

RatSWD Working Paper Series

www.ratswd.de

RatSWD ■

Rat für Sozial- und
Wirtschaftsdaten

194

Datenverfügbarkeit reicht nicht, um
Replikationsstudien zur Routine zu
machen

Gert G. Wagner und Denis Huschka

Januar 2012

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

Working Paper Series des Rates für Sozial- und Wirtschaftsdaten (RatSWD)

Die *RatSWD Working Papers* Reihe startete Ende 2007. Seit 2009 werden in dieser Publikationsreihe nur noch konzeptionelle und historische Arbeiten, die sich mit der Gestaltung der statistischen Infrastruktur und der Forschungsinfrastruktur in den Sozial-, Verhaltens- und Wirtschaftswissenschaften beschäftigen, publiziert. Dies sind insbesondere Papiere zur Gestaltung der Amtlichen Statistik, der Ressortforschung und der akademisch getragenen Forschungsinfrastruktur sowie Beiträge, die Arbeit des RatSWD selbst betreffend. Auch Papiere, die sich auf die oben genannten Bereiche außerhalb Deutschlands und auf supranationale Aspekte beziehen, sind besonders willkommen.

RatSWD Working Papers sind nicht-exklusiv, d. h. einer Veröffentlichung an anderen Orten steht nichts im Wege. Alle Arbeiten können und sollen auch in fachlich, institutionell und örtlich spezialisierten Reihen erscheinen. Die *RatSWD Working Papers* können nicht über den Buchhandel, sondern nur online über den RatSWD bezogen werden.

Um nicht deutsch sprechenden Nutzer/innen die Arbeit mit der neuen Reihe zu erleichtern, sind auf den englischen Internetseiten der *RatSWD Working Papers* nur die englischsprachigen Papers zu finden, auf den deutschen Seiten werden alle Nummern der Reihe chronologisch geordnet aufgelistet.

Einige ursprünglich in der *RatSWD Working Papers* Reihe erschienenen empirischen Forschungsarbeiten sind ab 2009 in der RatSWD Research Notes Reihe zu finden.

Die Inhalte der *RatSWD Working Papers* stellen ausdrücklich die Meinung der jeweiligen Autor/innen dar und nicht die des RatSWD.

Herausgeber der RatSWD Working Paper Series:

Vorsitzender des RatSWD (2007/2008 Heike Solga; seit 2009 Gert G. Wagner)

Geschäftsführer des RatSWD (Denis Huschka)

Datenverfügbarkeit reicht nicht, um Replikationsstudien zur Routine zu machen

Gert G. Wagner und Denis Huschka

Rat für Sozial- und Wirtschaftsdaten

Welcher Verhaltenskodex muss der Wissenschaft zugrunde liegen? Auf jeden Fall gehört Replizierbarkeit dazu. Sie ist das zentrale Merkmal von Wissenschaftlichkeit. Nur replizierbare Ergebnisse sind wirkliche wissenschaftliche Ergebnisse. Studien, die nicht replizierbar sind, sind streng genommen keine wissenschaftliche Forschung, sondern – wenn sie gut gemacht sind – eine Art Feuilleton. Dennoch – und dies mag auf den ersten Blick verwundern – ist für die meisten Wissenschaftler das Ermöglichen und vor allem das Durchführen von Replikationen alles andere als eine Selbstverständlichkeit.

In den Sozial-, Verhaltens- und Wirtschaftswissenschaften wird kaum repliziert und geprüft. Und dies obwohl inzwischen viele Datensätze frei zugänglich sind und von Hunderten von Wissenschaftlerinnen und Wissenschaftlern weltweit ausgewertet werden können. Eine Ausnahme sind Zeitreihenanalysen auf Basis der volkswirtschaftlichen Gesamtrechnung: hier liegen derart wenige Beobachtungspunkte vor, dass Replikationen bzw. die Verbesserung wissenschaftlicher Ergebnisse durch die Anwendung neuer Methoden notgedrungen zum wissenschaftlichen Alltag gehören.

Vor diesem Hintergrund ist eine aktuelle Erhebung der multidisziplinären Fachzeitschrift *Science* besonders interessant. Im Dezember 2011 widmete sich bereits zum zweiten Mal in diesem Jahr ein *Science*-Schwerpunkt dem Thema Forschungsdaten. Unter der Überschrift „Data Replication and Reproducibility“ werden Möglichkeiten der Replikation empirischer Studien in verschiedenen Bereichen erörtert (<http://www.sciencemag.org/site/special/data-rep/>).

