

# handbuch.io

■ CH

## Handbuch CoScience/Publication von Forschungsdaten

From Handbuch.io

< Handbuch CoScience

DOI: [10.2314/cosc2.53](https://doi.org/10.2314/cosc2.53)

Link zur Open Video Lecture "Open Access und offene Datenpublikation":

<http://dx.doi.org/10.5446/15243#t=30:11,44:47>

Autoren: Heinz Pampel, Janna Neumann

Kontributoren: Martin Fenner, Marco Tullney

Vermeehrt wird in der Wissenschaft über den Umgang mit Forschungsdaten diskutiert. Dank der rasanten technologischen Entwicklung können Forschungsdaten über das Internet verfügbar und nachnutzbar gemacht werden. Verbunden damit ist insbesondere die Erwartung, dass Forschungsergebnisse besser überprüft werden können und dass bereits erhobenen und verwendeten Daten für weitere Forschungsvorhaben nachgenutzt werden können. Zusammengefasst:

- **Nachnutzung:** Im Zentrum steht hier die Erwartung, dass einzigartige Daten in anderen Kontexten wiederverwertet werden. Eng verbunden ist damit die Hoffnung, die Effizienz der Forschung zu steigern.<sup>[1]</sup>
- **Nachprüfbarkeit:** Anliegen ist hier, die Transparenz der Forschung zu erhöhen, indem jede und jeder die Daten überprüfen kann.

Der folgende Text beschreibt zehn Punkte, die beim Umgang mit Forschungsdaten bedacht werden sollten.

### Contents

- 1 Was erwarten Förderorganisationen?
- 2 Welche Anforderungen haben Journals?
- 3 Wie organisiere ich meine Daten?
- 4 Wie sichere ich meine Daten?
- 5 Soll ich meine Daten veröffentlichen?
- 6 Welche Vorteile hat das Data Sharing?
- 7 Welche Nachteile hat das Data Sharing?
- 8 Wie finde ich ein Repositorium?
- 9 Welchen Vorteil bieten Data Journals?
- 10 Wo finde ich weitere Informationen?
- 11 Referenzen
- 12 Einzelnachweise

## Was erwarten Förderorganisationen?

Immer häufiger werden Forschende mit sogenannten Data Policies konfrontiert.<sup>[2]</sup> In diesen formulieren Förderorganisationen Anforderungen an wissenschaftliche Daten, die im Rahmen von geförderten Projekten entstehen.

Bereits 2003 veröffentlichte das National Institute of Health (NIH) (<http://www.nih.gov/>) eine 'NIH Data Sharing Policy'. Darin werden Antragstellerinnen und Antragsteller, die eine Zuwendung ab 500.000 US-Dollar beantragen, aufgefordert, Aussagen zur Art und Weise des Data Sharings zu tätigen: „The NIH expects and supports the timely release and sharing of final research data from NIH-supported studies for use by other researchers“.<sup>[3]</sup>

Die Deutsche Forschungsgemeinschaft (DFG) (<http://www.dfg.de/>) hat das Thema seit 2010 in ihrem 'Leitfaden für Antragsteller' verankert. In diesem Leitfaden werden Antragstellerinnen und Antragsteller aufgefordert, Maßnahmen zum Umgang mit den entstehenden Daten darzulegen. In der Version 04/14 des DFG-Leitfadens heißt es dazu:

„Wenn aus Projektmitteln systematisch Forschungsdaten oder Informationen gewonnen werden, die für die Nachnutzung durch andere Wissenschaftlerinnen und Wissenschaftler geeignet sind, legen Sie bitte dar, ob und auf welche Weise diese für andere zur Verfügung gestellt werden. Bitte berücksichtigen Sie dabei auch - sofern vorhanden - die in Ihrer Fachdisziplin existierenden Standards und die Angebote existierender Datenrepositorien oder Archive.“<sup>[4]</sup>

Auch informiert eine Webseite der DFG über „Anregungen und Best-Practice Beispiele“ zur Umsetzung dieser Vorgabe.<sup>[5]</sup>

Darüber hinaus gewinnen sogenannte 'Data Management Plans' an Bedeutung. In diesen müssen Antragsteller detaillierte Informationen zum Umgang mit den im Rahmen eines beantragten Projektes entstehenden Daten beschreiben. So wird z.B. im Rahmen des europäischen 'Rahmenprogramms für Forschung und Innovation HORIZON 2020' ein 'Research Data Pilot ([http://europa.eu/rapid/press-release\\_IP-13-1257\\_en.htm](http://europa.eu/rapid/press-release_IP-13-1257_en.htm))' erprobt. In den 'Guidelines on Data Management in Horizon 2020'<sup>[6]</sup> beschreibt die Europäische Kommission das Anliegen der Data Management Plans:

