausch?

DIW BERLIN

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 Frähdorf and Gerrit G. Wagner

»Most respondents are not willing to spend their time to conduct replication studies.«

Fecher et al., 2016

Warum Forschungsdatenaustausch?

Viewpoint: Replication in economics

Daniel S. Hamermesh *Department of Economics,
University of Texas at Austin*

Abstract. This examination of the role and potential for replication in economics points out the paucity of both pure replication – checking on others’ published papers using their data – and scientific replication – using data representing different populations in one’s own work or in a comment. Several controversies in empirical economics are used to illustrate how and how not to behave when replicating others’ work. The incentives for replication are examined, and proposals aimed at journal editors and authors are advanced that might stimulate an activity that most economists applaud but few perform. JEL classification: B41, A14, C59

Réplication en science économique. L’examen du rôle potentiel de la réplication en science économique souligne la pauvreté à la fois de la réplication pure (la vérification des résultats des textes publiés en utilisant les données utilisées dans ces travaux) et de la réplication scientifique (l’utilisation de données tirées de populations différentes pour valider des résultats). Plusieurs controverses en économie empirique sont utilisées pour illustrer comment se comporter et ne pas se comporter dans la réplication des travaux d’autrui. Les incitations à faire de la réplication sont analysées, et certaines suggestions sont faites aux rédacteurs de revues et aux auteurs afin de stimuler une activité plus grande dans ce registre – que les économistes applaudissent mais que bien peu acceptent de pratiquer.

1. Introduction: What is replication?

Economists treat replication the way teenagers treat chastity – as an ideal to be professed but not to be practised. Why is this? How much replication is done by

Edward Everett Hale Centennial Professor of Economics, University of Texas at Austin, and research associate, NBER and IZA. I thank Bernd Fitzenberger, Bruce McCullough, and Gerald Oettinger for helpful suggestions on an earlier draft, the authors of several of the studies cited here for useful clarifications of their views on the controversies in which they were involved, and several editors for their experiences at their journals. Email: hamermes@eco.utexas.edu

»Economists treat replication the way teenagers treat chastity – as an ideal to be professed but not to be practiced«

Hamermesh, 2007

Warum Forschungsdatenaustausch?

- Replizierbarkeit
- Synergien
- Neue Methoden

Warum Forschungsdatenaustausch?

OPEN ACCESS Freely available online



Data Sharing by Scientists: Practices and Perceptions

Carol Tenopir^{1*}, Suzie Allard¹, Kimberly Douglass¹, Arsev Umur Aydinoglu¹, Lei Wu¹, Eleanor Read², Maribeth Manoff¹, Mike Frame¹

1 School of Information Sciences, University of Tennessee, Knoxville, Tennessee, United States of America, **2** University of Tennessee Libraries, University of Tennessee, Knoxville, Tennessee, United States of America, **3** Center for Biological Informatics, United States Geological Survey, Oak Ridge, Tennessee, United States of America

Abstract

Background: Scientific research in the 21st century is more data intensive and collaborative than in the past. It is important to study the data practices of researchers – data accessibility, discovery, re-use, preservation and, particularly, data sharing. Data sharing is a valuable part of the scientific method allowing for verification of results and extending research from prior results.

Methodology/Principal Findings: A total of 1329 scientists participated in this survey exploring current data sharing practices and perceptions of the barriers and enablers of data sharing. Scientists do not make their data electronically available to others for various reasons, including insufficient time and lack of funding. Most respondents are satisfied with their current processes for the initial and short-term parts of the data or research lifecycle (collecting their research data; searching for, describing or cataloging, analyzing, and short-term storage of their data) but are not satisfied with long-term data preservation. Many organizations do not provide support to their researchers for data management both in the short- and long-term. If certain conditions are met (such as formal citation and sharing reprints) respondents agree they are willing to share their data. There are also significant differences and approaches in data management practices based on primary funding agency, subject discipline, age, work focus, and world region.

Conclusions/Significance: Barriers to effective data sharing and preservation are deeply rooted in the practices and culture of the research process as well as the researchers themselves. New mandates for data management plans from NSF and other federal agencies and world-wide attention to the need to share and preserve data could lead to changes. Large scale programs, such as the NSF-sponsored DataNET (including projects like DataONE) will both bring attention and resources to the issue and make it easier for scientists to apply sound data management principles.