1 Empirische Ergebnisse

Dass die Wiederholbarkeit empirischer Studien ein wichtiges wissenschaftliches Gut ist, welches aber nicht selbstverständlich „produziert“ wird, sondern welches durch verschiedene Mechanismen unterstützt und anerkannt werden muss, wurde durch eine innerhalb des Schwerpunkts „Data Replication and Reproducibility“ von *Science* geschalteten Online-Umfrage deutlich (vgl. auch den Anhang). Dabei wurde die Frage gestellt:

„Idealerweise sollten Wissenschaftler ihre eigenen (Roh-)Daten und Methoden vollständig veröffentlichen und auch Zeit darauf verwenden, andere Arbeiten zu replizieren. Was müsste getan werden um diese gute Praxis zu gewährleisten?“

Folgende Ergebnisse erbrachte die Umfrage:

- Anerkennung und Honorierung durch Institutionen: 6 %
- Zweckgebundene Fördermittelbereitstellung für Replikationsstudien: 19 %
- Mehr Veröffentlichung von Ergebnissen in Journalen, die vorangegangene Arbeiten bestätigen oder widerlegen: 45 %
- Die Hinterlegung von ausreichend Informationen zum Zweck der Replikation sollte durch die Förderinstitutionen bei der Bearbeitung nachfolgender Förderanträge honoriert bzw. bei Nichtbefolgung sanktioniert werden: 24 %

Diese Ergebnisse sind wissenschaftstheoretisch und -soziologisch interessant. Der allergrößte Teil der Umfrageteilnehmer – es dürfte sich in erster Linie um Wissenschaftler handeln – verlangt handfeste (materielle)

Anreize, um eigentlich selbstverständliches wissenschaftliches „Normalverhalten“ honoriert zu bekommen, nämlich „Reproducibility“. Dies ist bedenklich – Wissenschaftlichkeit muss sich offenkundig vor allem lohnen: materiell (Fördermittel) oder durch Reputationszugewinne (Publikationen). Der hehre Anspruch an wissenschaftliche Objektivität und Wiederholbarkeit *an sich* genügt nicht (mehr).

Freilich: das mangelnde Verständnis für den originären Wert der Wissenschaftlichkeit einfach auf die Wissenschaftler zu schieben wäre viel zu einfach. Offenbar gibt es im modernen Wissenschaftssystem weltweit eine der Wertsphäre „Lehrbuch-Wissenschaftlichkeit“ entgegenstehende „Wirklichkeits-Sphäre“, die von systemimmanenten Opportunitäten und Belohnungsstrukturen gekennzeichnet ist. Wissenschaftler, besonders junge, die sich Ihre Reputation erst erarbeiten und diese dann festigen müssen, unterliegen inzwischen weltweit den gleichen Anreizstrukturen: Publish or Perish – und zwar möglichst viele Publikationen in möglichst angesehenen Fachzeitschriften. Und in diese Topjournals kommt man i. d. R. nicht mit Replikationsstudien. So wichtig diese für die Wissenschaftlichkeit und dem Erhalt des gesamten Wissenschaftssystems sind, so sind mit Replikationen von Arbeiten anderer doch kaum Meriten zu gewinnen. Es muss schon ein handfester Verdacht auf Fälschung oder methodische Fehler vorliegen bevor ein publiziertes Ergebnis wirklich einem Replikationsversuch unterzogen wird. Eher schon werden verschiedene Studien für „Metaanalysen“ nebeneinandergestellt und kritisch gewürdigt. Metaanalysen werden als eigenständige Forschungsleistung angesehen; während Replikationsstudien als langweilig gelten. Metastudien enttarnen Fälschungen und schlecht gemachte Forschung aber keineswegs zwangsläufig, sondern merkwürdige Ergebnisse werden einfach als „Ausreißer“ hingenommen.