„A further new element in Horizon 2020 is the use of Data Management Plans (DMPs) detailing what data the project will generate, whether and how it will be exploited or made accessible for verification and re-use, and how it will be curated and preserved. The use of a Data Management Plan is required for projects participating in the Open Research Data Pilot. Other projects are invited to submit a Data Management Plan if relevant for their planned research.“

Die Vorgaben der Europäischen Kommission sehen vor, dass Antragstellerinnen und Antragsteller in den Data Management Plans Angaben zu den folgenden Punkten machen:

- Data set reference and name
- Data set description
- Standards and metadata
- Data sharing
- Archiving and preservation (including storage and backup)

Forschende sind daher aufgefordert, bereits bei der Konzeption eines Antrags zu überlegen, welche Maßnahmen ergriffen werden müssen, um den Zugang zu den Daten auch über das Projektende hinaus zu sichern (s.a. Abschnitt 3 Wie organisiere ich meine Daten?).

## Welche Anforderungen haben Journals?

Ähnliche Anforderungen werden in den *data policies* von wissenschaftlichen Zeitschriften formuliert. So erwartet z.B. der Open-Access-Verlag PLOS (<http://www.plos.org/>), dass die Daten, die Grundlage eines eingereichten Aufsatz sind, wenn möglich frei zugänglich gemacht werden:

„PLOS journals require authors to make all data underlying the findings described in their manuscript fully available without restriction, with rare exception.“<sup>[7]</sup>

Bei der Nature Publishing Group (<http://www.nature.com>) heißt es dazu: „authors are required to make materials, data and associated protocols promptly available to readers without undue qualifications“.<sup>[8]</sup>

Bei PLOS müssen Autorinnen und Autoren darüber hinaus in einem „Data Availability Statement“ die Zugänglichkeit und Nachnutzbarkeit der Daten beschreiben: „When submitting a manuscript online, authors must provide a Data Availability Statement describing compliance with PLOS’s policy. The data availability statement will be published with the article if accepted.“<sup>[9]</sup>

## Wie organisiere ich meine Daten?

Gerade in kleineren Forschungsprojekten ist der nachhaltige Umgang mit den erhobenen Daten häufig nicht geklärt. Verlässt eine Mitarbeiterin oder ein Mitarbeiter, die/der mit der Auswertung der Daten beschäftigt war, eine wissenschaftliche Einrichtung, gehen häufig auch die Daten - und damit auch das Wissen über die Daten - verloren. Vor diesem Hintergrund ist es wichtig, bereits bei Projektbeginn zu klären, wie mit den Daten umgegangen werden soll. Die folgenden fünf Fragen können dabei helfen, einen organisierten Umgang mit den Daten sicherzustellen:

- **Wer** soll die Daten nach Projektende noch nutzen können?
- **Wie** wird sichergestellt, dass die Daten auch nach Projektende noch nutzbar sind?
- **Was** gibt es im Speziellen zu beachten? (Z.B. juristische oder ethische Besonderheiten.)
- **Wann** sollen die Daten anderen Personen zugänglich gemacht werden?
- **Wo** sollen die Daten gespeichert und zugänglich gemacht werden?

In der jüngeren Vergangenheit wurden einige hilfreiche Tools entwickelt, die Forschende bei der Formulierung von Aussagen zum Umgang mit Daten unterstützen, z.B. das 'Data Management Plan Tool' (<https://dmp.cdlib.org/>) des Curation Center (<http://www.cdlib.org/>) der University of California oder das Tool 'DMPonline' (<https://dmponline.dcc.ac.uk/>) des britische Digital Curation Centre

(<http://www.dcc.ac.uk/>). Weitere hilfreiche Informationen inklusive einer Checkliste zum Thema finden sich im ‚Leitfaden zum Forschungsdaten-Management‘<sup>[10]</sup>.

## Wie sichere ich meine Daten?

Während in größeren Projekten der Umgang mit den erhobenen Daten häufig geklärt ist, stellt diese Frage gerade kleinere Projekte vor Herausforderungen. Häufig gibt es keine dauerhaften Infrastrukturen zur Speicherung der Daten.

In kollaborativen Projekten werden häufig gemeinsame Netzlaufwerke oder auch Cloud-Lösungen wie Dropbox verwendet, um Daten mit Kolleginnen und Kollegen gemeinsam zu verwalten. Diese Tools bieten jedoch mehrheitlich keine nachhaltige Lösung, die die dauerhafte Speicherung der Daten garantiert.

