

A hand is shown at the bottom, holding a globe. The globe is composed of a dense network of colorful lines (red, blue, yellow, green, purple) that represent data connections or a network. Small circular nodes in various colors are scattered across the globe, connected by the lines. The background is dark, making the colorful network stand out.

Forschungsdaten- management

Eine Handreichung

Arbeitsgruppe
Forschungsdaten der Schwerpunktinitiative
„Digitale Information“ der Allianz der
deutschen Wissenschaftsorganisationen

Diese Handreichung soll als Einstieg für Wissenschaftlerinnen und Wissenschaftler, die mit digitalen Daten arbeiten, sowie für alle an dieser Thematik Interessierten dienen und bietet darüber hinaus Hinweise zu weiterführender Information. Sie wurde von der Arbeitsgruppe „Forschungsdaten“ der Schwerpunktinitiative „Digitale Information“ der Allianz der deutschen Wissenschaftsorganisationen¹ verfasst.

¹ <http://www.allianzinitiative.de/>



Definition und Lebenszyklus von Forschungsdaten

Die Schwerpunktinitiative „Digitale Information“ der Allianz der deutschen Wissenschaftsorganisationen definiert Forschungsdaten als „... Daten, die im Zuge wissenschaftlicher Vorhaben z.B. durch Digitalisierung, Quellenforschungen, Experimente, Messungen, Erhebungen oder Befragungen entstehen.“²

Forschungsdatenmanagement umfasst den Prozess, diese Daten im Verlauf ihres gesamten Lebenszyklus, von der Planung über die Erzeugung, Auswahl, Auswertung und über die Speicherung bis hin zur Aufbereitung für die Nachnutzung, zu verwalten. Aufgrund der Heterogenität von Daten ist es notwendig, disziplinspezifische, gegebenenfalls auch datenspezifische Lösungen zu entwickeln.

Nicht alle Aufgaben, die in diesem Prozess anfallen, müssen zwingend von den Forschenden selber durchgeführt werden. Durch die Standardisierung von Einzelprozessen im Datenzyklus ist es möglich, Wissenschaftlerinnen und Wissenschaftler in ihrer Projektarbeit zu entlasten. So können z.B. personelle und fachliche Ressourcen für bestimmte Aufgaben wie Dokumentation, Speicherung oder Auswertung der Daten bereitgestellt werden, die solche Dienstleistungen zentral für alle wissenschaftlich Tätigen übernehmen.

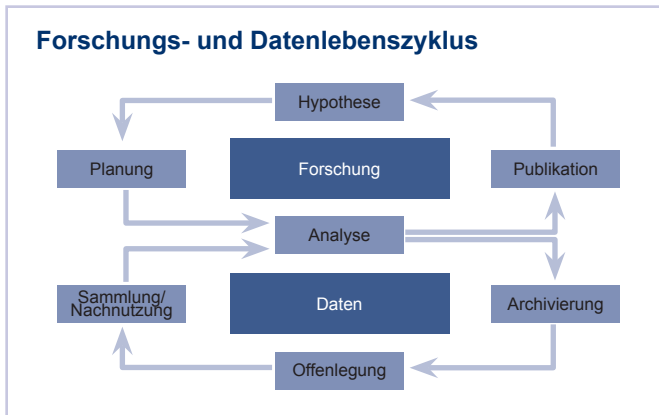


Abb. 1: Forschungs- und Datenzyklus in der Wissenschaft

² <http://www.allianzinitiative.de/de/>

Gründe für Forschungsdatenmanagement

Das Management von Forschungsdaten ist derzeit in vielen Wissenschaftsdisziplinen (noch) nicht verpflichtend. Dieser Prozess umfasst - wie in Abb.1 zu sehen - eine Reihe von Maßnahmen, von der Planung der Datenerhebung über die Datenverwaltung und Datenpublikation bis hin zu Löschkonzepten. Diese erfordern gut ausgebildetes Personal und kosten sowohl Zeit als auch Geld. Zudem gibt es häufig erhebliche Vorbehalte seitens der Forschenden, ihre Daten zur Nachnutzung freizugeben. Neben der Befürchtung, sich durch diese Transparenz angreifbar zu machen, werden als Gründe gegen die Bereitstellung die entstehenden Kosten (während andere sich kostenfrei „bedienen“ können), die fehlende Kontrolle darüber, was mit den Daten in der Nachnutzung geschieht und die Unkenntnis, wo die Daten bereitgestellt werden können, genannt. Darüber hinaus bestehen oft Unklarheiten in Bezug auf urheberrechtliche Fragestellungen.

