



Grundlagen zum Forschungsdatenmanagement im DLR

Zweck, Ziel und Geltungsbereich

Der Umgang mit Forschungsdaten hat sich aufgrund der Digitalisierung und der extrem anwachsenden Datenfülle grundlegend geändert. Hinzu kommen zunehmende Forderungen von nationalen und internationalen Behörden und Geldgebern, gewonnene Forschungsdaten langfristig aufzubewahren und der Öffentlichkeit zur Verfügung zu stellen.¹ Die Reproduzierbarkeit von wissenschaftlichen Experimenten und Datenerhebungen einerseits sowie die Möglichkeiten zur Nachnutzung (z. B. auch in anderen Fachdisziplinen) für neue Fragestellungen andererseits stellen den verantwortungsvollen, organisatorischen und technischen Umgang mit Daten an den Anfang der infrastrukturellen Aufgabenkette. Der sorgfältige Umgang mit Forschungsdaten ist außerdem Teil des Qualitätsverständnisses der Forschung am DLR.

Die Helmholtz-Gemeinschaft hat auf ihrer Mitgliederversammlung im September 2016 das Dokument „Die Ressource Information besser nutzbar machen“ Positionspapier zum Umgang mit Forschungsdaten in der Helmholtz-Gemeinschaft² verabschiedet. Darin appelliert die Mitgliederversammlung an die Zentren entsprechende Richtlinien für den Umgang mit Forschungsdaten aufzustellen. Ergänzend zu diesem Positionspapier wurden „Empfehlungen für Richtlinien der Helmholtz-Zentren zum Umgang mit Forschungsdaten“ verabschiedet³. Die DLR-Grundlagen des Forschungsdatenmanagements basieren auf diesen Empfehlungen und sind für die speziellen Anforderungen des DLR angepasst.

Die Allianz der Wissenschaftsorganisationen hat sich bereits mit den im Jahr 2010 verabschiedeten „Grundsätze[n] zum Umgang mit Forschungsdaten“ für die langfristige Sicherung von, den grundsätzlich offenen Zugang zu und die Berücksichtigung fachdisziplinärer Regularien im Umgang mit Forschungsdaten ausgesprochen⁴. Empfehlungen hierzu gibt es auch von der Deutschen Forschungsgemeinschaft⁵.

Die DLR-Grundlagen bilden den Rahmen für den Umgang mit Forschungsdaten im DLR. Sie regeln die Verantwortlichkeiten, den Prozess der Datenerhebung, der Datenarchivierung und Nachnutzung. Die Grundlagen gelten nicht rückwirkend. Die Grundlagen sind öffentlich zugänglich.

Die Grundlagen richten sich an alle Institute und Einrichtungen des DLR und sind verbindlich einzuhalten. Basierend auf diesen Grundlagen können je nach Bedarf oder Vorgaben zum Beispiel

¹ [EU Open Data Policy](https://ec.europa.eu/digital-single-market/en/policies/open-access)
<https://ec.europa.eu/digital-single-market/en/policies/open-access>
DFG Richtlinie
http://www.dfg.de/foerderung/programme/infrastruktur/lis/open_access/weitere_informationen/index.html

² [Helmholtz Positionspapier](https://www.helmholtz.de/fileadmin/user_upload/01_forschung/Open_Access/DE_AKOS_TG-Forschungsdatenleitlinie_Positionspapier.pdf)
https://www.helmholtz.de/fileadmin/user_upload/01_forschung/Open_Access/DE_AKOS_TG-Forschungsdatenleitlinie_Positionspapier.pdf

³ [Empfehlungen der Helmholtz für die Zentren](http://portal.dlr.de/Organisation/Organisationseinheiten/VorstandundStaebel/W/openaccess/Document%20Library/HGF-Forschungsdatenleitlinien-Empfehlungen.pdf)
<http://portal.dlr.de/Organisation/Organisationseinheiten/VorstandundStaebel/W/openaccess/Document%20Library/HGF-Forschungsdatenleitlinien-Empfehlungen.pdf>

⁴ [Grundsätze zum Umgang mit Forschungsdaten:](http://www.allianzinitiative.de/de/handlungsfelder/forschungsdaten/grundsaeetze/)
<http://www.allianzinitiative.de/de/handlungsfelder/forschungsdaten/grundsaeetze/>

⁵ [DFG Umgang mit Forschungsdaten](http://www.dfg.de/download/pdf/dfg_magazin/wissenschaftliche_karriere/heisenberg_treffen_2010/forschungsdaten.pdf)
http://www.dfg.de/download/pdf/dfg_magazin/wissenschaftliche_karriere/heisenberg_treffen_2010/forschungsdaten.pdf

für einzelne Forschungsbereiche, Institute/Einrichtungen, Forschungsvorhaben, Forschungsprojekte oder Forschungsinfrastrukturen detaillierte Richtlinien aufgestellt werden.

