

Offene Daten als Grundlage für wissenschaftliche Innovation und öffentliches Vertrauen



Positionspapier des German Reproducibility Networks, verfasst von Ulf Tölch und Felix Schönbrodt
([CC-BY 4.0](#))

Wissenschaftliche Erkenntnisse dienen als Grundlage für politische Entscheidungen in gesellschaftsrelevanten Bereichen wie der Pandemiebekämpfung oder der Bewältigung der Klimakrise. Hierbei sind *offen zugängliche* wissenschaftliche Daten ein zentraler Baustein, der auch im aktuellen Koalitionsvertrag besondere Erwähnung findet. Offene Daten – also Daten, die ganz frei oder mindestens für wissenschaftliche Nachnutzung zugänglich sind – ermöglichen Reproduzierbarkeit, Transparenz, Transfer und angewandte Innovationen und stärken das Vertrauen in wissenschaftliche Erkenntnisse in der Öffentlichkeit. Darüber hinaus ist es eine Frage der Effizienz und Verantwortlichkeit, dass mit Steuermitteln erhobene Daten zur Nachnutzung zur Verfügung stehen. Hierbei müssen selbstverständlich konkurrierende Güter abgewogen werden, die gegen eine Offenlegung sprechen, wie der Datenschutz bei personenbezogenen oder Schutzrechte bei potentiell kommerziell verwertbaren Daten. Unter diesen Voraussetzungen fordert bereits eine Vielzahl von Wissenschaftsorganisationen, Geldgebern und Fachzeitschriften die offene Bereitstellung von Forschungsdaten.

Unzugängliche Forschungsdaten zeigen Fehlentwicklungen im Wissenschaftssystem auf

Viele wertvolle Daten bleiben trotz dieser Vorteile und Notwendigkeiten unzugänglich, da in der Leistungsbewertung von Forschenden Publikationen im Fokus stehen. Diese bilden aber nur einen kleinen Ausschnitt wissenschaftlicher Arbeit ab. Die eigentlichen Grundlagen einer Publikation – wie zum Beispiel ausführliche Versuchsprotokolle sowie Daten und deren transparente Analyse – stehen als Indikatoren für die Bewertung von Forschungsleistungen im Hintergrund. Die kurze Laufzeit vieler Verträge in der Wissenschaft verschärft dieses Problem weiter (*#ichbinhanna*), da Zeit, die für sachgerechte Kuratierung und Offenlegung von Daten aufgewendet wird, nicht für neue Experimente und Publikationen zur Verfügung steht. Daten bereitzustellen kann unter einer publikationsfokussierten Anreizstruktur paradoxerweise zu einem Karrierehemmnis

werden. Unter den falschen Anreizbedingungen erfüllen jedoch die besten Dateninfrastrukturen nicht ihren Zweck, wenn diese nicht mit Daten befüllt werden.

Konkrete Schritte sind also notwendig, um mit Hilfe öffentlicher Gelder gewonnene Daten offen zur Verfügung zu stellen, um diese als einen Treiber für Innovation zu nutzen.

Konkrete Forderungen richten sich an Förderer, Forschungseinrichtungen und Politik:

(1) Die offene und zeitnahe Bereitstellung von Forschungsdaten soll der Standard bei allen wissenschaftlichen Publikationen und abgeschlossenen Projekten sein.

Mit jeder Publikation und nach Abschluss eines Forschungsprojektes sollten die jeweiligen Forschungsdaten offen verfügbar sein. Das heißt, sie sollten idealerweise sowohl durch Menschen als auch Algorithmen auffindbar und nachnutzbar sein ([FAIR](#)-Prinzipien), das heißt auch, entsprechend fachspezifischen Standards dokumentiert sein und in einem geeigneten, nicht-kommerziellen Repositorium langfristig gesichert werden. Sie sollten grundsätzlich so offen wie möglich sein, das heißt ohne Beschränkung in Bezug auf Zugang oder Nachnutzung. Die zeitnahe Bereitstellung ermöglicht der wissenschaftlichen Gemeinschaft, wie bereits heutzutage bei Preprints (Vorabveröffentlichungen) von Publikationen, früh Daten zu bewerten und entsprechend in eigenen Arbeiten darauf zu reagieren (wie es in den letzten beiden Jahren in der COVID-19-Forschung beispielhaft stattgefunden hat).

(2) Es sollen abgestufte Zugriffsklassenkonzepte angewandt werden, wenn Daten nicht vollständig offengelegt werden können.

Nur in begründeten Ausnahmefällen, wie der Wahrung von Schutzrechten, Datenschutz oder Persönlichkeitsrechten, sollten der Zugang und die Nachnutzung angemessen eingeschränkt werden. In diesen Fällen sollen abgestufte Zugriffsklassenkonzepte verwendet werden, so dass zumindest ein beschränkter Zugang für wissenschaftliche Zwecke ermöglicht wird. Die Entscheidung über den Zugang muss in diesen Fällen durch Institutionen und nicht durch individuelle Forschende getroffen werden (data use and access). Ausschüsse für Datennutzung und -zugang sollen zukünftig den Zugang zu sensitiven Daten regeln oder urheberrechtliche Bedenken prüfen. Dass dies ein gangbarer Weg unter Einbeziehung der Forschenden ist, zeigen große Drittmittelgeber wie der Wellcome Trust in Großbritannien und ab 2023 auch das National Institute of Health in den USA. Hier werden auch weitere – über die eigentliche Projektförderung hinausgehende – Mittel zur Verfügung gestellt, die nötig sind, um den zusätzlichen Aufwand und die Infrastruktur für Datenbereitstellung zu finanzieren.

