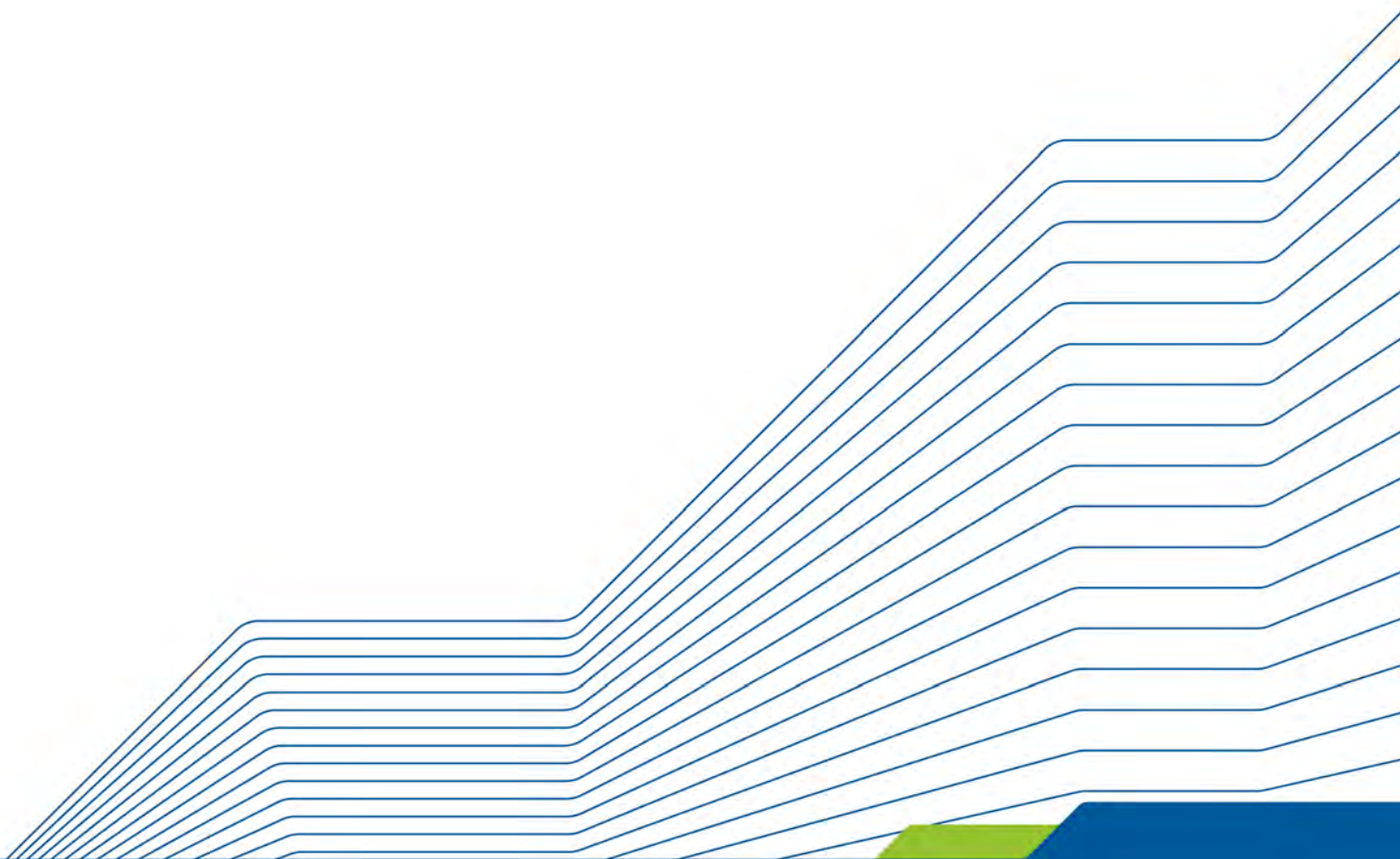


HELMHOLTZ

Open Science

Empfehlungen für Richtlinien der Helmholtz-Zentren zum Umgang mit Forschungsdaten

*In der 109. Mitgliederversammlung der Helmholtz-Gemeinschaft
am 13.-14. September 2017 beschlossen.*



Impressum

Die Onlineversion dieser Publikation finden Sie unter:

DOI: <https://doi.org/10.2312/os.helmholtz.002>

Herausgeber

Arbeitskreis Open Science der Helmholtz-Gemeinschaft

Redaktion

Helmholtz Open Science Koordinationsbüro

Kontakt

Helmholtz Open Science Koordinationsbüro

c/o Helmholtz-Zentrum Potsdam

Deutsches GeoForschungsZentrum GFZ

Telegrafenberg, 14473 Potsdam

E-Mail: open-science@helmholtz.de

Stand

Februar 2019

Lizenz

Alle Texte dieser Veröffentlichung, ausgenommen Zitate, sind unter einem Creative Commons Attribution 4.0 International“ (CC BY 4.0) Lizenzvertrag lizenziert. Siehe: <https://creativecommons.org/licenses/by/4.0>



Inhaltsverzeichnis

Präambel	2
Zielsetzung	2
Forschungsdaten und Forschungsdatenmanagement.....	3
Verantwortlichkeit	4
Offener Zugang	5
Qualität im Rahmen guter wissenschaftlicher Praxis	5
Wissenschaftliche Anerkennung	6
Langzeitverfügbarkeit	7
Forschungsdaten-Infrastrukturen	8
Qualifizierung.....	9
Finanzierung.....	9
Rechtsfragen	9
Anhang: Rechtsfragen.....	11

Präambel

Die vorliegenden Empfehlungen für Richtlinien der Helmholtz-Zentren zum Umgang mit Forschungsdaten wurden am 13./14. September 2017 durch die Mitgliederversammlung der Helmholtz-Gemeinschaft Deutscher Forschungszentren e.V. verabschiedet. Dabei wurde betont, dass Open Access und Technologietransfer für die Helmholtz-Gemeinschaft gleichrangige Ziele darstellen, deren ggf. divergierenden Erfordernisse im Einzelfall gegeneinander abzuwägen sind.