Citation: Tenopir C, Allard S, Douglass K, Aydinoglu AU, Wu L, et al. (2011) Data Sharing by Scientists: Practices and Perceptions. PLoS ONE 6(6): e21101. doi:10.1371/journal.pone.0021101

Editor: Cameron Neylon, Science and Technology Facilities Council, United Kingdom

Received: January 3, 2011 **Accepted:** May 20, 2011 **Published:** June 29, 2011

Copyright: © 2011 Tenopir et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The project was funded as part of the National Science Foundation Division of Cyberinfrastructure, Data Conservation Network for Earth (DataONE) NSF award #0330944 under a Cooperative Agreement. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: tenopir@u.tkn.edu

Introduction

Data are the infrastructure of science. Sound data are critical as they form the basis for good scientific decisions, wise management and use of resources, and informed decision-making. Moreover, “science is becoming data intensive and collaborative” [1]. The amount of data collected, analyzed, re-analyzed, and stored has increased enormously due to developments in computational simulation and modeling, automated data acquisition, and communication technologies [2]. Following the previous research paradigms (experimental, theoretical, and computational), the new era has been called “the fourth paradigm: data-intensive scientific discovery” where “all of the science literature is online, all of the science data is online, and they interoperate with each other” [3]. Digital data are not only the outputs of research but provide inputs to new hypotheses, enabling new scientific insights and driving innovation [4].

As science becomes more data intensive and collaborative, data sharing becomes more important. Data sharing includes the deposition and preservation of data; however, it is primarily

associated with providing access for use and reuse of data. Data sharing has many advantages, including:

- re-analysis of data helps verify results/data, which is a key part of the scientific process;
- different interpretations or approaches to existing data contribute to scientific progress—especially in an interdisciplinary setting;
- well-managed, long-term preservation helps retain data integrity;
- when data is available, (re-)collection of data is minimized; thus, use of resources is optimized;
- data availability provides safeguards against misconduct related to data fabrication and falsification;
- application studies serve as training tools for new generations of researchers [5][6][7].

Additionally, researchers, data managers and publishers in the PARSE survey overwhelmingly agreed that public funding was the

»60% to 90% of respondents in all disciplines agree with the statement that they would use other researchers' datasets if their datasets were easily accessible.«

Tenopir et al., 2011

The great potential of data sharing

RatSWD
Working
Paper
Series

http://www.ratswd.de

RatSWD.
German Data Forum

215

Open Access to Data: An Ideal
Professed but not Practised

Patrick Andreoli-Versbach
Frank Mueller-Langer

February 2013

»We find that 8.81% of researchers share some of their data whereas only 2.05% fully share.«

Andreoli-Versbach & Mueller-Langer, 2013

Warum Forschungsdatenaustausch?

- Replizierbarkeit
- Synergien
- Neue Methoden

Warum Forschungsdatenaustausch?

Challenges of Open Data in Medical Research

Ralf Floca

Abstract The success of modern, evidence based and personalized medical research is highly dependent on the availability of a sufficient data basis in terms of quantity and quality. This often also implies topics like exchange and consolidation of data. In the area of conflict between data privacy, institutional structures and research interests, several technical, organizational and legal challenges emerge. Coping with these challenges is one of the main tasks of information management in medical research. Using the example of cancer research, this case study points out the marginal conditions, requirements and peculiarities of handling research data in the context of medical research.

Introduction

First the general importance of data exchange and consolidation will be discussed. In the second section, the important role of the patient in medical research will be addressed and how it affects the handling of data. The third section focuses on the question what the role of open data could be in this context. Finally, the fourth section tackles the topic of challenges of open data in the context of medical (research) data. It tries to illustrate why it is a problem and what the obstacles are.

R. Floca (✉)
German Cancer Research Center, Heidelberg, Germany
e-mail: r.floca@dkfz.de

S. Bartling and S. Friesike (eds.), *Opening Science*,
DOI: 10.1007/978-3-319-00026-8_22, © The Author(s) 2014

297

»The current situation can [...] be characterized as too few cases for too many factors. The size of sample sets even large institutions can collect is too small for evidence-based and highly stratified medicine.«

Floca, 2014

Warum Forschungsdatenaustausch?