(3) Die Bereitstellung von Daten soll als zusätzliches Bewertungskriterium für Forschungsleistung herangezogen werden.

Wissenschaftliche Leistung zeigt sich nicht nur in Publikationen, sondern in einer [Vielzahl von Aktivitäten](#), wie z.B. der Erstellung und Pflege von Software, Methodenentwicklung, gesellschaftlichem Engagement oder der Lehre. Daher soll die Veröffentlichung von Daten als Teil einer guten wissenschaftlichen Praxis in die Bewertung von Forschungsleistung einfließen, zum Beispiel bei der Evaluierung von Förderanträgen oder

Berufungskommissionen. Als wichtiger Baustein einer Vermessung *verantwortungsvoller Wissenschaft* bilden offene und nachnutzbare Daten die Grundlage für komplementäre Bewertungsdimensionen wie zum Beispiel Innovationspotenzial und Exzellenz.

Strukturierte Lebensläufe sollten neben der üblichen Auflistung der, zum Beispiel, 10 besten Publikationen auch eine Auflistung der veröffentlichten Datensätze beinhalten, mit einer kurzen Beschreibung der Nachnutzung.

(4) Drittmittelgeber sollten offene Forschungsdaten zur Bedingung für Förderungen machen, die erfolgreiche Bereitstellung abfragen und stichprobenartig prüfen.

Die großen öffentlichen Drittmittelgeber wie die Deutsche Forschungsgemeinschaft oder das Bundesministerium für Bildung und Forschung haben Richtlinien für die Veröffentlichung von Forschungsdaten für geförderte Projekte entwickelt. Die tatsächliche Bereitstellung sollte zukünftig im Abschlussbericht abgefragt werden und die Umsetzung zumindest stichprobenartig geprüft werden. Eine Nichtbeachtung würde einen Verstoß gegen die gute wissenschaftliche Praxis darstellen.

(5) Universitäten und Forschungseinrichtungen müssen rechtliche (z.B. in Bezug auf Datenschutz, Lizenzen, Fragen des Urheberrechts) und technische (z.B. in Bezug auf Metadatendokumentation) Unterstützung anbieten.

Auf institutioneller Ebene werden die Voraussetzungen für die Offenlegung von Daten gelegt. Hierfür müssen Institutionen sicherstellen, dass Infrastrukturen vorhanden und zugänglich sind (z.B. European Open Science Cloud, Nationale Forschungsdateninfrastruktur, re3data.org). Durch die verschiedenen Anforderungen sind hier spezialisierte Beratungsangebote und Unterstützung erforderlich. Ein Vorreiter ist hier die TU Delft, die sogenannte [data stewards](#) implementiert hat. Diese unterstützen die Forschenden aktiv bei der Kuratierung und späteren Veröffentlichung ihrer Daten.

Das GRN fordert alle angesprochenen Akteure in Politik und Institutionen auf, sich zeitnah mit der Umsetzung dieser Forderungen zu beschäftigen und steht gerne als Berater und Ansprechpartner insbesondere für die Anwenderperspektive zur Verfügung.

Über das German Reproducibility Network (GRN)

Das interdisziplinäre Deutsche Reproduzierbarkeitsnetzwerk (GRN) will die Vertrauenswürdigkeit und Transparenz wissenschaftlicher Forschung erhöhen, indem es die Faktoren untersucht und fördert, die zu robuster Forschung beitragen. Wir fördern Schulungsaktivitäten, führen meta-wissenschaftliche Forschung durch und arbeiten mit verschiedenen Interessengruppen zusammen (z. B. Politik, Geldgeber, Verlage, Institutionen und akademische Gesellschaften), um die Koordination der Bemühungen sicherzustellen.

Kontakt: info@reproducibilitynetwork.de

PD Dr. Ulf Tölch: Berlin Institute of Health at Charité – Universitätsmedizin Berlin, QUEST Center for Responsible Research

Prof. Dr. Felix Schönbrodt: Ludwig-Maximilians-Universität München, LMU Open Science Center

Links:

- Rechtliche Beratungsangebote und -strukturen im universitären Forschungsdatenmanagement: <https://zenodo.org/record/5878779#.YfFUFi9XaXo>
- <https://osp.od.nih.gov/2022/01/25/gearing-up-for-2023-implementing-the-nih-data-management-and-sharing-policy/>
- <https://wellcome.org/what-we-do/our-work/open-research>
- <https://www.tudelft.nl/en/library/research-data-management/r/support/data-stewardship>
- <https://osec2022.eu/paris-call/>