Zielsetzung

Aufgrund des raschen digitalen Wandels und der zunehmenden Datenfülle ändert sich der Umgang mit Forschungsdaten und deren Nutzung grundlegend. Die Helmholtz-Gemeinschaft hat deswegen auf ihrer Mitgliederversammlung im September 2016 das Dokument Die Ressource Information besser nutzbar machen! Positionspapier zum Umgang mit Forschungsdaten in der Helmholtz-Gemeinschaft¹ (im Folgenden Positionspapier) verabschiedet. Darin appelliert die Mitgliederversammlung an die Zentren entsprechende Richtlinien für den Umgang mit Forschungsdaten aufzustellen und beauftragte Herrn Marquardt in seiner Funktion als Betreuer des Arbeitskreises Open Science, in Abstimmung mit dem Arbeitskreis die Operationalisierung des Papiers einzuleiten.

Das vorliegende Dokument richtet sich an die Verantwortlichen der Zentren, die für die Einhaltung der im Positionspapier verabschiedeten Grundsätze zuständig sind.²

Ziel der vorliegenden Empfehlungen ist es daher, die Zentren bei der Formulierung ihrer jeweiligen Richtlinien zu unterstützen. Diese Richtlinien sollen allen Interessierten öffentlich zugänglich gemacht werden. Dabei versteht es sich von selbst, dass die Umsetzung der in diesem Papier vorgestellten Empfehlungen in der Verantwortung der Zentren liegt und an die jeweiligen forschungs- und disziplinspezifischen Gegebenheiten angepasst werden muss. Mit den Empfehlungen wird vorgeschlagen, was in der Zentren-spezifischen Umsetzung geregelt werden sollte, ohne Vorgaben zu machen, wie die einzelnen Aspekte ausgestaltet werden.

Ein wesentliches Ziel der vorliegenden Empfehlungen ist es darüber hinaus, eine gewisse Harmonisierung der Richtlinien zum Umgang mit Forschungsdaten an den Zentren der Helmholtz-Gemeinschaft zu erreichen und dazu beizutragen, einen „Helmholtz-Standard“ entstehen zu lassen. Ein solches Qualitätsniveau soll sich auch vor dem Hintergrund sich bildender fachspezifischer Verfahrensweisen, Standards in den verschiedenen Forschungsfeldern und Publikationsorganen sowie der Anforderungen großer Forschungsförderer entwickeln. In diesem Zusammenhang sind insbesondere die *Guidelines on FAIR Data Management in Horizon 2020*³ der EU-Kommission sowie die

¹ Erstellt vom AK Open Science.

² Siehe auch Kapitel „Verantwortlichkeit“

³ http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

HELMHOLTZ

Open Science

Leitlinien zum Umgang mit Forschungsdaten der DFG⁴ zu nennen, die schon bei der Antragstellung Hinweise bzw. klare Pläne verlangen, wie mit Daten aus diesen öffentlich geförderten Forschungsprojekten umgegangen werden soll und wie sie für eine Nachnutzung aufbereitet werden können.

Durch die Umsetzung der vorgelegten Empfehlungen soll letztlich Wissenschaftlerinnen und Wissenschaftlern die Arbeit erleichtert werden, indem nicht nur ein gemeinsamer „Helmholtz-Standard“ die Einhaltung der jeweiligen Vorgaben der Publikationsorgane und Förderorganisationen garantiert, sondern auch Sorge dafür getragen wird, dass die notwendigen Rahmenbedingungen am Zentrum geschaffen und unterhalten werden.

Die Helmholtz-Gemeinschaft sieht sich unter anderem auch dem gesellschaftlichen Ziel verpflichtet, qualitätsgesicherte Forschungsdaten auch langfristig für die Gesellschaft zu erhalten und nutzbar zu machen. Die Zentren sollen mit ihren Richtlinien die Notwendigkeit eines zusätzlichen Aufwands für das Forschungsdatenmanagement anerkennen und entsprechend dessen Durchführung sowohl strukturell und finanziell als auch durch wissenschaftliche Anerkennung unterstützen.

Forschungsdaten und Forschungsdatenmanagement

Als **Forschungsdaten** werden hier alle Daten bezeichnet, die während des Forschungsprozesses entstehen, dazu herangezogen werden oder dessen Ergebnis sind. Sie werden abhängig von der spezifischen **Forschungsfrage** und unter Anwendung verschiedener Methoden erzeugt bzw. beschafft oder gesammelt, beobachtet, simuliert, abgeleitet, validiert, bearbeitet, weiterverarbeitet, analysiert und schließlich archiviert. Forschungsdaten treten demnach in jeder Wissenschaftsdisziplin und je nach Stadium im Lebenszyklus der Daten in unterschiedlichen Medientypen und Formaten, Aggregations- und Qualitätsstufen auf.

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.⁵

⁴ DFG-Leitlinien zum Umgang mit Forschungsdaten (30. September 2015)

https://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten.pdf

⁵ Die Herausforderungen, Chancen und Rahmenbedingungen, die hieraus erwachsen, werden in der Task Group „Zugang zu und Nachnutzung von wissenschaftlicher Software“ des AK Open Science erarbeitet.