Vor diesem Hintergrund empfiehlt sich die Suche nach einem verlässlichen Repositorium, dass die Daten dauerhaft und verlässlich speichert (s. Abschnitt 8: Wie finde ich ein Repositorium?).

## Soll ich meine Daten veröffentlichen?

Um den Zugriff und die Nachnutzung von Forschungsdaten zu gewährleisten, ist es sinnvoll, ergänzend zu einem 'traditionellen' wissenschaftlichen Artikeln auch die dazugehörigen Daten, auf denen die Forschungsergebnisse basieren, zu veröffentlichen.

Die Publikation und Auffindbarkeit Ihrer Forschungsdaten über nationale und internationale Informationsportale kann einen wesentlichen Beitrag zum Reputationsgewinn der Datenproduzierenden - auch über die Grenzen Ihrer eigenen wissenschaftlichen Fachcommunity hinaus - leisten.

Durch die Veröffentlichung der Daten, die Grundlage eines Aufsatzes sind, können diese zeitnah nachgenutzt und zitiert werden. Die Zitierung der Daten durch andere Forschende erhöht die Sichtbarkeit und kann die Reputation der Forschung stärken.

Es ist stets zu klären, wie eine mögliche Nachnutzung von Forschungsdaten durch Dritte rechtlich abgesichert werden kann. Die zugelassene Nachnutzung kann von einer völlig freien bis zu einer eher eingeschränkten Nutzung variieren und auf verschiedene Weise festgelegt werden. Dabei müssen insbesondere datenschutzrechtliche und ethische Aspekte beachtet werden (s. a. Freie Lizenzen und Nachnutzung).

## Welche Vorteile hat das Data Sharing?

Welche Vorteile das Teilen und Nachnutzen von Forschungsdaten hat, wird in der Illustration des ‚Journals of Open Archaeology Data‘ dargestellt:

Bei der Abwägung der Vor- und Nachteile kommt es nicht nur darauf an, die Vorteile des Einzelnen, sondern auch die der (Fach-)Community und der interessierten Öffentlichkeit zu beachten.

Sobald Forschungsdaten geteilt werden, wird die eigene Forschung für andere nachvollziehbar. Sie wird reproduzierbar und kann verifiziert werden. Dies ist ein zentraler Punkt bei der Umsetzung der guten wissenschaftlichen Praxis.

Zusätzlich bietet der offene Zugang zu Forschungsdaten neue Möglichkeiten für die wissenschaftliche Kollaborationen. Dies ist ein wesentlicher Gesichtspunkt beim Thema 'CoScience'. Zum Beispiel können Forschende weltweit über das Netz Daten aus der eigenen oder einer verwandten Disziplin suchen und auffinden und darüber hinaus mit den jeweiligen Daten-ProduzentInnen in den Dialog treten, um z.B. gemeinsam Forschungsvorhaben weiterzuentwickeln. Die Zusammenarbeit kann so auch über die eigene Fachdisziplin hinaus gefördert und um interdisziplinäre Ideen angereichert werden.

Ein weiterer Vorteil des Data Sharing liegt in der Möglichkeit der Nachnutzung gut dokumentierter Daten. Dieses Vorgehen kann beispielsweise bei der wissenschaftlichen Arbeit Zeit einsparen und damit einen effizienteren Umgang mit Forschungsergebnisse ermöglichen, da diese nicht immer wieder von Neuem erhoben werden müssen (s. a. Nutzen Sie Forschungsdaten nach).

Error creating thumbnail: File missing

Abbildung: Benefits for data sharing - Published by Ubiquity Press  
(<http://openarchaeologydata.metajnl.com/about%7C>) - Creative Commons Attribution 3.0 Unported (CC BY 3.0) (<http://creativecommons.org/licenses/by/3.0/%7C>)

## Welche Nachteile hat das Data Sharing?

Neben den Vorteilen für das Data Sharing soll hier ein kurzer Blick auf mögliche Nachteile geworfen werden. Ob die genannten Nachteile jedoch ein tragendes Argument gegen das Teilen von Forschungsdaten sind, muss eine Wissenschaftlerin bzw. ein Wissenschaftler im Einzelfall entscheiden. Hier sollte immer die kritische Frage im Fokus stehen, ob das Teilen einen Mehrwert für die 'Community' darstellt oder ob dem Teilen nur der eigene 'Forschungsegoismus' im Weg steht.

