

Monika Jungbauer-Gans, Corinna Kleinert, Jürgen Schupp, Mark Trappmann, Tobias Wolbring

Bekanntmachung: Regelung zur Archivierung von Daten und Analysefiles quantitativer Studien

Darüber, was gute Forschung ausmacht, wurde und wird viel gestritten. Eine Anforderung, der jedoch die meisten Wissenschaftlerinnen und Wissenschaftler – unabhängig von ihrer methodologischen Ausrichtung – zustimmen werden, ist, dass intersubjektiv nachprüfbar sein sollte, wie empirische Befunde zustande gekommen sind. Datengrundlage und analytisches Vorgehen sollten demgemäß in Publikationen oder dem entsprechendem Begleitmaterial transparent und präzise beschrieben werden. Im Idealfall werden andere Forschende dadurch in die Lage versetzt, auf Grundlage derselben Datenbasis zu den gleichen Ergebnissen zu gelangen (Reproduktion) bzw. auf Grundlage einer modifizierten, erweiterten oder neuen Datenbasis vergleichbare Analysen durchzuführen (Replikation).

Eine solche Reproduktions- und Replikationskultur ist – gerade in Zeiten von alternativen Fakten und Angriffen auf die Wissenschaft – aus mehreren Gründen wichtig. Erstens helfen Reproduktionen und Replikationen dabei, Fehler bei der Datenaufbereitung und -auswertung aufzudecken. Zweitens wird bei der wiederholten Nutzung von Sekundärdaten Zeit und Geld gespart – nicht jeder und jede Forschende muss das Rad neu erfinden; stattdessen kann kumulativ auf den Arbeiten Anderer aufgebaut werden. Drittens ist ein intersubjektiv nachvollziehbares Vorgehen der Aufdeckung grundsätzlich nicht ausschließbarer Datenfälschungen, Fehldarstellungen oder Ergebnismanipulationen dienlich. Die Dokumentation von Fällen, in denen publizierte Beiträge zurückgezogen wurden (siehe die Zusammenfassung unter <http://retractionwatch.com/>), zeigt, dass auch in der sozialwissenschaftlichen Forschung Betrug und Täuschung vorkommen, jedoch eher selten aufgedeckt werden. Häufiger als eine eklatante Manipulation ist aber die Generierung gefälliger, meist positiver Resultate durch Zuschneiden der Stichprobe, selektive Auswahl von Zitaten oder die flexible Nutzung von Kontrollvariablen. Ein Desiderat sozialwissenschaftlicher Forschung ist daher, Reproduktionen und Replikationen als festen Bestandteil des Forschungsprozesses zu etablieren und durch eine intersubjektiv nachvollziehbare Darstellung zu erleichtern.

Wer sich jedoch selbst schon einmal an einer Replikation versucht hat, weiß, dass es mit der Präzision des geschriebenen Worts manchmal so eine Sache ist. Beschreibungen sind vielfach unpräzise oder missverständlich, wichtige Details werden implizit als Selbstverständlichkeiten angenommen und manchmal wird auch einfach vergessen, bestimmte, randständig erscheinende Aspekte des Forschungsprozesses zu erläutern. Zudem erfordert es sowohl Zeit als auch Raum und trägt nicht unbedingt zur guten Lesbarkeit wissenschaftlicher Fachbeiträge bei, wenn alle

getroffenen Entscheidungen in aller Ausführlichkeit dargelegt werden. Gerade bei komplexen Analysedatensätzen, wie sie heutzutage häufig vorkommen, würde eine exakte Beschreibung aller Arbeitsschritte den Rahmen von Fachaufsätzen sprengen. In Folge ist es alleine auf Grundlage publizierter Artikel meist nicht möglich, Resultate einer Studie exakt zu reproduzieren. Kurzum: Auch wenn Autorinnen und Autoren darum bemüht sind, das Vorgehen transparent und vollständig im Manuskript darzustellen, bleiben der wissenschaftlichen Community oft Informationen über bestimmte (für die Ergebnisse potentiell folgenreiche) Entscheidungen der Forschenden vorenthalten.

Ein einfacher, aber effektiver Ansatz zur Sicherung der Reproduzierbarkeit und Schaffung von Transparenz im quantitativen Bereich ist es, die Autorinnen und Autoren empirischer Studien dazu zu verpflichten, die Rohdaten sowie die Dateien zur Datenaufbereitung und -auswertung interessierten Forschenden zur Verfügung zu stellen. Eine entsprechende Selbstverpflichtungserklärung wurde etwa von der *Zeitschrift für Soziologie* im Jahr 2002 eingeführt. Wie ein Lehrforschungsprojekt von Josef Brüderl (2013) an der Universität Mannheim und mehrere dadurch angeregte Replikationsseminare an der LMU München gezeigt haben, ist die Institution der Selbstverpflichtung jedoch bislang ein weitgehend zahnloser Tiger gewesen. Obwohl sich die *ZfS*-Autorinnen und Autoren schriftlich dazu verpflichtet hatten, die für eine Reproduktion notwendigen Dateien zu archivieren und Dritten auf Anfrage zugänglich zu machen, hatte ein Großteil der kontaktierten Autorinnen und Autoren entweder die Dateien nicht mehr vorrätig oder war nicht bereit, diese preiszugeben. Wurden die angeforderten Informationen zugänglich gemacht, war es den Studierenden und auch den Lehrenden dennoch nicht in allen Fällen möglich, die Befunde zu reproduzieren. Vereinzelt wurden auch fragwürdige Entscheidungen und eindeutige Fehler bei der Analyse offenbar, deren Behebung zu substanziel anderen Resultaten führte als in der als Vorbild dienenden Publikation. Die *ZfS* hat kürzlich auf diese Befunde reagiert und die reine Selbstverpflichtung durch ein verbindlicheres Verfahren ersetzt (siehe auch Hinz 2015). Dies illustriert, ebenso wie kürzlich erschienene Positionspapiere der Akademie für Soziologie (2019) und der Deutschen Gesellschaft für Soziologie (2019), trotz einiger offener Punkte eine Verbesserung der Replikationskultur in der deutschen Soziologie.