Diesen Bedenken steht eine Reihe von Vorteilen gegenüber:
Um die Regeln zur guten wissenschaftlichen Praxis³ zu erfüllen, müssen

schon jetzt die in Projekten generierten Daten über das Projektende hinaus so archiviert werden, dass ihre Aussagekraft erhalten bleibt. Sie sind ein wichtiger Bestandteil der wissenschaftlichen Wertschöpfungskette. Mit Metadaten versehene, langzeitarchivierte, qualitätsgesicherte Daten, zu denen ein ungehinderter Zugang besteht, haben mittel- und langfristig großen Nutzen für die Wissenschaft. So müssen Untersuchungen und Experimente nicht zeit- und kostenintensiv wiederholt werden. Durch die Offenlegung von Daten wird die Transparenz der Forschung erhöht, ist eine verlässliche Qualitätskontrolle möglich und nicht zuletzt kann so eine Absicherung der eigenen Ergebnisse gewährleistet werden. Perspektivisch ermöglicht Forschungsdatenmanagement die Anwendung zukünftiger Methoden auf bestehende Daten und bietet so Investitionssicherheit bei aktuellen Projekten.

Daten gelten als „Forschungsprodukte“, die zunehmend zu Evaluierungszwecken herangezogen werden. Sie können, entsprechend gespeichert und präsentiert, wie Publikationen zitiert werden

³ http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_1310.pdf

und erhöhen damit die wissenschaftliche Sichtbarkeit und Reputation sowohl des einzelnen Forschenden als auch die seiner Forschungseinrichtung. Es ist nachgewiesen, dass Publikationen, die mit den dazugehörigen Forschungsdaten veröffentlicht werden, häufiger zitiert werden als Arbeiten ohne diesen Beleg.⁴ Nicht zuletzt kann ein effektives Forschungsdatenmanagement dazu beitragen, dass Forschungsergebnisse zeitnah auf anerkannten Plattformen, wie z.B. Repositorien, bereitgestellt und somit schneller in der Praxis angewendet werden können (Translation).

Forschungseinrichtungen können sich mit einem effizienten, strukturierten Angebot von qualitätsgesicherten Forschungsdaten als aktiver Player im nationalen und internationalen wissenschaftlichen Umfeld positionieren und tragen auf diese Weise wesentlich zur Vernetzung von Institutionen und Forschenden bei.

Viele Stakeholder, wie beispielsweise Wissenschaftseinrichtungen und Forschungsförderer, haben mittlerweile verbindliche Anforderungen in Bezug auf Forschungsdaten etabliert. In der Regel finden sich Aussagen dazu in der Datenrichtlinie einer Institution oder in denen der Förderorganisationen. Auch viele Wissenschaftsverlage fordern mittlerweile die Bereitstellung der zur Publikation gehörenden Forschungsdaten. Dies wird in den Autoren-Verlags-Verträgen geregelt. Drittmittelgeber, wie die Deutsche Forschungsgemeinschaft (DFG), das Bundesministerium für Bildung und Forschung (BMBF) oder die Europäische Union in ihrem Rahmenprogramm für Forschung und Innovation - Horizont 2020, bieten die Möglichkeit, explizit Gelder für das Forschungsdatenmanagement innerhalb eines geförderten Projektes zu beantragen. Näheres hierzu regeln die entsprechenden Calls und Merkblätter, in denen in der Regel unter anderem auch Datenmanagementpläne gefordert werden.

⁴ Piwowar HA, Day RS, Fridsma DB (2007) Sharing Detailed Research Data Is Associated with Increased Citation Rate. PLoS ONE 2(3): e308. doi:10.1371/journal.pone.0000308

Datenmanagementpläne

Ein Datenmanagementplan (DMP) ist ein Instrument zur strukturierten Erfassung und Verarbeitung aller relevanten Informationen über die in einem Forschungsprojekt verwendeten Daten. Durch ihn können bereits in einem sehr frühen Stadium eventuelle Probleme identifiziert werden, wie etwa der Bedarf an geeigneten Einverständniserklärungen bei Befragungen oder erhöhte Hardwareanforderungen bei Computersimulationen. Ebenso ist er für die verschiedenen Projektbeteiligten hilfreich, um sich einen gemeinsamen Überblick über die Arbeit mit den Daten zu verschaffen. Der DMP wird sinnvollerweise zur Projektplanung angefertigt und sollte im Verlauf des Projekts an sich ändernde Bedingungen angepasst werden. Immer häufiger verlangen auch Förderorganisationen von den Zuwendungsempfängern, derartige DMPs zu erstellen, um das Management der Projektdaten zu dokumentieren. Ein valider DMP kann in diesem Kontext dazu beitragen, zusätzliche Kosten für das Datenmanagement plausibel zu machen und einwerben zu können.