Definitionen

Forschungsdaten

Als Forschungsdaten werden hier Daten bezeichnet, die während des Forschungsprozesses entstehen. Die detaillierte Definition von Forschungsdaten liegt in der Verantwortung des Institutes / der Einrichtung.

Lebenszyklus von Forschungsdaten

Daten werden im Hinblick auf vorliegende Forschungsfragen erhoben und können anschließend maschinell verarbeitet werden, bevor sie auf ein spezifisches Forschungsinteresse analysiert und interpretiert werden. Danach werden die Daten archiviert und je nach Voraussetzungen und Möglichkeiten Dritten Zugang zu den Daten eingeräumt. So können bei Bedarf Daten entsprechend nachgenutzt und in die Erstellung neuer Forschungsvorhaben einbezogen werden. Die vereinfachte Form eines Lebenszyklus von Forschungsdaten lässt sich in folgende Phasen unterteilen⁶:

- Planung und Erstellung;
- Auswahl und Bewertung;
- Übernahme in ein Archiv;
- Sichere Speicherung;
- Erhaltungsmaßnahmen;
- Zugriff und Nutzung.

Forschungsdatenmanagement

Der Begriff Forschungsdatenmanagement bezeichnet den Umgang mit Forschungsdaten während des gesamten Lebenszyklus unter Beachtung der Regeln zur Sicherung der guten wissenschaftlichen Praxis. Darüber hinaus beinhaltet der Begriff die fachspezifische Dokumentation der Prozesse im Kontext ihrer Erhebung. Datenmanagementpläne erleichtern die Dokumentation dieser Prozesse und die Beschreibung der Daten.

Datenmanagementplan

Ein Datenmanagementplan ist ein Dokument zur Beschreibung des Lebenszyklus der Forschungsdaten. Förderorganisationen verlangen inzwischen Aussagen dazu, wie mit den erstellten Daten im beantragten Projekt verfahren wird, d.h. sie fordern einen Datenmanagementplan ein.

Nachnutzung von Forschungsdaten

In der Forschungspraxis stehen Forschungsdaten häufig noch nicht zur Nachnutzung zur Verfügung. Forschungsdaten sollen/müssen auch nach ihrer Erstverwertung für eine Nachnutzung langfristig⁷ in

⁶ [Lebenszyklus von Daten https://data.uni-bielefeld.de/de/node/342](https://data.uni-bielefeld.de/de/node/342)

⁷ Wie lange die Forschungsdaten gespeichert werden müssen ist fachspezifisch und wurde bewusst nicht vorgegeben.

geeigneten, vertrauenswürdigen Forschungsdaten-Infrastrukturen nachhaltig abgelegt und gesichert sowie ggf. veröffentlicht werden. Für die Nachnutzung von Forschungsdaten ist es notwendig, Metadaten zu erfassen sowie den Entstehungskontext und die benutzten Werkzeuge bzw. Software zu dokumentieren. In diesem Zusammenhang ist auch der langfristige Zugang zu und die Nachnutzung von wissenschaftlicher Software essentiell.

Metadaten

Metadaten oder *Metainformationen* sind Daten, die Informationen über Merkmale anderer Daten enthalten, um die Zugänglichkeit und Nutzungsmöglichkeiten von Daten zu dokumentieren. Ein Standard ist „Dublin Core“⁸. Welche Metadaten verwendet werden hängt aber auch von der Fachdisziplin ab. In einigen Disziplinen wurden bereits spezifische Standards festgelegt, wie z. B. bei Fernerkundungsdaten⁹.

Verantwortung und Festlegung

Das Management von Forschungsdaten umfasst den gesamten Lebenszyklus. Im Rahmen von Forschungskollaborationen gelten diese Grundsätze, soweit die anderen Beteiligten keine strengeren Vorgaben treffen. Die Verantwortung für das Management von Forschungsdaten, insbesondere die in Projekten und Vorhaben entstehen, ist zu Beginn vom verantwortlichen Institutsleiter/Einrichtungsleiter festzulegen und gilt für den gesamten Lebenszyklus.

Einhaltung von rechtlichen Rahmenbedingungen

Gesetzliche Vorgaben und etwaige fachbezogene Grundsätze sind einzuhalten. Dabei sind insbesondere das Datenschutz- und Urheberrecht, der Geheimschutz, die Vorgaben der Exportkontrolle und vertragliche Vorgaben zu beachten. Besonders schützenswert sind personenbezogene und personenbeziehbare Daten.

Informationen und Hilfestellungen geben der Datenschutzbeauftragte des DLR (in allen Fragen rund um den Datenschutz) und die DLR-Rechtsabteilung (zu allen weiteren rechtlichen Fragen).

Datenmanagementplan

Es wird empfohlen Datenmanagementpläne (DMP) für alle Aktivitäten zu erstellen, bei denen Forschungsdaten entstehen. Gibt es bei Drittmittelprojekten Vorgaben zu Datenmanagementplänen (DMP), so sind diese vorrangig einzuhalten, sollten aber den Grundlagen des Forschungsdatenmanagements entsprechen. Fragen der Datenart und -menge, Informationen zur Art der Datenerzeugung etc. und ob und welche vorher schon bestehenden Daten in welcher Qualität eingebracht worden, sind ebenso wie etwaige Eigentumsverhältnisse der Daten und Urheberrechte zu klären und Bestandteil des DMP.