Dieses strukturelle Set-up hat Auswirkungen auf das Verhalten empirisch arbeitender Forscher: Wenn es also in vielen Fällen überhaupt keine Nachfrage an meinen Forschungsdaten und Auswertungsprozeduren zum Zwecke der Replikation gibt, dann wird natürlich auch möglichst wenig Zeit darauf verschwende, diese aufzubereiten und zugänglich zu machen. Und wenn man mit bloßer Replikation keinen Blumentopf gewinnen kann, dann werden auch kaum Replikationsversuche unternommen. Dies ist – systemimmanent – voll und ganz rational.

Gerade die Sozial- und Wirtschaftswissenschaften machen das Problem sehr deutlich: in diesen Disziplinen gibt es zum Teil seit Jahrzehnten Datenarchive und interessante Datensätze, die Replikationsstudien auf vergleichsweise einfache Art und Weise erlauben würden. Freilich wird

kaum – im Wortsinne – nachgerechnet. Die trotz sehr guter Datenverfügbarkeit auch in den Sozial- und Wirtschaftswissenschaften unzureichende Replikation von Ergebnissen unterstreicht: die freie und einfache Verfügbarkeit von Daten reicht nicht aus um Replikationen zu fördern und damit einen wichtigen Beitrag zur vernachlässigten Qualitätssicherung wissenschaftlicher Forschung zu leisten. Es müssen auch die individuellen Anreize stimmen.

2 Wissenschaftstheoretische und wissenschaftsorganisatorische Schlussfolgerungen

Die *Science*-Umfrage zu „Data Replication and Reproducibility“ zeigt deutlich: Wissenschaftler aller Disziplinen erwarten „Rewards“ für das eigentlich Selbstverständliche, nämlich Wissenschaftlichkeit ihrer Arbeitsweise und ihrer Ergebnisse. Mehr (Forschungs-)Geld oder mehr Publikationen. An die simple „Anerkennung“ denken nur 6 %.

Die Ergebnisse zeigen nicht, dass Wissenschaftler keine Berufsethik hätten, sondern die Antworten spiegeln ganz deutlich die Opportunitätsstrukturen im Wissenschaftssystem wieder. Und sie benennen die Währung, in der gezahlt werden muss, will man mehr Wissenschaftlichkeit mit Hilfe von systematischen Replikationsstudien erreichen: Publikationen. Denn im Rekrutierungsverfahren für Senior-Stellen im Wissenschaftssystem nimmt die möglichst lange Publikationsliste eine zentrale Position ein, nicht der bloße Nachweis einer wissenschaftlichen Arbeitsweise. Diese unterstellt man – bis zum Beweis des Gegenteils. Wobei dieser Beweis zunehmend geführt wird: bekannt gewordene Fälschungsskandale haben in den letzten Jahren zugenommen. Insbesondere – so hat man den Eindruck – in den Lebenswissenschaften.

Die Publikationslisten und daraus berechenbaren Zitations-Maßzahlen einzelner Forscher sind Objektivierungsversuche, wissenschaftliche Leistung einfach zählbar zu machen. Der prominente Zeitschriftentitel ist dabei für die Leser ein Indikator für die Relevanz und die Stimmigkeit der Ergebnisse. Doch dies ist gefährlich. Denn so gut wie kein Gutachter rechnet die Ergebnisse in den eingereichten Manuskripten nach. Fast alle Gutachter beurteilen die Plausibilität eines Ergebnisses und vertrauen auf die Redlichkeit der Autoren. Und in prominenten Journals publizierte

Ergebnisse werden dann oft jahre- oder jahrzehntelang unkritisch zitiert, da kritische Replikationen kaum Gewinn versprechen (da zu selten gravierende Mängel entdeckt werden dürften). Und wenn tatsächlich einmal repliziert wird, so zeigt sich gerade in den Sozial-, Verhaltens- und Wirtschaftswissenschaften, dass die publizierten Ergebnisse in den meisten Fällen – wie Tests immer mal wieder zeigen – nicht wirklich exakt reproduziert werden können, weil die Dokumentation der Ergebnisgenese unvollständig ist oder weil die zugrunde liegenden Daten nicht dokumentiert oder gar nicht zur Verfügung gestellt werden. Streng genommen sind solche Publikationen unwissenschaftlich. Und wenn eine Replikation gelingt, so hat man – außer im Falle der seltenen Aufdeckung spektakulärer Fälschungen – kaum eine Chance, die Replikationsergebnisse zu publizieren.