Der Begriff **Forschungsdatenmanagement** bezeichnet den Umgang mit Forschungsdaten von der Planung, deren Generierung und Verarbeitung bis zur Langzeit-Archivierung bzw. Löschung unter Beachtung der Regeln zur Sicherung der guten wissenschaftlichen Praxis. Das Management von Forschungsdaten betrifft demnach den gesamten Lebenszyklus von Forschungsdaten. 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.

Im fachspezifischen Kontext und vor dem Hintergrund der Besonderheiten der jeweiligen Forschungseinrichtung sind geeignete Regeln zu definieren, die Umfang und Detaillierungsgrad der zu erhebenden Daten sowie den konkreten Anwendungsbereich der Richtlinien für das Zentrum festlegen. Dies ist allein schon deshalb notwendig, weil effektives Forschungsdatenmanagement nicht ohne eigens dafür bereitgestellte Ressourcen auskommt.

In Anlehnung an das Verfahren im Programm Horizon2020 wird empfohlen, bereits in Anträgen oder Plänen zu Forschungsvorhaben, Forschungsprojekten oder Forschungsinfrastrukturen (also z.B. Drittmittelanträge, aber auch alle Helmholtz- und Zentren-internen Verfahren bis hin zur Programm-orientierten Förderung und strategischen Investitionen) das Erstellen einer kompakten Datenmanagementstrategie und im weiteren Verlauf die Pflege eines detaillierteren Datenmanagementplans, der den Umgang mit den im Rahmen des Forschungsvorhabens bzw. Projektes entstandenen Forschungsdaten beschreibt, verpflichtend zu machen. Bereits im Antrag über Forschungsvorhaben⁶ sollte auf dazu benötigte Ressourcen und die Urheber-, Nutzungs- und Zugangsrechte sowie die Aufbewahrung während und nach Ende des Forschungsvorhabens eingegangen werden. Den Projekten sollte dabei die Möglichkeit gegeben werden, sich eines Standardplans des Zentrums bzw. einer Organisationseinheit zu bedienen.

Verantwortlichkeit

Die Verantwortung für das Management der Daten aller Forschung muss in den Zentren eindeutig geregelt sein – bis zur Ebene der Forschungsvorhaben, Forschungsprojekte und Forschungsinfrastrukturen. In Abstimmung mit den beteiligten Wissenschaftlerinnen und Wissenschaftlern sowie unter Berücksichtigung der einschlägigen Anforderungen der Forschungsförderer und -partner werden sämtliche Schritte des Forschungsdatenmanagements festgelegt, dokumentiert und entsprechend umgesetzt. Dabei werden ethische, rechtliche und Verwertungsaspekte sowie die Einhaltung der Regeln der guten wissenschaftlichen Praxis berücksichtigt.

⁶ Unter Forschungsvorhaben werden hier auch Projekte und Infrastrukturen verstanden.

Offener Zugang

In Übereinstimmung mit internationalen Organisationen auf dem Gebiet der Förderung und Durchführung von Forschungsaufgaben unterstützt die Helmholtz-Gemeinschaft die langfristige Sicherung und den grundsätzlich offenen Zugang zu Forschungsdaten aus öffentlich geförderter Forschung gemäß den Grundsätzen von Open Access, wie sie in der „Berliner Erklärung über offenen Zugang zu wissenschaftlichem Wissen“ von 2003 niedergelegt sind. Dabei sind gesetzliche Regelungen, insbesondere der Schutz personenbezogener Daten, ebenso wie die wissenschaftlichen Interessen und vertraglichen Vereinbarungen mit Kooperationspartnern sowie gegebenenfalls Verwertungsinteressen zu berücksichtigen. Ferner ist zu berücksichtigen, dass u.U. eine zumindest zeitweilige Vertraulichkeit von Forschungsdaten eine unabdingbare Voraussetzung für eine spätere Kommerzialisierung darstellt. Angemessene Embargofristen zur exklusiven Erstnutzung können nach Maßgaben des Zentrums gesetzt werden und sind dann als Teil der Richtlinien zum Umgang mit Forschungsdaten eines Zentrums auch öffentlich zu machen.

Bei der Bestimmung der Dauer dieser Frist sollen sich die Zentren an entsprechenden Vorgaben von Fachgesellschaften, großen Forschungsverbänden und Forschungsförderern orientieren. Derzeit fehlen für viele Forschungsfelder solche Vorgaben. Gleiches gilt für die Bestimmung des Zeitpunktes, zu dem die Laufzeit der Frist beginnen soll. Die derzeit noch bestehende Unsicherheit soll nicht zu einem Verzicht auf die Fristsetzung führen. Vielmehr sollen sich die Zentren aktiv am Findungsprozess beteiligen und die Angemessenheit getroffener Entscheidungen in regelmäßigen Abständen prüfen.

Die Entscheidung für einen grundsätzlich offenen Umgang mit Forschungsergebnissen schließt deren Kommerzialisierung nicht aus. In Übereinstimmung mit der Mission der Helmholtz-Gemeinschaft im Bereich Technologietransfer⁷ und den entsprechenden Formulierungen in Satzungen der einzelnen Zentren ist eine kommerzielle Nachnutzung der Forschungsergebnisse ebenfalls grundsätzlich zu ermöglichen.

Qualität im Rahmen guter wissenschaftlicher Praxis

Die nachhaltige Nutzbarmachung von Forschungsdaten bedarf eines sichergestellten Qualitätsmanagements. Dieses soll den gesamten Lebenszyklus der Daten abdecken und somit von der Datenerhebung über Workflows und Methoden der Bearbeitung, deren Aufbewahrung und Sicherung, bis zum kontrollierten Löschen der Daten reichen.