Ein oft hervorgebrachter Nachteil beim Data Sharing ist, dass öffentlich zugängliche Daten von anderen für weitere Interpretationen offenstehen und damit verbunden die Befürchtung, dass andere Forscherinnen und Forscher basierend auf den offenen Daten eigene Publikationen produzieren. Deshalb ist es wichtig, dass die Daten auf vertrauenswürdigen Infrastrukturen (Repositorien) zugänglich gemacht werden, bei denen deutliche Zitationshinweise angegeben sind (z. B. die Nennung des Digital Object Identifier - DOI - eines Datensatzes). Dank eines solchen Zitationshinweises kann sichergestellt werden, dass die/der ursprüngliche Daten-Produzentin oder Daten-Produzent Anerkennung - im Rahmen einer Zitation - für sein Data Sharing erhält. Darüber hinaus kann durch die Anwendung von freien Lizenzen der Grad der Nachnutzung festgelegt werden.

Weiter müssen beim Data Sharing rechtliche und ethische Aspekte berücksichtigt werden. So müssen z.B. personenbezogenen Daten auf Grund des Datenschutzes anonymisiert werden, bevor sie geteilt werden können (s. a. Freie Lizenzen und Nachnutzung). Dieser Prozess kann sehr aufwändig sein und somit eine Hürde beim Data Sharing darstellen.

## Wie finde ich ein Repository?

Auf nationaler und internationaler Ebene haben sich in einigen Fachdisziplinen bereits Repositorien zur dauerhaften Speicherung von Forschungsdaten etabliert. Z.B.: Dryad (<http://datadryad.org/>) in der Biodiversitätsforschung, Genbank (<https://www.ncbi.nlm.nih.gov/genbank/>) in der Genetik oder PANGAEA (<http://www.pangaea.de/>) in den Geowissenschaften.

Wissenschaftlerinnen und Wissenschaftler sollten bei der Auswahl eines geeigneten Repositoriums u.a. folgende Aspekte berücksichtigen:

- Passt die Ausrichtung des Forschungsdatenrepositoriums auf die Datensätze der eigenen Fachdisziplin?
- Ist eine langfristige Verfügbarkeit des Repositoriums (z.B. im Hinblick auf Finanzierung und beteiligte Institute/Personen) gewährleistet?
- Welche Services (z.B. Open Access, persistente Identifikatoren) bietet das Repository an?
- Welche Datenüberlassungsvereinbarungen (für den Datengeber) und Datennutzungsvereinbarungen (für den Datennutzer) gibt es?
- Welche Metadatenstandards werden verwendet?
- Wie ist die Anbindung des Repositoriums an (fachspezifische) Suchportale?

Mit der Zunahme der Archivierungsmöglichkeiten für wissenschaftliche Datensätze stehen Forschenden zunehmend auch übergreifende Informationsportale wie z. B. das re3data.org - Registry of Research Data Repository (<http://www.re3data.org/>) zur Verfügung. Dieses Portal erlaubt u. a. die disziplinspezifische Suche nach geeigneten Repositorien und gibt z. B. Auskunft über Akkreditierung, Verfügbarkeit sowie angebotene Servicemodelle eines Repositoriums.<sup>[11]</sup>

## Welchen Vorteil bieten Data Journals?

Data Journals widmen sich der Beschreibung von Datensätzen. Diese Beschreibungen werden häufig Data Papers oder auch Data Descriptors genannt. In diesen Artikeln werden zentrale Parameter eines Datensatzes beschrieben. So dokumentieren *Data Papers* z.B. den Erhebungsprozess der Daten und geben detaillierte Auskunft über verwendeten Instrumente und Methoden. Darüber hinaus stellen diese Artikel einen beschriebenen Datensatz in Beziehung zu weiteren relevanten Forschungsergebnissen.

Im Rahmen eines Peer-Review-Verfahrens stellen die Data Journals dank Expertinnen und Experten die wissenschaftliche Qualität der beschriebenen Daten auf Basis der jeweiligen Community-Standards sicher. Die Kriterien für die Einreichung von Data Papers und den zugehörigen Daten werden in Policies für die Publikation festgelegt.

Zentrales Anliegen der Data Papers ist es, möglichst alle relevanten Angaben für eine Nachnutzung eines beschriebenen Datensatzes zu liefern. Der Datensatz selbst wird auf einem Repository gespeichert. Das Data Paper verweist auf das Repository und den dort gespeicherten Datensatz. Werden die Daten nachgenutzt, muss der Nachnutzende das Data Paper und den Datensatz zitieren. Dank dieser Praxis erhält die wissenschaftliche Leistung der Person, die die Daten teilt, Anerkennung.