- Replizierbarkeit  Qualität
- Synergien  Quantität/Effizienz
- (Neue) Methoden  Innovation

Zwischen Ideal und Wirklichkeit

»The 'dirty little secret' behind the promotion of data sharing is that not much sharing may be taking place.«

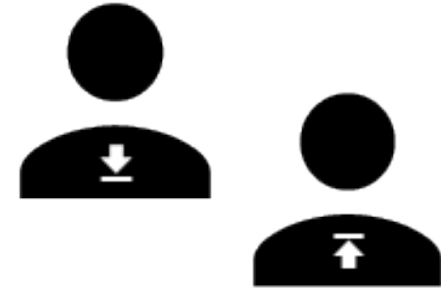
Borgman 2012: 1059

Zwischen Ideal und Wirklichkeit

- Wissenschaftler
- Forschungsförderer
- Fachzeitschriften

Zwischen Ideal und Wirklichkeit

- Wissenschaftler
- Forschungsförderer
- Fachzeitschriften

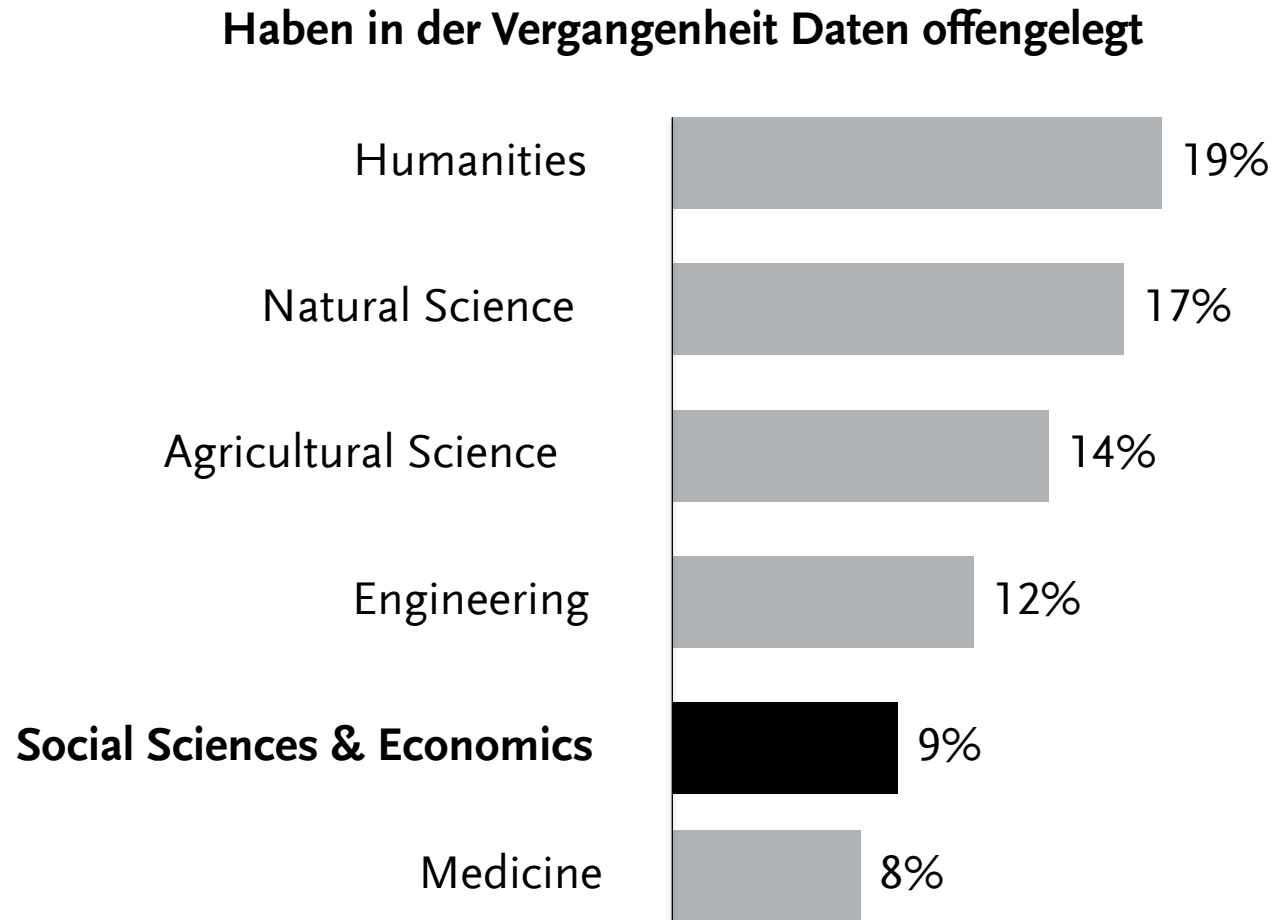


Zwischen Ideal und Wirklichkeit

- 83% sagen, offener Zugang zu Forschungsdaten ist ein wesentlicher Beitrag zum wissenschaftlichen Fortschritt; 13% haben in der Vergangenheit Daten offengelegt
- „Falls andere vor mir publizieren können“ größtes Hindernis (4.25)
- „Kritik/Falsifikation“ ist kein Hindernis (1.99)
- Datenzitation größter Anreiz (4.56)
- 61% der Naturwissenschaftler sagen, sie wissen, wo und wie man Daten veröffentlichen kann; 42% der Ökonomen