Bei der Datenerhebung sind neben der Beschreibung der reinen Prozedur der Datenerfassung auch Rahmenparameter (sogenannte Metadaten) zu erfassen, die Aussagen über die Qualität der erhobenen Daten ermöglichen. Welche Metadaten zu diesem Zwecke erhoben werden müssen hängt im Einzelfall jeweils vom konkreten Forschungsvorhaben/Projekt ab. Informationen zur Qualitäts-

⁷ "Wir tragen bei zur Gestaltung unserer Zukunft durch Verbindung von Forschung und Technologieentwicklung mit innovativen Anwendungs- und Vorsorgeperspektiven."

sicherung können dabei sowohl in der Form von Laborbüchern als beispielsweise auch in der Dokumentation datenerzeugender Prozesse festgehalten werden.

Im Rahmen moderner („digitaler“) Wissenschaft ist jedoch der expliziten und maschinenlesbaren Kodierung dieser Informationen der Vorzug zu geben und diese – in Theorie, Standardisierung und effizienzsteigernder, nicht belastender, Praxis – voranzutreiben.

Spätestens in Zusammenhang mit automatisierter Datenanalyse wird diese Maschinenlesbarkeit notwendig, ebenso wie die explizite *und* maschinenlesbare Angabe von Fehlermaßen (sei es als Teil der Daten oder der Metadaten).

Teil eines umfangreichen Metadatensatzes sollen auch Informationen über die verwendeten Datenformate sein. Nach Möglichkeit sollen offene und freie Datenformate Verwendung finden, da sich für diese auch nach einem Wegfall der ursprünglich eingesetzten Applikation noch Implementierungen realisieren lassen. Für eine Nachnutzung, insbesondere auch mit digitalen Methoden, ist darauf zu achten, dass auch qualitätssichernde Metadaten – nach Möglichkeit in digitalisierter Form und algorithmisch zugänglich – abgelegt werden.

Wissenschaftliche Anerkennung

Die Generierung von Forschungsdaten ist eine zentrale und unverzichtbare Leistung im Forschungsprozess, welche der Wissenschaft und mittelbar der Gesellschaft insgesamt zugutekommt. Qualitätsgesicherte Forschungsdaten zählen somit zum wissenschaftlichen Output der Zentren und ihre Erzeugung ist entsprechend zu würdigen.

Das Generieren von Forschungsdaten wird heute trotzdem vielfach als nachrangig zu deren Analyse betrachtet. Diese Differenzierung ist angesichts der Kompetenzen, die für die Erhebung und Aufbereitung von Daten benötigt werden, oft nicht mehr angemessen. Eine verbesserte Anerkennung der wissenschaftlichen Leistung, die in der Erhebung von Forschungsdaten zum Ausdruck kommt, beginnt deshalb mit der Entwicklung eines neuen Blicks auf alle Arbeitsprozesse und den jeweils daran Beteiligten, die nur in einer kooperativen Anstrengung exzellente Forschungsergebnisse realisieren können.

Ansätze, Leistungen in diesem arbeitsteiligen Prozess durch etablierte Methoden der Leistungsmessung – vor allem das Publizieren von Artikeln in renommierten Zeitschriften – zu erfassen und zu bewerten, können nur ein Zwischenschritt sein, weil sie die Defizite der etablierten Evaluierungsmetriken perpetuieren und vielfach für die Erfassung der erbrachten Leistung ungeeignet sind.⁸

Es ist klar, dass die im Publikationswesen genutzten bibliometrischen Indikatoren in diesem Kontext nicht ausreichen, ebenso wenig wie allgemeine, unspezifische Indikatoren wie z. B. Anzahl oder Volumen. Die Helmholtz-Gemeinschaft setzt sich deshalb zusätzlich das Ziel, geeignete Indikatoren

⁸ Vgl. beispielsweise The Metric Tide, Report of the Independent Review of the Role of Metrics in Research Assessment and Management, July 2015. DOI: 10.13140/RG.2.1.4929.1363.
http://www.hefce.ac.uk/media/HEFCE,2014/Content/Pubs/Independentresearch/2015/The,Metric,Tide/2015_metric_tide.pdf

HELMHOLTZ

Open Science

zu entwickeln, mit denen wissenschaftliche Leistung, die im Rahmen des Managements von Forschungsdaten erbracht wird, sichtbar gemacht werden kann.

Unabhängig von der Frage der Evaluierung gewährleistet der Weg der zitierbaren Datenpublikation nicht nur eine nachvollziehbare wissenschaftliche Anerkennung, sondern auch die Reproduzierbarkeit von darauf aufbauenden Untersuchungen. Zudem wird der offene Zugang zu den Forschungsdaten dokumentiert.

Folgende aktuelle Beispiele (2016) belegen die zunehmende Anerkennung der wissenschaftlichen Bedeutung von Forschungsdaten;

- Die amerikanische National Science Foundation erwartet in ihren Proposal Guidelines seit 2013 nicht mehr eine (möglichst lange) Liste von Artikeln, sondern je 5 „Produkte“ – dies können Artikel, Daten, Software, Patente ... sein – welche die bedeutendsten persönlichen Leistungen bzw. die für den Fortschritt der Community wichtigsten Ergebnisse des Antragstellers wiedergeben.⁹
- Die Entwicklung des Datenjournals „Earth Systems Science Data“ zeigt eindrücklich, wie auch über den klassischen Weg der Publikationen die wissenschaftliche Bedeutung von Daten sichtbar gemacht werden kann.¹⁰
- Zunehmend fordern auch weithin bekannte, etablierte Journale je Artikel ein „Data Availability Statement“.¹¹ Diese Forderung wird sich ohne Zweifel zu einem Standard entwickeln; es wird an Herausgebern und Gutachtern liegen, welche Art „availability“ schließlich als akzeptabel gelten wird.