Der systematische Fehler in dieser Art von Objektivierung geistiger Leistung ist, dass eine formalisierte Qualitätskontrolle, das Review-Verfahren, welches den Herausgebern von Fachjournalen die Auswahl an Texten erleichtern soll, kontraindizierte Auswirkungen hat: Es unterdrückt konsequente Wissenschaftlichkeit, weil das Veröffentlichungssystem darauf ausgelegt ist, möglichst immer neue und innovative Analysen zu publizieren. Dies ist auch betriebswirtschaftlich aus der Sicht der Verlage sinnvoll. Für langweilige Replikationen ist – banalerweise – ganz simpel kein Platz. Als Forscher sollten wir aber genau diese nachfragen.

Immerhin: eine Art Quasi-Replikation von Forschungsergebnissen kann man in den Ingenieurwissenschaften und in den angewandten Naturwissenschaften beobachten: Die Replikation von Ergebnissen ergibt sich durch ihre Anwendungen im realen Leben sozusagen nebenbei. Wenn eine neuartig konstruierte Brücke stehen bleibt, dann ist die Tauglichkeit einer neuen Methode durch „Replikation“, nämlich den Brückenbau, bewiesen. Oder die Brücke stürzt ein; dann konnte die Tauglichkeit einer neuen Konstruktionsmethode nicht repliziert werden.

Aber: eine solche „Replikation durch Anwendung“ ist in der Grundlagenforschung nur selten möglich. Und in den Sozial-, Verhaltens- und Wirtschaftswissenschaften, deren Ergebnisse häufig Prognosen sind, haben Replikationsstudien keine Tradition, da Prognosen ohnehin sehr ungenau bzw. Anwendungen sehr selten sind. Wenn ein publiziertes Ergebnis nicht replizierbar ist, dann wird dies vielfach schulterzuckend hingenommen. Mit der impliziten Unterstellung, dass die mangelnde Replizierbarkeit nicht an einer schlechten Analyse liegt, sondern daran, dass die Wirklichkeit sich zwischenzeitlich verändert hat. Gerade in den Sozialwissenschaften wird munter – ggf. im Feuilleton – gestritten und es werden immer weitere neue

Hypothesen aufgestellt statt ernsthaft nach den Gründen für unterschiedlichen und prognostisch schwachen empirischen Befunde zu suchen.

Eine bessere Anerkennung von Replikationsstudien als wissenschaftliche Leistung und bessere Chancen der Publikation dieser, trügen dazu bei die Wissenschaftlichkeit empirischer Forschung deutlich zu erhöhen. Und diese Aufwertung von Replikationsstudien hätte einen weiteren, nicht zu unterschätzenden Nebeneffekt: jede Menge interessante Forschungsdaten würden quasi als Nebenprodukt für alle verfügbar und damit nicht nur für Replikationen verwendbar. Die Sekundäranalyse von Forschungsdaten unter Zuhilfenahme neuer Methoden oder unter anderen Annahmen und Theoriegerüsten kann in vielen Fällen genauso innovative Ergebnisse hervorbringen wie die Primärforschung, in deren Rahmen die Daten ursprünglich erhoben wurden.

Doch um dies zu erreichen, muss man alle relevanten Akteure ins Boot holen:

- Die **Wissenschaftler**: Sie haben als Einzelne nicht die Kraft sich aus den Zwängen ihrer Karrierebedingungen zu befreien. Ohne die Vielzahl der einzelnen Wissenschaftler kann aber kein besseres System entwickelt werden. Hier sind intensive Diskussionen innerhalb der Wissenschaftscommunities notwendig, auch im Hinblick auf die verschiedenen Rollen als Autoren und Reviewers, aber vor allem auch mit anderen Akteuren, insb. Forschungsförderern und Verlagen.
- Die **Forschungsförderer**: Sie müssen endlich konsequent darauf achten, dass am Ende einer Förderung auch sichergestellt wird, dass die Ergebnisse auch replizierbar sind (und erst damit wirklich wissenschaftliche Ergebnisse werden). Dies bedeutet, dass die Forschungsförderer in ihren Antragsverfahren vermehrt auf das Forschungsdatenmanagement achten müssen. Es sollte aber kein zusätzliches Geld für die Datenbereitstellung und Dokumentation für bereits weit fortgeschrittene oder abgeschlossene Projekte eingeworben werden müssen. Vielmehr sollte bereits in der Antragsstellungsphase darauf geachtet werden, dass es für jedes Projekt einen Datenmanagementplan gibt, in welchem der Umgang mit den Projektforschungsdaten und der Modus ihrer Dokumentation und Verfügbarmachung, einschließlich der Finanzierung dieser

Aufgaben, überzeugend dargelegt wird. Und die Nichteinhaltung eines Datenmanagementplans muss auch sanktioniert werden. Im schlimmsten Falle müsste ein Forscher von der weiteren Förderung ausgeschlossen werden bis er seine Daten archiviert und zugänglich gemacht hat.

