

Leibniz-Zentrum für Marine Tropenforschung (ZMT) GmbH

Open Data Policy

Beschlossen am 28. September 2015

Präambel

Wissenschaftliche Forschungsergebnisse mit Partnerinstitutionen in den Tropen und darüber hinaus zu teilen, gehört zur Identität des ZMT. Dieser Gedanke wurde während der Evaluierung des ZMT im Jahr 2013 aufgegriffen und mündete in der Empfehlung, die **Forschungsdaten des Instituts der wissenschaftlichen Gemeinschaft durch Open Access zur Verfügung zu stellen**. Das ZMT folgt dieser Empfehlung und strebt verstärkt an, Originaldaten mit Partnerinstitutionen und anderen Leibniz-Einrichtungen zu teilen. Außerdem ist geplant, die Informationsinfrastrukturen der Leibniz-Gemeinschaft intensiver zu nutzen. Als Grundlage für die Umsetzung der Empfehlung hat das ZMT diese Open Data Policy entwickelt und im Herbst 2015 verabschiedet.

Die Veröffentlichung von Forschungsergebnissen in digitaler und frei verfügbarer Form beinhaltet nicht nur Fachpublikationen sondern auch Originaldaten.¹ Qualitätsgesicherte Forschungsdaten bilden einen Grundpfeiler wissenschaftlicher Erkenntnisse und können unabhängig von ihrem ursprünglichen Erhebungszweck vielfach Grundlage weiterführender Forschung sein. Eine **nachhaltige Sicherung und Bereitstellung von Forschungsdaten** dient daher auch der Gewinnung künftiger Forschungsergebnisse.

Forschungsdatenmanagement am ZMT

Das ZMT hat 2012 mit dem "**Archive of Scientific Data**" ein zentrales Forschungsdatenarchiv eingeführt, das aktuell über das Intranet des ZMT aufgerufen werden kann. Das ZMT hält

¹ Vgl. Leitlinie zu Open Access der Leibniz-Gemeinschaft, verabschiedet am 23.11.2007; URL: http://www.leibniz-gemeinschaft.de/fileadmin/user_upload/downloads/Infrastruktur/OpenAccess-Leitlinie.pdf (zuletzt abgerufen am 19.08.2015)

seine Wissenschaftler ausdrücklich dazu an, Forschungsdaten in diesem Forschungsdaten-repositorium zur Verfügung zu stellen. So wird die dauerhafte Verfügbarkeit der Daten gemäß den von der Deutschen Forschungsgemeinschaft (DFG) verabschiedeten Empfehlungen zur **Sicherung guter wissenschaftlicher Praxis** gewährleistet.²

Die DFG schlägt u.a. vor, dass "Primärdaten als Grundlage für Veröffentlichungen [...] auf haltbaren und gesicherten Trägern in der Institution, wo sie entstanden sind, **zehn Jahre lang aufbewahrt werden [sollen]**".³ Außerdem ist für Projektbeantragungen bei der DFG und der EU eine Erörterung zum geplanten Datenmanagement gefordert.^{4,5}

Ein verlässliches Forschungsdatenmanagement muss vielfältige technische und organisatorische Anforderungen erfüllen, die vom wissenschaftlichen Personal gemeinsam mit den wissenschaftsunterstützenden Einrichtungen - vor allem IT und Bibliothek - definiert und umgesetzt werden.⁶

Jeder Wissenschaftler ist für die Speicherung der einem Projekt oder einer Publikation zugrunde liegenden Originaldaten verantwortlich. Es ist unbedingt erforderlich, alle relevanten Metadaten, die zum Verständnis und zur Weiterbearbeitung der Daten notwendig sind, mit anzugeben. Idealerweise geschieht dies mit der Erarbeitung eines Datenmanagementplans.