Aktuell gibt es eine Gründungswelle von Data Journals. Beispiele für Data Journals sind:

- Earth System Science Data (ESSD) (<http://www.earth-system-science-data.net/>) im Verlag Copernicus Publications (<http://publications.copernicus.org/>)
- Journal of Open Archaeology Data (JOAD) (<http://openarchaeologydata.metajnl.com/>) im Verlag Ubiquity Press (<http://www.ubiquitypress.com/>)
- Scientific Data (<http://www.nature.com/scientificdata/about/>) im Verlag Nature Publishing Group (<http://www.nature.com/>)

Eine Liste von Data Journals findet sich im Wiki Forschungsdaten.org ([http://www.forschungsdaten.org/index.php/Data\\_Journals](http://www.forschungsdaten.org/index.php/Data_Journals)).

## Wo finde ich weitere Informationen?

Weiterführende Informationen zum Thema finden sich u.a. im Wiki Forschungsdaten.org (<http://www.forschungsdaten.org/%7C>) oder auch im Handbuch Forschungsdatenmanagement (<http://www.forschungsdatenmanagement.de/%7C>).

## Referenzen

- Pampel, H., & Dallmeier-Tiessen, S. (2014). Open Research Data: From Vision to Practice. In S. Bartling & S. Friesike (Eds.), *Opening Science. The Evolving Guide on How the Internet is Changing Research, Collaboration and Scholarly Publishing* (pp. 213–224). Heidelberg: Springer. doi:10.1007/978-3-319-00026-8.
- Tenopir, C. et al. (2011). Data Sharing by Scientists: Practices and Perceptions, *PLOS One* 6, no. 6 (2011): e21101. doi:10.1371/journal.pone.0021101
- The Royal Society. (2012). Science as an open enterprise. The Royal Society Science Policy Centre report 02/12. [http://royalsociety.org/uploadedFiles/Royal\\_Society\\_Content/policy/projects/sape/2012-06-20-SAOE.pdf](http://royalsociety.org/uploadedFiles/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf)

## Einzelnachweise

1. Organisation for Economic Co-operation and Development. (2007). OECD Principles and Guidelines for Access to Research Data from Public Funding. Retrieved from <http://www.oecd.org/dataoecd/9/61/38500813.pdf>
  2. Pampel, Heinz; Bertelmann, Roland (2011). „Data Policies“ im Spannungsfeld zwischen Empfehlung und Verpflichtung. In S. Büttner, H.-C. Hobohm, & L. Müller (Eds.), *Handbuch Forschungsdatenmanagement* (pp. 49–61). Bad Honnef: Bock + Herchen. urn:nbn:de:kobv:525-opus-2287 (<http://nbn-resolving.de/urn:nbn:de:kobv:525-opus-2287>)
  3. NIH (2003). Final NIH Statement on Sharing Research Data. <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html>
  4. Leitfaden für die Antragstellung. DFG-Vordruck 54.01 – 04/14. [http://www.dfg.de/formulare/54\\_01/](http://www.dfg.de/formulare/54_01/)
  5. [http://www.dfg.de/foerderung/antragstellung\\_begutachtung\\_entscheidung/antragstellende/antragstellung/sicherung\\_forschungsdaten/](http://www.dfg.de/foerderung/antragstellung_begutachtung_entscheidung/antragstellende/antragstellung/sicherung_forschungsdaten/) Zugriff: 30.04.2014
  6. Europäische Kommission (2014): Guidelines on Data Management in Horizon 2020. [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf)
  7. PLOS (2013): Data Access for the Open Access Literature: PLOS's Data Policy. <http://www.plos.org/data-access-for-the-open-access-literature-ploss-data-policy/>
  8. Nature (2013). Availability of data and materials. <http://www.nature.com/authors/policies/availability.html>
  9. PLOS (2013): Data Access for the Open Access Literature: PLOS's Data Policy. <http://www.plos.org/data-access-for-the-open-access-literature-ploss-data-policy/>
  10. Ludwig, Jens; Enke, Harry (Eds.). (2013). Leitfaden zum Forschungsdaten-Management. Handreichungen aus dem WissGrid-Projekt. Glückstadt: Verlag Werner Hülsbusch. Retrieved from [http://www.wissgrid.de/publikationen/Leitfaden\\_Data-Management-WissGrid.pdf](http://www.wissgrid.de/publikationen/Leitfaden_Data-Management-WissGrid.pdf)
  11. Pampel H. et al. (2013) Making Research Data Repositories Visible: The re3data.org Registry. *PLoS ONE* 8(11): e78080 doi:10.1371/journal.pone.0078080
- Retrieved from 'https://test.handbuch.tib.eu/w/index.php?title=Handbuch\_CoScience/Publikation\_von\_Forschungsdaten&oldid=3400'