Langzeitverfügbarkeit

Forschungsdaten sollen zumeist auch nach ihrer Erstverwertung für eine Nachnutzung langfristig in einer geeigneten, vertrauenswürdigen Forschungsdaten-Infrastruktur nachhaltig abgelegt und gesichert sowie ggf. veröffentlicht werden.

⁹ Vgl. beispielsweise The Metric Tide, Report of the Independent Review of the Role of Metrics in Research Assessment and Management, July 2015. DOI: 10.13140/RG.2.1.4929.1363.
http://www.hefce.ac.uk/media/HEFCE,2014/Content/Pubs/Independentresearch/2015/The,Metric,Tide/2015_metric_tide.pdf

¹⁰ Vgl. The National Science Foundation, Proposal and Award Policies and Procedures Guide, Part I – Grant Proposal Guide (October 25, 2016), Chap. II, C.2.f(c). https://www.nsf.gov/pubs/policydocs/pappg17_1/nsf17_1.pdf

¹¹ Earth System Science Data. The Data Publishing Journal. <http://www.earth-system-science-data.net>.

Bei Forschungsdaten, auf denen Veröffentlichungen aufbauen, ist aufgrund der Regeln der „Guten wissenschaftlichen Praxis“ ohnehin eine 10-jährige Aufbewahrung zu gewährleisten. Darüber hinaus wird von Journalen zunehmend direkter Zugang zu ihnen eingefordert.

Bei der Archivierung und gegebenenfalls Zugänglichmachung von Forschungsdaten müssen auch die Implikationen, die sich aus dem geltenden Recht, bzw. vertraglich begründeten Rechten Dritter ergeben, beachtet werden.

Sofern die notwendigen Festlegungen im Datenmanagementplan fixiert wurden, sollte der Implementierung einer solchen Langzeitverfügbarkeit durch eine geeignete Dateninfrastruktur grundsätzlich nichts im Wege stehen. Rechte – einschließlich moralischer Rechte - oder Nutzungsszenarien nachträglich zu ermitteln, ist dagegen so gut wie unmöglich.

- Die langfristige Archivierung setzt unter Umständen auch eine mehrfache Verlagerung der Daten auf Medien unterschiedlichen Typs voraus. Parameter, welche Einfluss auf die Qualität der Daten haben können, sind dabei zu erfassen und festzuhalten. Dies kann beispielsweise der Fall sein, wenn Methoden der Komprimierung eingesetzt werden, die nicht verlustfrei sind.
- Die Sicherstellung der Nachnutzbarkeit macht es insbesondere bei Daten, die in proprietären Datenformaten abgelegt werden, notwendig, den Zugriff auf Software zum Lesen und Verarbeiten dieser Daten zu gewährleisten. Die Archivierung der einschlägigen Software – ob offen oder proprietär – ist daher seitens einer Forschungsdaten-Infrastruktur in die Überlegungen zu einer langfristigen Nutzungsperspektive einzubeziehen.
- Zudem ist auch ein möglicher Ersatz von Formaten und durchaus auch von Metadatenschemata durch neue Formate zu beachten. „Alte“ Daten sind ggf. dieser Entwicklung anzupassen.
- Wesentlicher Bestandteil der für Archivierung und Zugänglichmachung notwendigen Metadaten ist in jedem Fall das Rechteschema, welches eine Zusammenstellung der Zugriffsrechte über den gesamten Lebenszyklus der Daten hinweg beinhalten sollte.
- Detailliertere Festlegungen zum jeweiligen Ende des Lebenszyklus' der Daten sind auch vor dem Hintergrund der immensen Steigerung des weltweiten Datenaufkommens von Bedeutung. Es wird wirtschaftlich nicht möglich sein, jedes Datum auf unbestimmte Zeit (und zum Beispiel über technische und ontologische Weiterentwicklungen hinweg) zu bewahren.
- Teil eines solchen Rechteschemas kann auch eine gezielte Löschung von Daten nach einer festgelegten Periode sein. Vor diesem Hintergrund sind ggf. Regelungen zu treffen, in welcher Form Daten nach Ablauf des Lebenszyklus zu löschen sind und wie deren Löschung zu dokumentieren ist. (s.a. Kapitel Rechtsfragen)

Forschungsdaten-Infrastrukturen

Ein nachhaltiges Forschungsdaten-Management stellt vielfältige technische und organisatorische Anforderungen u.a. an die Forschungsdaten-Infrastrukturen. Sie sollen effektives Arbeiten mit den Daten, deren langfristige Archivierung, den Austausch mit Kooperationspartnern sowie die Veröffentli-

chung von Daten gewährleisten. Die Speicherung und Archivierung der Forschungsdaten soll in vertrauenswürdigen Helmholtz-internen oder -externen Forschungsdaten-Infrastrukturen erfolgen.¹²

Der Begriff Forschungsdaten-Infrastrukturen meint hier immer die Kombination aus Hardware, Software und kompetentem Personal. Die zum gesicherten Betrieb, zur Weiterentwicklung, zur Qualitätssicherung der Daten sowie zur Schaffung der Forschungsdaten-Infrastrukturen aufgebauten Kompetenzen sind zugleich ein wertvolles Reservoir für die Qualifizierung von Mitarbeitern der Zentren, Partner und Communities.

Qualifizierung

Aufgrund der hohen Bedeutung und des steil wachsenden Bedarfs an qualifiziertem Personal für Forschungsdaten-Management und -Analyse in Forschung und Industrie ist aktuell dringend der Aufbau von Ausbildungsangeboten erforderlich.