In einem **Datenmanagementplan** werden in der Projektplanungsphase die wichtigsten Festlegungen im Umgang mit Forschungsdaten zusammengefasst. Dies beginnt mit einer Abschätzung der anfallenden Datenmengen und ihrer inhaltlichen Klassifizierung. Es sind zudem Angaben zur Richtigkeit, Vollständigkeit, Integrität, Authentizität und Vertraulichkeit der Daten zu machen.

Mit dem Eintrag in das "Archive of Scientific Data" gewährt der Wissenschaftler dem ZMT das Recht, die zugehörigen Metadaten kostenlos online zugänglich zu machen. Ein Zurückhalten der Veröffentlichung der Daten bis zum Ablauf von Offenlegungsfristen ist möglich.

² Vgl. DFG, Sicherung guter wissenschaftlicher Praxis, ergänzte und aktualisierte Denkschrift, verabschiedet am 03.07.2013; URL: http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_1310.pdf (zuletzt abgerufen am 19.08.2015)

³ DFG, Sicherung guter wissenschaftlicher Praxis, ergänzte und aktualisierte Denkschrift, verabschiedet am 03.07.2013, S. 21.

⁴ Vgl. DFG-Vordruck: Leitfaden für die Antragstellung (54.01 - 06/2014), S. 5-6; URL: http://www.dfg.de/formulare/54_01/54_01_de.pdf (zuletzt abgerufen am 19.08.2015)

⁵ Vgl. Guidelines on Data Management in Horizon 2020, Version 1.0, 11.12.2013; URL: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf (zuletzt abgerufen am 19.08.2015)

⁶ Vgl. <http://www.allianzinitiative.de/de/handlungsfelder/forschungsdaten/grundsaeetze/> (zuletzt abgerufen am 13.08.2015)

Die Forschungsdaten werden zeitlich unbegrenzt gespeichert. Der Dateneintrag in das Forschungsdatenarchiv soll möglichst zeitgleich mit der Veröffentlichung der zugehörigen Publikation, spätestens mit Ablauf des Projektes oder vor dem Ausscheiden aus dem ZMT, geschehen.

Neben Forschungsdaten, die im Rahmen von Forschungsarbeiten am ZMT erhoben worden sind, sollen auch **weitere Daten und Produkte dem ZMT zur einfachen Nutzung unbeschränkt zur Verfügung gestellt** werden. Zusammenfassend sind folgende Produkte zur Archivierung und einfachen Nutzung vorgesehen:

- Fachpublikationen in PDF-Form zur internen Speicherung
- Metadaten der zu Forschungsdaten gehörenden Publikationen
- ggf. nicht publizierte Forschungsdaten
- Abschlussarbeiten (Bachelor-, Master-, Doktorarbeiten)
- Bildmaterial, das in Verbindung mit den Forschungsarbeiten entstanden ist
- Videomaterial, das in Verbindung mit den Forschungsarbeiten entstanden ist

Der Wissenschaftler ist verantwortlich für die Prüfung der Zulässigkeit der Datenfreigabe bezüglich personenbezogener Daten, Urheberrechten und eventueller Interessen Dritter. Die Verwaltung steht für Rückfragen als Ansprechpartner zur Verfügung.

Checkliste für das Forschungsdatenmanagement

Ohne definierte Arbeitsabläufe für das Datenmanagement können nur geringe Datenmengen mit unklarer Qualität gesichert werden. Für die sachgerechte Planung eines Forschungsvorhabens wird deshalb der Einsatz einer Checkliste empfohlen:

- **Planung und Erstellung** (*Organisation durch den Wissenschaftler*)
 - Sind alle Rahmendaten des Projekts dokumentiert (Projektname, Finanzgeber, Laufzeit, Leiter, Partner, Ziele)?
 - Wie können bereits existierende Daten integriert/nachgenutzt werden?
 - Welche Bedeutung haben die Daten für die Projektziele?
 - Wie lassen sich die verwendeten/erzeugten Daten charakterisieren (Datenarten, Formate, Reproduzierbarkeit)?
 - Wie werden die Daten erfasst/erstellt?
 - Wie groß ist die Datenmenge/Produktionsrate?
 - Wie erfolgt die Qualitätssicherung?
- **Auswahl und Aufbewahrungsdauer** (*Organisation durch den Wissenschaftler*)
 - Wieso müssen welche Daten aufbewahrt werden?
 - Wann, durch wen und womit erfolgt die Datenauswahl?
- **Einspeisen und Verantwortungsübernahme** (*Organisation durch den Wissenschaftler in Rücksprache mit der IT*)
 - Wann und wie werden die Daten übergeben/übertragen?
 - Wann und von wem werden welche Metadaten erfasst?
 - Wie werden Daten und Metadaten validiert?
 - Wie wird mit sensiblen Daten umgegangen?

- **Speicherung und Infrastruktur** (*Organisation durch die IT in Rücksprache mit dem Wissenschaftler*)
 - Wer ist während des Projekts und danach für die Speicherung der Daten verantwortlich?
 - Mit welchen Technologien und an welchen Orten werden die Daten gespeichert?
 - Gibt es besondere Infrastruktur-Anforderungen für Datentransfer, -zugriff und -nutzung? (Netzwerk-Bandbreite, Hardware etc.)

- **Erhaltungsmaßnahmen** (*Organisation durch den Wissenschaftler in Zusammenarbeit mit der IT*)
 - Sind die eingesetzten Technologien sowie Abhängigkeiten, Nutzungszielgruppe und -anforderungen dokumentiert?
 - Wird regelmäßig überprüft, ob sich die Anforderungen, verfügbaren Technologien oder Abhängigkeiten verändert haben?
 - Wird die Aufbewahrungswürdigkeit regelmäßig überprüft?
 - Gibt es eine Nachfolgeregelung bei einem Wechsel der aufbewahrenden Institution?

- **Zugriff und Nutzung** (*Organisation durch den Wissenschaftler in Zusammenarbeit mit der IT*)
 - Gibt es Verpflichtungen, Daten freizugeben oder nicht freizugeben?
 - Wie werden Veröffentlichung, Suchbarkeit und Zugriff realisiert?
 - Zu welchem Zweck und mit welcher Software werden die Daten voraussichtlich genutzt?
 - Wie werden ggf. Erstnutzungsrecht, Nutzungsbeschränkungen oder Lizenzbedingungen durchgesetzt?

- **Management, Organisation und Policies** (*Organisation durch den Wissenschaftler*)
 - Haben alle Beteiligten eingewilligt und sind deren Beiträge definiert?
 - Ist der Ablauf des Datenmanagements beschrieben?
 - Sind alle Ressourcen beschrieben und abgeschätzt?
 - Welche Anforderungen/Auflagen/Policies müssen von wem berücksichtigt und umgesetzt werden?
 - Wie wird die Einhaltung überprüft?

- **Rechtliche Aspekte von Forschungsdaten** (*Prüfung durch den Wissenschaftler ggf. In Rücksprache mit dem Forschungsdatenbeauftragten, Meldung an die IT*)
 - Sind die Daten aufgrund des Datenschutzes oder aus anderen Gründen sensibel?
 - Werden fremde Daten oder Software verwendet, welche dem Urheberrecht, dem Patentrecht o. Ä. unterliegen?
 - Sind die Nutzungsbedingungen und Rechte für eigenes geistiges Eigentum geklärt?
 - Sind evtl. auslaufende Schutzfristen zu berücksichtigen?

- **Metadaten** (*Organisation durch den Wissenschaftler*)
 - Welche Voraussetzungen bestehen für die Verarbeitung und das Verständnis der Metadaten?

- **Identifikatoren und Informationsobjekte** (*Prüfung durch den Wissenschaftler ggf. in Rücksprache mit dem Forschungsdatenbeauftragten, Meldung an die IT*)
 - In welchem Verhältnis stehen die Informationsobjekte zueinander?
 - Für welche Informationsobjekte werden dauerhafte Identifikatoren benötigt und welche werden benutzt (DOI, URN, Handle etc)?
 - Wer pflegt und aktualisiert die Identifikatoren?