In der Regel sollte ein DMP folgende Fragen beantworten:

- Welche Arten von Daten werden wie erhoben, in welchen Mengen und Formaten?
- Wie und wo werden die Daten gesichert und welche zusätzlichen Informationen (Metadaten) sind notwendig?
- Welche rechtlichen, ethischen oder fachlichen Regelungen sind bei der Erhebung, Speicherung oder einer etwaigen Weitergabe zu beachten?
- Sollen Daten nach Projektabschluss veröffentlicht werden und wenn ja, wo und zu welchen Konditionen?

Die Erstellung eines DMP kann mittels einer darauf spezialisierten Software oder entsprechender Internetangebote unterstützt werden. So empfiehlt etwa die Europäische Kommission das Werkzeug des britischen Digital Curation Center.⁵ An deutsche Wissenschafts- und Förderstrukturen angepasste Tools sind derzeit in der Entwicklung (z.B. RDMO⁶). Viele wissenschaftliche Einrichtungen haben auch bereits zentrale Stellen eingerichtet, welche den Forschenden Hilfestellung bei der Realisierung des Datenmanagements und dem Erstellen von DMPs bieten.

⁵ <https://dmponline.dcc.ac.uk/>

⁶ <http://rdmorganiser.github.io/>

Speicherung von Forschungsdaten

Für die Speicherung von Forschungsdaten stehen verschiedene Möglichkeiten zur Verfügung. Sie können in Datenarchiven oder Daten-Repositories gespeichert werden. Als erstes sollte geklärt werden, ob die eigene Einrichtung eine solche Plattform betreibt (institutionelles Datenarchiv oder -repository) oder mit einem Datenzentrum zusammenarbeitet, in dem die Daten gegebenenfalls verpflichtend gespeichert werden müssen. Ist dies nicht der Fall, ist es, insbesondere unter dem Aspekt der Nachnutzung, sinnvoll zu überprüfen, ob für die jeweilige Fachdisziplin ein oder sogar mehrere fachspezifische Repositorien existieren, auf denen die Daten abgelegt werden könnten. Eine gute Recherchemöglichkeit nach fachspezifischen Repositorien bietet Re3data (Registry of Research Data Repositories⁷), welches mittlerweile über 1500 Repositorien aus allen Fachgebieten auflistet. Es werden nicht nur frei zugängliche, sondern auch Repositorien mit eingeschränkten oder ohne Zugriffsrechte auf die Daten aufgeführt.

Ein Repository für Forschungsdaten aus allen Fachgebieten ist Zenodo⁸. Zenodo wurde im Rahmen des OpenAire Projektes⁹ der Europäischen Kommission vom CERN entwickelt und bietet allen Forschenden die Möglichkeit, ihre Daten langfristig zu speichern und zur Verfügung zu stellen. Ein Repository speziell für die Daten, die einer Publikation zugrunde liegen und welche in Form von sogenannten „Supplementary Material“ mittlerweile von fast allen wissenschaftlichen Verlagen bei der Einreichung von Artikeln gefordert werden, ist Dryad¹⁰.

Da es sich bei Forschungsdaten in der Regel um Daten mit einem großen Wert handelt, sollte die Auswahl des Datenrepositorys sehr sorgfältig und unter Berücksichtigung der jeweiligen spezifischen Interessen erfolgen. Die Eignung kommerzieller Cloud-Angebote zur Datenspeicherung ist im Einzelfall differenziert zu betrachten. Kriterien, die für die Auswahl je nach individueller Anforderung herangezogen werden können, sind beispielsweise:

⁷ <http://www.re3data.org/>

⁸ <https://zenodo.org>

⁹ <https://www.openaire.eu/>

¹⁰ <https://datadryad.org/>

- Von wem wird das Repositorium betrieben? Insbesondere unter dem Aspekt der langfristigen Verfügbarkeit ist gegebenenfalls eine große öffentlich geförderte Einrichtung sicherer als ein kommerzieller Anbieter.
- Wird eine geeignete Auswahl von verschiedenen Lizenzmodellen, die die Nachnutzung der Daten festlegen, angeboten?
- Wird ein kollaboratives Arbeiten mit den Daten ermöglicht?
- Fallen Gebühren für die Speicherung an?
- Erfolgt eine fachgerechte Erfassung und Aufbereitung der Daten? Ist beispielsweise eine gute Recherchemöglichkeiten und gute Zitierbarkeit der Daten gewährleistet, etwa durch DOI-Vergabe¹¹?