Die Zentren stellen – ggf. in Zusammenarbeit mit Hochschulen und/oder Firmen – ihren Mitarbeiterinnen und Mitarbeitern ein professionelles Qualifizierungs- und Unterstützungsangebot zur Verfügung, das an den spezifischen Erfordernissen ihrer Fachdisziplinen und Infrastruktureinrichtungen ausgerichtet ist.

Finanzierung

Die Helmholtz-Zentren sehen die Notwendigkeit eines dauerhaften zusätzlichen Aufwands für das Forschungsdaten-Management und entsprechen dem durch strukturelle und finanzielle Maßnahmen.¹³ Drittmittel ergänzen, bilden aber nicht die Grundlage der Finanzierung des Forschungsdaten-Managements.

Rechtsfragen

Die Realisierung des offenen Zugangs zu Forschungsdaten (s. Kapitel "Offener Zugang") macht es notwendig, den einschlägigen rechtlichen Kontext mit in den Blick zu nehmen.

Sowohl eine Zugänglichmachung von Forschungsdaten im Sinne von Open Science als auch deren wirtschaftliche Nutzung erfordern eine Prüfung der notwendigen Verfügungsberechtigung.¹⁴ Dabei sind u.a. das Arbeitsrecht, das Arbeitnehmererfindergesetz und das Grundgesetz (Freiheit der

¹² <http://www.re3data.org>

¹³ Rat für Informationsinfrastrukturen (3.V.2016): Leistung aus Vielfalt. Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland, Göttingen, Rat für Informationsinfrastrukturen, S. 63ff. www.rfii.de/?wpdmdl=1998

¹⁴ Hier wird der Begriff Verfügungsberechtigung benutzt, weil bei nicht-materiellen Gütern der Eigentumsbegriff nicht wie bei materiellen Gütern zur Anwendung kommt. Im Text wird der Begriff "Eigentum" trotzdem teilweise verwendet, weil er umgangssprachlich richtig vermittelt, dass es darum geht, die natürliche oder juristische Person zu benennen, die gegebenenfalls eine beabsichtigte Nutzung erlauben kann.

Wissenschaft) zu beachten. Zu den eigentumsrechtlichen Fragen kann auch zählen, ob zu archivierende Inhalte urheberrechtlich geschützt sind, ob es sich um Betriebs- oder Geschäftsgeheimnisse handelt, oder mit welcher Lizenz Inhalte gegebenenfalls exklusiv weitergegeben oder veröffentlicht werden sollen. Weitere rechtliche Rahmenbedingungen können sich aus so unterschiedlichen Feldern wie dem Datenschutz oder Exportkontrollen ergeben.

Der Schutz personenbezogener Daten ist selbstverständlich insbesondere bei (bio-)medizinischen Daten von hoher Bedeutung. Die gesetzlichen Regelungen hierzu sind zurzeit durch die neue europäische Datenschutz-Grundverordnung im Umbruch begriffen. Klare und einfache Aussagen sind daher schwierig bis unmöglich. Allerdings arbeitet der Rat für Informations-Infrastrukturen derzeit an einer Empfehlung zu dem Thema. Den Zentren wird empfohlen, dieses Papier bei der Ausgestaltung ihrer Forschungsdaten-Richtlinien zu berücksichtigen.

Die hier genannten Punkte sind Beispiele. Sie sollen die Zentren zur Prüfung sensibilisieren, welche Tatbestände bei ihnen von Relevanz sind und wie damit umgegangen werden soll. Im Anhang „Rechtsfragen“ werden im Wesentlichen detailliertere Informationen zu eigentumsrechtlichen Fragen im Kontext der Archivierung und Zugänglichmachung von Forschungsfragen gegeben. Der Arbeitskreis Open Science strebt an, diesen Anhang mit Hilfe des Arbeitskreises Recht zu ergänzen. Aufgrund der Komplexität der Fragestellungen kann jedoch derzeit nicht gesagt werden, bis wann dies realisiert werden kann.

Anhang: Rechtsfragen

Die Realisierung des offenen Zugangs zu Forschungsdaten (s. Kapitel "Offener Zugang") macht es notwendig, den einschlägigen rechtlichen Kontext mit in den Blick zu nehmen. Rechtsfragen werden hier mit dem Ziel erläutert, eine möglichst breite Nachnutzung dieser Daten durch Dritte zu erleichtern; dies wird auch als rechtliche Interoperabilität bezeichnet.¹⁵

Sowohl eine Zugänglichmachung von Forschungsdaten im Sinne von Open Science als auch deren wirtschaftliche Nutzung erfordern eine Prüfung der notwendigen Verfügungsberechtigung.¹⁶ Dabei sind u.a. das Arbeitsrecht, das Arbeitnehmererfindergesetz und das Grundgesetz (Freiheit der Wissenschaft) zu beachten. Zu den eigentumsrechtlichen Fragen kann auch zählen, ob zu archivierende Inhalte urheberrechtlich geschützt sind, ob es sich um Betriebs- oder Geschäftsgeheimnisse handelt, oder mit welcher Lizenz Inhalte gegebenenfalls exklusiv weitergegeben oder veröffentlicht werden sollen. Weitere rechtliche Rahmenbedingungen können sich aus so unterschiedlichen Feldern wie dem Datenschutz oder Exportkontrollen ergeben.

Für eigentumsrechtliche Überlegungen ist es irrelevant, ob es sich bei Daten um Forschungsdaten handelt. Daten – ob Forschungsdaten oder nicht – sind gemeinfrei, d.h. an ihnen besteht kein Eigentumsrecht. Dies gilt nicht, wenn sie sich als Werk im Sinne des Urheberrechtes qualifizieren (s. Kasten „Schöpfungshöhe und Daten“).