Eine solche Kultur ist dabei fraglos sowohl für qualitative als auch quantitative Studien wünschenswert. Während man sich jedoch im quantitativen Bereich bereits seit einiger Zeit auf Standards und Verfahren der Archivierung und Bereitstellung von Daten und Analysefiles verständigt hat, ist dieser Prozess für die qualitative Forschung noch im Gange und keineswegs abgeschlossen. Die Herausgeberinnen und Herausgeber der *Sozialen Welt* haben vor diesem Hintergrund entschieden, vorerst nur eine Archivierungspflicht für Forschungsergebnisse einzuführen, die auf standardisierten quantitativen Daten basieren. Wie die Nachvollziehbarkeit qualitativer Ergebnisse unter adäquater Berücksichtigung der methodologischen Überlegungen

qualitativer Forschung institutionell abgesichert werden kann, wird noch zu formulieren sein.

Die Vorgabe der Archivierung quantitativer Daten und Analysen folgt dem Vorbild anderer internationaler Fachzeitschriften, wie zum Beispiel *Sociological Methods and Research*, und geht über eine reine Selbstverpflichtung hinaus. Mit sofortiger Wirkung sind für Beiträge in der Sozialen Welt *vor der Publikation* die verwendeten Daten zusammen mit den Aufbereitungs- und Auswertungsdateien auf einen Replikationsserver hochzuladen. GESIS stellt bspw. die hierfür notwendige IT-Infrastruktur bereit und regelt den Zugang (www.replikationsserver.de). Auch andere Forschungsdatenzentren, die vom FDI-Ausschuss des Rates für Wirtschafts- und Sozialdaten akkreditiert wurden, verfügen über vergleichbare Replikationsserver. Dort hochgeladene Daten und Analysefiles stehen nicht frei verfügbar im Netz, sondern werden individuell jeweils auf Anfrage und üblicherweise erst nach Vorliegen des erneuten Einverständnisses der Autorinnen und Autoren bereitgestellt. Im Falle der Nutzung von für Forschungszwecke zugänglicher Scientific-Use-Files oder anderer Daten mit alternativen Archivierungskonzepten (z.B. Inhouse-Zugang zu Replikationszwecken) entfällt die Archivierung der Daten, aber nicht die Bereitstellung der Dateien, die von den Rohdaten zu den publizierten Resultaten führen. Zudem muss die eindeutige Zitation des verwendeten Datensatzes möglichst mit Hilfe von persistenten Identifikatoren (z.B. DOI oder handle) in der Publikation und den Aufbereitungsskripten enthalten sein. Im Falle datenschutz- oder eigentumsrechtlicher Einschränkungen, welche die Weitergabe von Rohdaten verhindern, sind alternative Wege der Ermöglichung von Replikationen mit der Redaktion im Vorfeld abzuklären und schriftlich festzuhalten. Nutzerinnen und Nutzer des bereit gestellten Materials werden im Falle von Publikationen, in denen Daten, Operationalisierungen und Analyse-Codes ganz oder teilweise übernommen werden, dazu angehalten, dem Urheberrecht durch Zitation Rechnung zu tragen und die verwendeten Quellen in ihrem Literaturverzeichnis aufzuführen.

Schließlich möchten wir unsere Leserinnen und Leser dazu ermutigen, das auf diesem Wege gesammelte Material zu nutzen, Reproduktionen und Replikationen publizierter Studien zu wagen und übereinstimmende sowie abweichende Befunde entweder bei der *Sozialen Welt* oder auch anderen Zeitschriften einzureichen. Die *Soziale Welt* lebt, ebenso wie Wissenschaft im Allgemeinen, von Kritik und Debatten. Die hier vorgestellte Regelung zur Archivierung kann eine solche Kultur befördern. Um jedoch – abgesehen von einer abschreckenden Wirkung gegenüber bewusst vorgenommenen Fälschungen – den intendierten Effekt zu erzielen, ist es unabdingbar, dass inhaltlich und methodisch geschulte Forscherinnen und Forscher einen genaueren Blick auf Daten, Operationalisierung und Analysen werfen und die Belastbarkeit von Befunden kritisch hinterfragen.

Literatur

Akademie für Soziologie (2019): Wissenschaftliche Daten sind kein Privateigentum einzelner For- schender, sondern ein kollektives Gut. Die Bereitstellung von Forschungsdaten zur Nachnut- zung und Replikation muss auch in der Soziologie die Norm sein. Positionspapier, Januar 2019.

Brüderl, Josef (2013): Sind die Sozialwissenschaften wissenschaftlich? Ergebnisse eines Replikati- onsexperiments. Vortrag am 19.11.2013 auf der Tagung „Rational-Choice-Soziologie“ an der Venice International University.

Deutsche Gesellschaft für Soziologie (2019): Bereitstellung und Nachnutzung von Forschungsda- ten in der Soziologie. Stellungnahme des Vorstands und Konzils der DGS, Stand: 8.1.2019.

Hinz, Thomas / Rössel, Jörg / Maiwald, Kai-Olaf / Vollmer, Hendrik / Wobbe, Theresa (2015): Editorial 2015, in: Zeitschrift für Soziologie 44(1), S. 2-5.