¹¹s. Kapitel „Datenzitation“, S. 11

Offene Daten

Die Berliner Erklärung zu Open Access¹², die bereits 2003 von allen führenden Wissenschaftsorganisationen unterzeichnet wurde, umfasst nicht nur den offenen Zugang zu Publikationen, sondern schließt auch die Forschungsdaten mit ein. Der Wert, den Forschungsdaten für den wissenschaftlichen Forschungsprozess insgesamt haben, führt dazu, dass immer mehr Förderorganisationen den offenen Zugang zu den Daten, die in von ihnen geförderten Projekten entstanden sind, fordern. So hat beispielsweise die Europäische Kommission in „Horizon 2020“ - dem Rahmenprogramm für Forschung und Innovation - einen „Open Research Data Pilot“¹³ integriert. Hier wird von den Zuwendungsempfängern gefordert, dass die Daten, die in den geförderten Projekten entstanden sind, frei zugänglich und nachnutzbar sein sollen. Für die Forschenden bietet der offene Zugang zu ihren Daten den Vorteil, dass sie von anderen nachgenutzt werden können und sie dadurch zitiert werden. Es gibt allerdings auch gute Gründe, die gegen eine Offenlegung von For-

schungsdaten sprechen. Es ist offensichtlich, dass nicht alle Daten generell frei zugänglich gemacht werden dürfen. Unter Umständen müssen Vorgaben von Projektpartnern berücksichtigt oder rechtliche Rahmenbedingungen eingehalten werden, beispielweise datenschutzrechtliche Bestimmungen bei personenbezogenen Daten. Aus diesem Grund muss vor der Speicherung der Daten in einem Repositorium sorgfältig abgewogen werden, ob der Zugang dazu offen, nur eingeschränkt oder gar nicht möglich sein darf. Auch der Open Research Data Pilot bietet in solchen Fällen eine Opt-out-Möglichkeit. Dies bedeutet, dass in begründeten Ausnahmefällen die Forschungsdaten entgegen den offiziellen Vorgaben nicht frei zugänglich zur Verfügung gestellt werden müssen.

Eine gute Möglichkeit festzulegen, wie die eigenen Forschungsdaten nachgenutzt werden können, bieten die Creative Commons Lizenzen¹⁴. Mit ihnen kann der Datenproduzent festlegen, unter welchen Bedingungen seine Daten weiter verwendet werden dürfen.

¹² <https://openaccess.mpg.de/Berliner-Erklaerung>

¹³ http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

¹⁴ <http://de.creativecommons.org/>



Datenzitation

Gemäß den Empfehlungen zur Sicherung guter wissenschaftlicher Praxis sind Urheber von Informationen, die man selbst nicht generiert hat sondern aus anderen Zusammenhängen in den Kontext der eigenen Arbeit rückt, zu zitieren. Über die Informationen aus wissenschaftlichen Publikationen hinaus werden heute auch Forschungsdaten als zitierwürdig eingestuft. Durch die Zitation wird sichergestellt, dass die Leistung des Datenproduzenten anerkannt wird. Die Zitationen von Forschungsdaten werden in einigen Datenbanken und bibliometrischen Analysetools schon heute gezählt und sichtbar gemacht (z.B. im Data Citation Index oder in Tools der alternativen Metriken wie Altmetric oder

Impact Story). Um eine verlässliche und nachhaltige Zitierung zu gewährleisten, wird empfohlen, die Daten unter anderem mit einem persistenten Identifikator zu versehen. Für Forschungsdaten wird dazu häufig ein DOI (Digital Object Identifier) verwendet. In einigen Repositorien wird er beim Hochladen des Datensatzes automatisiert generiert oder er kann beispielsweise auch von der den Datensatz publizierenden Einrichtung vergeben werden. Im Zweifelsfall sollte man in der Bibliothek der eigenen Wissenschaftseinrichtung nachfragen, ob DOIs über die Bibliothek oder eine andere Stelle im Haus bezogen werden können.