Source: Own survey among 1564 German researchers

Zwischen Ideal und Wirklichkeit



Source: Own survey among 1564 German researchers

Zwischen Ideal und Wirklichkeit

- Wissenschaftler
- Forschungsförderer
- Fachzeitschriften



Zwischen Ideal und Wirklichkeit

Forschungsförderer

- von 36 erwähnen 18 (51%) Datendokumentation in ihren Förderrichtlinien
- 12 (34%) haben verpflichtende Veröffentlichung, 11 (31%) empfehlen, 12 (34%) keine Policy
- 14 (40%) haben „Embargo-Period“

Source: Own analysis of 36 research funders' guidelines

Zwischen Ideal und Wirklichkeit

- Wissenschaftler
- Forschungsförderer
- Fachzeitschriften



Zwischen Ideal und Wirklichkeit


Fachzeitschriften

- 346 Fachzeitschriften (economics and business studies)
- 49 (14%) haben eine Daten-Policy
- 37 (11%) verlangen Daten und Supporting Material

- Von den 126,500 Studien, die in den Top-50 Wiwi.-zeitschriften zw. 1974-2014 veröffentlicht wurden, waren 131 Replikationsstudien
- Mehr als die Hälfte der haben noch nie eine Replikationsstudie veröffentlicht

Source:
Vlaeminck & Herrmann, 2015
Mueller-Langer et al, 2015

Zwischen Ideal und Wirklichkeit

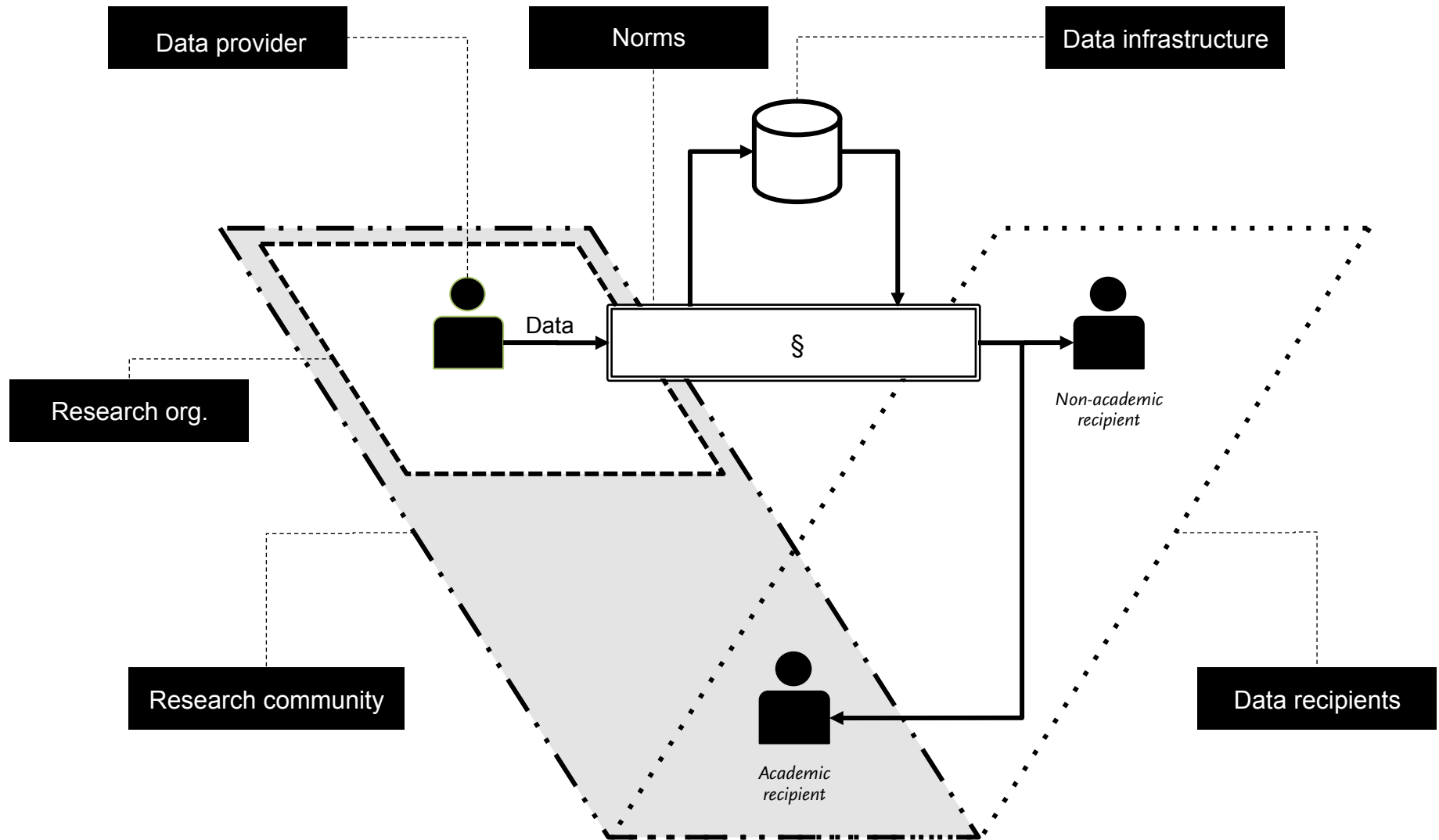
- Diskrepanz zwischen dem Ideal und der Wirklichkeit beim offenen Zugang zu Forschungsdaten
- „Reputationsökonomie“: Für Forschende sind Daten eine Investition in die Artikelpublikation; Datenpublikationen geben kaum Anerkennung
- Forschungsförderer ändern allmählich ihre Policies
- Fachzeitschriften 