Zum Vergleich wird auf die Checkliste zum Forschungsdaten-Management des deutschen WissGrid-Projekts verwiesen.⁷ Weitere Informationen finden sich auch in Beispielen und Checklisten des britischen Digital Curation Centre (DCC).^{8,9}

Über die Webanwendung der Forschungsdatenbank soll im Sinne dieser Open Data Policy künftig ein externer Zugriff auf die Forschungsdaten ermöglicht werden. Die Umsetzung wird die folgenden abgestuften Zugriffsrechte beinhalten, welche dem Wissenschaftler bei der Dateneinspeisung zur Auswahl stehen werden:

- Vollzugriff auf alle Metadaten und Daten ohne Authentifikation.
- Zugriff auf die Metadaten inkl. Beschreibung. Vollzugriff nur nach Authentifikation, womit protokolliert werden kann, wer wann welche Daten genutzt hat.
- Zugriff auf die Metadaten inkl. Beschreibung. Vollzugriff nur nach Genehmigung des Eigentümers. Eine entsprechende Anfrage wird nach Authentifikation ermöglicht.

Der Versand schützenswerter Daten an Kooperationspartner über Cloud-Dienste externer Anbieter ist kritisch zu betrachten. Der hauseigene Cloud-Dienst ZMT-Owncloud (<https://zmtcloud.zmt-bremen.de/owncloud/>) ist zu bevorzugen.

Struktur der Forschungsdatenbank

Für die Archivierung von Forschungsdaten sind alle relevanten Metadaten anzugeben. Dies umfasst abhängig von der Forschungsdisziplin i.d.R. die folgenden Attribute:

- **Projekttitle** (Titel des Projekts in dem die Daten erhoben wurden, in max. 200 Zeichen, englisch)
- **Kurztitel/Akronym** (Titel des Projekts in dem die Daten erhoben wurden, in max. 80 Zeichen, englisch)
- **Kontaktperson** (Name des projektbetreuenden Wissenschaftlers)
- **Beteiligte Wissenschaftler** (Namen der an dem Projekt beteiligten Wissenschaftler)

⁷ Vgl. WissGrid: Checkliste zum Forschungsdaten-Management, 2013, S. 83-97; URL: http://www.wissgrid.de/publikationen/Leitfaden_Data-Management-WissGrid.pdf (zuletzt abgerufen am 19.08.2015)

⁸ Vgl. DCC: Data plan guidance and examples, 2013; URL: <http://www.dcc.ac.uk/resources/data-management-plans/guidance-examples> (zuletzt abgerufen am 19.08.2015)

⁹ Vgl. DCC: Checklist for a Data Management Plan, Version 4.0, 2013; URL: http://www.dcc.ac.uk/sites/default/files/documents/resource/DMP/DMP_Checklist_2013.pdf (zuletzt abgerufen am 19.08.2015)

- **Abteilung/Arbeitsgruppe** (Nennung der verantwortlichen Abteilung und Arbeitsgruppe)
- **Schlagwörter** (suchbare Deskriptoren zur inhaltlichen Beschreibung der Daten, max. 10)
- **Taxon** (Angaben der lateinischen Namen beteiligter Spezien, max. 10)
- **Studientyp** (Nennung der Studienart, z.B. Experiment, Umfrage, Feldstudie)
- **Ort der Erhebungen** (Nennung von Stadt, Land, Gebiet)
- **Koordinate 1** (Koordinate 1 des Ortes, an dem die Daten erhoben wurden)
- **Koordinate 2** (Koordinate 2 des Ortes, an dem die Daten erhoben wurden)
- **Beginn der Studie** (einheitliches Datumsformat)
- **Ende der Studie** (einheitliches Datumsformat)
- **Datenkontext** (z.B. Doktorandenprojekt, Masterarbeit)
- **Abstract** (Eingabe einer zusammenfassenden Beschreibung des Projekts in dem die Daten erhoben wurden)
- **Methodenbeschreibung** (Beschreibung aller verwendeten Materialien, Messgeräte und Methodenparameter, so dass ein Experiment ggf. anhand dieser Beschreibung reproduziert werden kann)
- **Dateien** (mindestens die Rohdaten mit den Metadaten, die für die Verwendung der eigentlichen Daten notwendig sind: Struktur der Informationen, Inhalte etc., sowie andere hilfreichen Daten wie GIS Dateien etc.)
- **Publikationen**, denen diese Daten zugrunde liegen (Hochladen als PDF-Datei möglich)