- Der Umstand, dass Daten gemeinfrei sind, bedeutet nicht, dass keine Kontrolle über deren Nutzung ausgeübt werden kann. Diese Kontrollmöglichkeit basiert dann jedoch nicht auf einem Eigentumsrecht an den Daten, sondern auf einem oder mehreren Vertragsverhältnissen. Im Kontext der Datenerhebung sind dies typischerweise Arbeits- oder Werkverträge. In Fällen, in welchen Daten von Dritten „gekauft“ – juristisch präziser dem „Käufer“ die Daten zur Nutzung überlassen – werden, werden die Nutzungsbedingungen einschließlich der Höhe einer möglichen Nutzungsgebühr in einem Lizenzvertrag geregelt. Aus dem Lizenzvertrag kann sich beispielsweise ergeben, wie lange die Daten gespeichert und an wen sie weitergegeben werden dürfen.

Schöpfungshöhe und Daten

Nach deutschem(!) Recht entsteht ein urheberrechtlich geschütztes Werk als Resultat einer schöpferischen Leistung mit einer Mindestschöpfungshöhe. Das Werk muss die schöpferische Leistung verkörpern. Dabei schützt das Urheberrecht nicht den Inhalt sondern die Form des Werkes. Auch wenn für das Generieren von Daten im Sinne von faktischen oder deskriptiven Information viel Kreativität bzw. Schöpfungskraft aufgewendet wurde, ist diese im urheberrechtlichen Sinne in den Daten nicht verkörpert.

¹⁵ RDA-CODATA Legal Interoperability Interest Group (20.10.2016): Legal Interoperability of Research Data: Principles and Implementation Guidelines, <https://zenodo.org/record/162241>.

¹⁶ Hier wird der Begriff Verfügungsberechtigung benutzt, weil bei nicht-materiellen Gütern der Eigentumsbegriff nicht wie bei materiellen Gütern zur Anwendung kommt. Im Text wird der Begriff "Eigentum" trotzdem teilweise verwendet, weil er umgangssprachlich richtig vermittelt, dass es darum geht, die natürliche oder juristische Person zu benennen, die gegebenenfalls eine beabsichtigte Nutzung erlauben kann.

- Innerhalb der Helmholtz-Gemeinschaft werden diese Lizenzverträge in der Regel zwischen einem Zentrum und dem Datenlieferanten geschlossen. Lizenznehmer und damit im Rahmen der Vertragsbedingungen Verfügungsberechtigter ist damit das jeweilige Zentrum. Werden Daten von Mitarbeitern von Zentren im Rahmen ihrer dienstlichen Tätigkeit erhoben, wird ebenfalls in der Regel das Verfügungsrecht beim Arbeitgeber, sprich dem Zentrum, liegen. Diese Situation kann komplexer werden, wenn Mitarbeiter die Daten erheben, die gleichzeitig bei mehr als einem Arbeitgeber angestellt sind, oder Daten im Rahmen von Projekten mit Dritten erhoben werden.
- Die Feststellung, dass Daten – wenn sie sich nicht als Werk im urheberrechtlichen Sinne qualifizieren – gemeinfrei sind, bedarf einer Ergänzung. Werden Daten innerhalb einer Datenbank gemäß des Urheberrechtes (§§ 87a ff.) vorgehalten, gilt in der Europäischen Union ein Datenbankschutz. Der Datenbankschutz ist ein Leistungsschutzrecht für die Investition in die Herstellung der Datenbank. Es etabliert kein Eigentumsrecht an deren Inhalt. Es bindet die Nutzung der Datenbank und das Kopieren von wesentlichen Teilen ihres Inhaltes an die Genehmigung des Datenbankherstellers. Datenbankhersteller ist die natürliche oder juristische Person, die die Erstellung der Datenbank bezahlt hat. Innerhalb der Helmholtz-Gemeinschaft kann deshalb im Regelfall davon ausgegangen werden, dass ein Helmholtz-Zentrum allein oder zusammen mit Projektpartnern Datenbankhersteller ist und damit über das Recht verfügt, über die Nutzungsbedingungen dieser Datenbanken zu entscheiden.

Aus den Erläuterungen oben folgt: Bereits bei der Planung von Forschungsprojekten sollen – am besten innerhalb von Datenmanagementplänen – die Informationen zusammengestellt werden, die zur Klärung der Eigentums- bzw. Verfügungsrechte benötigt werden. Sollen die Daten Dritten zugänglich gemacht werden, werden mit der Weitergabe oder Veröffentlichung der Daten als Metadatum Informationen ergänzt, wer Verfügungsberechtigter der Daten und gegebenenfalls Datenbankhersteller ist und unter welchen Bedingungen die Daten bzw. die Datenbank genutzt werden dürfen. Dafür sollen, soweit vorhanden, international verbreitete Standardlizenzen genutzt werden.

- In Fällen, in welchen ein urheberrechtlicher Schutz der Daten besteht, wird eine Nachnutzungsmöglichkeit besonders gut durch die Lizenzierung mit der Lizenz CC BY erreicht. Diese Lizenz schließt ab der aktuellen Version 4.0 auch die einfache Übertragung etwaig bestehender Datenbankrechte ein. Die Lizenz erlaubt eine kommerzielle Nutzung der lizenzierten Werke. Soll diese ausgeschlossen werden, wird die Lizenz um den Zusatz Non-Commercial ergänzt.
- In Fällen, in welchen Daten veröffentlicht werden sollen, die keinen urheberrechtlichen Schutz genießen – der Regelfall! –, sollten diese mit der Lizenz CC 0 (zero) versehen werden. Mit dieser wird erklärt, dass gegebenenfalls bestehende Rechte an den so lizenzierten Inhalten, soweit dies gesetzlich möglich ist, aufgegeben werden. Sollen diese Daten in Form einer geschützten Daten-