Ein Markt für Forschungsdaten

Forschungspolitische Maßnahmen sollte daher darauf abzielen,

- den **Reputationsnutzen** für die Produktion, Bereitstellung und Nachnutzung hochwertiger Datensätze zu **erhöhen**,
- den **Aufwand der Bereitstellung** und (damit die Transaktionskosten) möglichst **gering zu halten**,
- **Daten auffindbar und nachnutzbar zu machen** und
- rechtliche **Unsicherheiten zu beseitigen**.

Ein Markt für Forschungsdaten



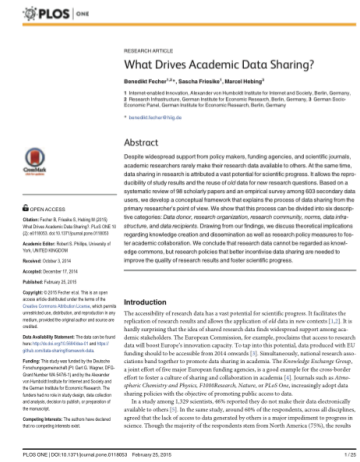
Source: Own systematic review



Fecher B, Fräßdorf M, Wagner GG. *Perceptions and Practices of Replication by Social and Behavioral Scientists: Making Replication a Mandatory Element of Curricula Would Be Useful.* DIW Discussion Papers. 2016;



Fecher B, Friesike S, Hebing M, Link S, Sauermann A (2015) *A Reputation Economy: Results from an Empirical Survey on Academic Data Sharing.* DIW Berlin Discussion



Fecher B, Friesike S, Hebing M (2015) *What Drives Academic Data Sharing?* PLoS ONE 10(2): e0118053.

References

- Cassidy, John (2013) The Reinhart and Rogoff Controversy: A Summing Up. <http://www.newyorker.com/news/john-cassidy/the-reinhart-and-rogooff-controversy-a-summing-up>.
- Fecher, Benedikt, Sascha Friesike, and Marcel Hebing (2015) What Drives Academic Data Sharing? Robert S. Phillips, ed. PLOS ONE 10(2): e0118053.
- Fecher, Benedikt, Sascha Friesike, Marcel Hebing, Stephanie Linek, and Armin Sauermann 2015 A Reputation Economy: Results from an Empirical Survey on Academic Data Sharing. DIW Berlin Discussion Paper 1454. <http://dx.doi.org/10.2139/ssrn.2568693>.
- Hamermesh, Daniel (2012) Six Decades of Top Economics Publishing: Who and How? w18635. Cambridge, MA: National Bureau of Economic Research. <http://www.nber.org/papers/w18635.pdf>, accessed April 12, 2015.
- Ian, Sample (2015) Study Delivers Bleak Verdict on Validity of Psychology Experiment Results. <http://www.theguardian.com/science/2015/aug/27/study-delivers-bleak-verdict-on-validity-of-psychology-experiment-results>.
- Tenopir, Carol, Suzie Allard, Kimberly Douglass, et al. (2011) Data Sharing by Scientists: Practices and Perceptions. Cameron Neylon, ed. PLoS ONE 6(6): e21101.
- The Economist (2013) Problems with Scientific Research: How Science Goes Wrong. <http://www.economist.com/news/leaders/21588069-scientific-research-has-changed-world-now-it-needs-change-itself-how-science-goes-wrong>.