Zur **Langzeitspeicherung und Weitergabe der archivierten Daten** sollten alle Dateien in offenen Standardformaten gespeichert werden, mindestens jedoch sollten solche Formate verwendet werden, die von der Mehrheit der Forschungsgemeinschaft genutzt und akzeptiert sind.

Für die Auswahl geeigneter Dateiformate wird auf die Liste des Portals Forschungsdaten-Bildung¹⁰ verwiesen. Empfohlen werden abhängig vom Dateityp:

Quantitative Daten (Statistik-Softwareformate)

Bevorzugte Formate:

- STATA (*.dta)
- R (*.R)
- SAS Transport (*.sas)
- geschützte Formate von Statistikpaketen, z.B. SPSS (*.sav), Stata (*.dta)

Akzeptierte Formate:

- OpenDocument-Tabellendokumente (*.ods), MS Excel (*.xls, *.xlsx), MS Access (*.mdb, *.accdb)
- CSV-Formate ohne zusätzliche Datendefinitionsdatei

¹⁰ Vgl. Empfohlene Dateiformate für die Übermittlung der Forschungsdaten-Dateien; URL: <http://www.forschungsdaten-bildung.de/formate> (zuletzt abgerufen am 22.09.2015)

- Tabulator-, Komma- oder Spaltengetrennte Textdateien (CSV) mit zusätzlicher Datendefinitionsdatei (Setup, Syntax, Command file)
- Column Binary-Formate oder Card-Image Formate

Textdateien

Bevorzugte Formate:

- PDF/A (*.pdf)
- Plain Text-Formate (ASCII)

Akzeptierte Formate:

- OpenDocument Text (*.odt)
- PDF (*.pdf)
- MS Word (*.doc, *.docx)
- RichTextFormat (*.rtf)
- HTML (*.htm, *.html)

Bilder

Bevorzugte Formate:

- TIFF Version 6 unkomprimiert (*.tif)

Akzeptierte Formate:

- JPEG (*.jpg, *.jpeg), PNG (*.png), GIF (*.gif), BMP (*.bmp)
- PDF/A, PDF (*.pdf)

Audio

Bevorzugte Formate:

- MPEG-1 Audio Layer 3 (*.mp3)

Akzeptierte Formate:

- Wave Audio Format WAV (*.wav)

Video

Bevorzugte Formate:

- MPEG-4 (*.mp4), MPEG-2 (*.mpg)

Akzeptierte Formate:

- Audio Video Interleave AVI (*.avi)
- Windows Media Video WMV (*.wmv)

Die **Datensicherheit von personenbezogenen und sensiblen Daten** wird gewährleistet, indem diese Daten nicht für alle Nutzer sichtbar gespeichert werden und ggf. nur auf Anfrage verfügbar sind.

Um den langfristigen Zugriff auf Forschungsdaten zu gewährleisten, wird es zukünftig eine **DOI-Vergabe** (Digital Object Identifier) für die Daten im "Archive of Scientific Data" geben. Der DOI identifiziert die Forschungsdaten direkt, ist ihnen permanent zugeordnet und für die Quellenangabe geeignet.

Ansprechpartner am ZMT sind Hauke Reuter (wissenschaftliche Betreuung), Thomas Rau (IT Betreuung) und Christina Schrader (mediale Betreuung).