Lizenz CC 0

Die Creative Commons Lizenz Zero (CC 0) ist die Erklärung, dass der Rechteinhaber u.U. bestehende Urheberrechte aufgibt. Gemeinfreie Daten benötigen solch eine Lizenz nicht. Die Nutzung der Lizenz wird hier trotzdem empfohlen, weil sie, wenn die Gemeinfreiheit der Daten nicht angezweifelt wird, nicht schadet und anderenfalls die gewünschte Freigabe bewirkt.

bank weitergegeben werden, gilt die Lizenz CC 0 auch für diese. Theoretisch kann die Datenbank auch separat lizenziert werden. In Frage kommen Lizenzen wie CC BY oder Open Database License v1.0 (ODbL). Eine begründete Empfehlung für eine bestimmte Lizenz zur Lizenzierung der Datenbank bedarf einer Prüfung der jeweiligen Umstände. Auch bei der Weitergabe gemeinfreier Daten kann vertraglich eine kommerzielle Nutzung ausgeschlossen werden.

- Ein möglicher Nachteil, der mit der Nutzung der Lizenz CC 0 verbunden ist, scheint aus der Perspektive von Wissenschaftlerinnen und Wissenschaftlern ein Verzicht auf die Pflicht zur Nennung der Quelle. Der Wegfall dieser Pflicht aus urheber- bzw. vertragsrechtlichen Gründen entbindet die Nutzer der Daten jedoch nicht von den Verpflichtungen, die sich aus der guten wissenschaftlichen Praxis ableiten. Diese gebietet die Nennung von Quellen.
- Schließlich ist es möglich, dass für die Nutzung der Daten und/oder der Datenbank auch Software zur Verfügung gestellt werden soll. Software ist urheberrechtlich geschützt und bedarf deshalb einer separaten Lizenzierung, z.B. durch die GNU GPL oder Apache Lizenzen. Auch hier gilt, dass für die Empfehlung einer Lizenz die Umstände geprüft werden müssen.¹⁷

Liegen Gründe vor, die gegen die oben erläuterte standardisierte Lizenzierung für jedermann sprechen, soll darauf geachtet werden, dass eine exklusive Lizenzierung die Forschungsdaten zumindest für wissenschaftliche Zwecke frei nachnutzbar belässt.

Ein zentraler Punkt für die Policy jeder Forschungsdaten-Infrastruktur ist die Entscheidung darüber, ob Daten, über deren weitere Nutzung die Infrastruktur nicht autonom entscheiden kann, angenommen werden sollen. Auch in diesem Zusammenhang ist sicher zu stellen, dass geklärt und in den Metadaten dokumentiert ist, wer die Verfügungsgewalt bzw. die Rechte an den Daten innehat und auf welcher Rechtsgrundlage (Rechtekette) diese Lage beruht.

Einfacher stellt sich die eigentumsrechtliche Frage für Betreiber von Großgeräten dar, die lediglich Daten, die bei der Nutzung ihrer Geräte gewonnen werden, behalten und öffentlich zugänglich machen wollen, d.h. die keine externen Daten übernehmen wollen. Die durch die Nutzung der Geräte gewonnenen "Messwerte" sind gemeinfrei. Die Betreiber der Großgeräte können in ihren Nutzungsordnungen festlegen, dass die Nutzung der Geräte mit einer Zustimmung der Nutzer zur freien Nutzung der gewonnenen Daten durch den Betreiber verbunden ist. Sollen Embargofristen gelten, sind deren Länge und der Zeitpunkt des Fristbeginns in der Nutzungsordnung festzulegen. Wichtig ist, die Zustimmung der Nutzer zur Nutzungsordnung zu dokumentieren und diese Dokumentation aufzubewahren.

Zur Minimierung möglicher Haftungsrisiken im Kontext einer Weitergabe sollten die Daten mit einer Erklärung versehen werden, aus der hervorgeht, dass keine Garantie für deren Richtigkeit (ohne Gewähr) übernommen wird und die Nutzung der Daten auf eigenes Risiko erfolgt. Der genaue Wortlaut der Erklärung soll in Abstimmung mit der jeweiligen Rechtsabteilung des Zentrums festgelegt

¹⁷ Die Unterschiede der Lizenzen werden z. B. in Wikipedia erläutert: <https://de.wikipedia.org/wiki/Apache-Lizenz>

werden. Die Zentren sollen die entsprechenden Erklärungen untereinander soweit wie möglich harmonisieren.

Akzeptieren Forschungsdaten-Infrastrukturen der Zentren personenbezogene Daten, muss die Einhaltung der datenschutzrechtlichen Bestimmungen sichergestellt werden. Die Zentren als Betreiber solcher Infrastrukturen sind dann für die Einhaltung der Datenschutzbestimmungen verantwortlich. Daraus können sich bereits beim Daten-Ingest¹⁸ Anforderungen an die Datenlieferanten ergeben, die in einer Policy der Forschungsdaten-Infrastruktur erläutert werden sollen. Forschungsdaten-Infrastrukturen, die zum Daten-Ingest ein breites Spektrum an Daten annehmen, bei denen nicht immer offensichtlich ist, ob diese auch Personenbezug beinhalten, sollten sich von den Datenlieferanten gegebenenfalls bestätigen lassen, dass die importierten Daten keinen Personenbezug enthalten.

¹⁸ Der Begriff Daten-Ingest bezeichnet den gesamten Vorgang der Überführung von Daten vom Datenlieferanten in die Daten-Infrastruktur. Dies kann ein einfaches Hochladen sein. Es kann sich dabei aber auch um einen komplexen Prozess handeln, der mit einem Vertragsabschluss und/oder der Kuratierung der Daten verbunden ist.

HELMHOLTZ
Open Science