- Die **Verlage** (zusammen mit den **Reviewern**): Wenn es handfester Anreize bedarf, wissenschaftlich korrekt zu arbeiten, dann muss man diese auch in der Währung der betreffenden Community auszahlen: Reputationsgewinne durch Publikationen. Man kann sogar provokativ fragen, ob Journals, die nicht darauf achten und durch ihre Policy dafür sorgen, dass die den Artikeln zu Grunde liegenden Daten für Replikationsstudien zur Verfügung stehen (und diese im Zweifel auch eine Chance haben, publiziert zu werden) überhaupt wissenschaftliche Journale sind. Nun, es gibt wenige Journale die genau dies tun und darauf achten, dass die Daten abgelegt und zugänglich gemacht werden. Aber es ist nach wie vor nahezu unmöglich eine Datensatzbeschreibung oder eine Replikationsstudie, die keine spektakulären Ergebnisse erbringt, in *Science* oder einem Triple A-Journal einer Fachdisziplin unterzubringen. Dies sind aber die Orte, an denen man als Forscher publizieren muss, will man seine Karriere befördern. Auch eine Zeitschrift wie *Science* sollte sich also fragen, ob sie eine Sektion für Replikationsstudien einrichtet. Und zwar nicht nur online, da das nicht genug Reputation bringt! Sondern in der Printausgabe!

3 Schlussfolgerungen

Um „Data Replication and Reproducibility“ von Forschungsergebnissen systematisch so sicherzustellen wie dies notwendig ist und in den Lehrbüchern steht, muss im Selbstverständnis und Agieren aller am Forschungs- und Publikationsprozess beteiligten Akteure sich etwas ändern, Und dafür engagieren sich Organisationen wie der RatSWD und seine Partner¹. Durch das Zusammenbringen aller Akteure wird versucht gemeinsame und dadurch von allen getragene Lösungen zu finden und Entwicklungen anzustoßen.

In einem eigenen Forschungsprojekt (EDaWax) versuchen Mitglieder des RatSWD zu ergründen unter welchen Bedingungen die oben genannten Akteure zusammengebracht werden können (vgl.: <http://www.edawax.de>).

In der Zusammenarbeit von Forschern, Forschungsförderern und Fachzeitschriften liegt der Schlüssel – ansonsten werden wir weiter ein Karussell der Begründungen dafür erleben, warum Wissenschaft im 21. Jahrhundert oft noch unwissenschaftlich ist und trotzdem niemand Schuld hat: Die Wissenschaftler alleine können es nicht ändern, weil sie Anreizsystemen ihrer Disziplinen unterliegen, diese sind wiederum mit den wirtschaftlichen Interessen der jeweiligen (Fach)verlage verknüpft, in deren Journals man publizieren muss. Dies hinzunehmen sollte der akademischen Elite eigentlich nicht ausreichen.

¹ Vgl. German Data Forum (ed.): Building on Progress: Expanding the Research Infrastructure for the Social, Economic and Behavioral Sciences. Budrich UniPress, Opladen & Farmington Hills, MI, 2011

Anhang

<http://www.sciencemag.org/site/special/data-rep/> (heruntergeladen am 9. Januar 2012)

TAKE THE POLL

Ideally, scientists would fully disclose their own raw data and methods and also spend time replicating others' work. What would best ensure this good behavior? (Poll Closed)

- Recognition and rewards from institutions 6.07%
- Funding earmarked for replication studies 19.1%
- More publication by journals of data that confirm or refute previous work 44.86%
- Rewards from funders on subsequent grant applications for depositing sufficient details for replication (or penalties for noncompliance) 24.49%
- Other (add your ideas in the comments section) 5.48%