Datenjournale

Wer seine hochwertigen Forschungsdaten aufwändig mit erklärenden Metadaten versieht und veröffentlicht, möchte dafür auch die entsprechende Anerkennung in der Community erhalten. Eine Möglichkeit, dies neben der bereits erwähnten Offenlegung auf Datenrepositorien zu erreichen, wird immer häufiger genutzt: die Publikation von Datenartikeln (Data Paper) in gemischten oder reinen Datenzeitschriften (Data Journals). Ein Datenartikel enthält keine wissenschaftlichen Erkenntnisse, sondern ist eine reine Beschreibung der Daten, der dazu gehörenden Methoden, wie sie generiert wurden sowie ggf. der Art und Weise der Bearbeitung der Daten. Dazu verweist der Artikel natürlich auf die Daten selbst, welche üblicherweise in einem externen Datenrepositorium abgelegt sind (siehe Speicherung von

Forschungsdaten). Der Datenartikel kann aus den Forschungspublikationen heraus referenziert werden und dient als semantische Verbindung zwischen Publikation und Datensatz. Er kann von Forschenden, welche die Daten nachnutzen möchten, zitiert werden und bedient damit den eingespielten Reputationsmechanismus. Wie auch von den herkömmlichen Publikationen bekannt, durchlaufen Datenartikel in der Regel ebenfalls ein Peer-Review-Verfahren. Datenzeitschriften sind überwiegend im Open Access verfügbar, d.h. es fallen keine Subskriptionskosten an, gegebenenfalls werden aber Article Processing Charges (APC) für die Veröffentlichung verlangt.



Recherche von Forschungsdaten

Je umfangreicher und vielfältiger die Menge an qualitativ hochwertigen und verfügbaren Forschungsdaten wird, desto attraktiver ist es, diese Daten für die eigene Forschung nachzunutzen. Während in einigen Fachdisziplinen, etwa der Genforschung, die Daten in wenigen, zentralen Archiven abgelegt werden¹⁵, finden sich die Daten in anderen Fächern weltweit in unterschiedlichen Repositorien verstreut. Plattformen, welche Informationen über diese Daten zusammentragen, sind bisher selten. Zu erwähnen sein hier der kommerzielle Data Citation Index als Modul des Web of Science und die Metadaten-Su-

che von DataCite¹⁶, dem Dachverband der Organisationen, welche Digital Object Identifier (DOIs) für Forschungsdaten vergeben. Nationale Initiativen, Forschungsdaten über einheitliche Angebote zugänglich zu machen, gibt es bereits in Australien¹⁷ und Großbritannien¹⁸. Auf europäischer Ebene fassen CESSDA¹⁹ sozialwissenschaftliche und Europeana²⁰ geisteswissenschaftliche Forschungsdaten zusammen. Manche Forschungseinrichtungen bieten auch hier bereits zentrale Dienste an, die bei der Suche nach speziellen Daten unterstützen.

¹⁵ <http://www.ebi.ac.uk/>

¹⁶ <http://search.datacite.org/>

¹⁷ <https://researchdata.andis.org.au/>

¹⁸ <http://ckan.data.alpha.jisc.ac.uk/>

¹⁹ <http://cessda.net/>

²⁰ <http://www.europeana.eu/>

Impressum

Die Online-Version dieser Publikation finden Sie unter:

<http://doi.org/10.2312/allianzoa.029>

Herausgeber

Arbeitsgruppe Forschungsdaten der Schwerpunktinitiative Digitale Information der Allianz der deutschen Wissenschaftsorganisationen

Redaktion

Michael Franke (Max Planck Digital Library)

Henriette Senst (Robert Koch-Institut)

Dagmar Sitek (Deutsches Krebsforschungszentrum)

Kontakt

Dr. Stefan Winkler-Nees (Deutsche Forschungsgemeinschaft)

Stefan.Winkler-Nees@dfg.de

Stand

1. Auflage, Februar 2018

Lizenz

Alle Texte dieser Veröffentlichung, ausgenommen Zitate und Umschlagfoto (© Vege / Fotolia), sind unter einem Creative Commons Attribution 4.0 International (CC BY 4.0) Lizenzvertrag lizenziert.

<http://creativecommons.org/licenses/by/4.0/>



DFG Deutsche
Forschungsgemeinschaft

 **Fraunhofer**

HELMHOLTZ SPITZENFORSCHUNG FÜR
GROSSE HERAUSFORDERUNGEN

HRK Hochschulrektorenkonferenz
Die Stimme der Hochschulen


Leibniz-Gemeinschaft

 **Leopoldina**
Nationale Akademie
der Wissenschaften

DAAD Deutscher Akademischer Austauschdienst
German Academic Exchange Service



MAX-PLANCK-GESELLSCHAFT

WR | WISSENSCHAFTSRAT