Hilfestellung beim Aufstellen eines DMP findet man unter „Anleitung zum Erstellen eines DMP“¹⁰.

⁸ [Dublin Core Standard https://data.uni-bielefeld.de/de/node/344](https://data.uni-bielefeld.de/de/node/344)

⁹ Auf europäischer Ebene existiert das Geographic Data Description Directory (GDDD) als Umsetzung des CEN „Metadata Standard on Geographic Information“ (MDS) <https://www.fgdc.gov/metadata>

¹⁰ [Anleitung zum Erstellen eines DMP](#)

Daten-Repository (edata.dlr.de)¹¹

Analog zu elib für die Publikationen sind Informationen zu den Forschungsdaten im DLR-Daten-Repository, *edata.dlr.de* zu hinterlegen. Der Eintrag in das Repository ist für alle Institute und Einrichtungen des DLR verpflichtend. Informationen zu allen Forschungsdaten sind einzutragen, wobei in Abhängigkeit der Datenerhebung, der Datenvielfalt und Datenstruktur auch eine Aggregation möglich ist.

Verwendet man zur Erstellung des DMPs ein geeignetes Tool (wie z.B. das von der DFG angebotene Tool RDMO)¹², kann edata teilweise automatisch generiert werden.

Qualitätsmanagement, Metadaten

Die Grundsätze guter wissenschaftlicher Praxis sind auch beim Datenmanagement einzuhalten¹³. Dies beinhaltet ein sichergestelltes Qualitätsmanagement, das den gesamten Lebenszyklus der Daten abdecken soll. Neben der Beschreibung der reinen Prozedur der Datenerhebung und der Datenerfassung sind auch Metadaten zu erfassen. Zum Teil existieren fachbereichsspezifische Standards für Metadaten¹⁴, die dann verwendet werden müssen. Sind keine Standards vorhanden, so sollten die Metadaten neben Angaben über Datenformate auch Angaben über die Qualität der Daten enthalten und maschinen-lesbar sein.

Datensicherung, Langzeitarchivierung

Forschungsdaten inklusive der Metadaten sind in geeigneten, vertrauenswürdigen Forschungsdaten-Infrastrukturen zu speichern. Forschungsdaten sind entsprechend der „Vorschläge zur Sicherung guter wissenschaftlicher Praxis“ der DFG nach Abschluss des Vorhabens zu sichern.

Offener Zugang

Forschungsdaten sollen auch für eine Nachnutzung nachhaltig abgelegt, gesichert und im Sinne von Open-Access, wie es in der „Berliner Erklärung über offenen Zugang zu wissenschaftlichem Wissen“ von 2003 niedergelegt ist, öffentlich zugänglich gemacht werden.

Dabei sind gesetzliche Regelungen, insbesondere der Schutz personenbezogener Daten, ebenso wie die wissenschaftlichen Interessen und vertraglichen Vereinbarungen mit Kooperationspartnern sowie gegebenenfalls Verwertungsinteressen zu berücksichtigen.

Es ist möglich eine Embargofrist¹⁵ zur exklusiven Erstnutzung oder kommerziellen Verwertungsmöglichkeit festzulegen. Die Embargofrist ist vom Instituts- / Einrichtungsleiter in Abstimmung mit dem DLR-Forschungsdatenmanager festzulegen.

In zu begründenden Ausnahmefällen kann der Vorstand im Interesse des DLR auch aus anderen Gründen eine Embargofrist festlegen.

<https://www.forschungsdaten.info/themen/planen-und-strukturieren/datenmanagementplan/>

¹¹ „edata“ ist der Name des Repositoriums zum 1.6.2018; sollte er zu einem späteren Zeitpunkt geändert werden, wird das an Inhalt und Funktion des Repositoriums nichts ändern.

¹² <https://rdmorganiser.github.io/>

¹³ Gute Wissenschaftliche Praxis http://www.dfg.de/foerderung/grundlagen_rahmenbedingungen/gwp/

¹⁴ Metadaten Standards <https://data.uni-bielefeld.de/de/node/344>

¹⁵ Embargofrist = Zeitliche Veröffentlichungssperre

Finanzierung

Das Forschungsdatenmanagement inklusive Langzeitarchivierung ist mit Kosten und Folgekosten verbunden. Die Kosten fallen während des gesamten Lebenszyklus der Daten an. Bei der Beantragung eines Projektes und Vorhabens sind Kosten für Datenmanagement und Archivierung zu berücksichtigen und entsprechende Verabredungen zu treffen.

Inkrafttreten

Die Grundlage für das Forschungsdatenmanagement tritt am 01.07.2018 durch VO-Beschluss vom 25.06.2018 in Kraft.