

Einführung in das Forschungsdatenmanagement

RDA-DE-Trainings-Workshop A1, 8. Juni 2017

Kerstin Helbig

Humboldt-Universität zu Berlin, Computer- und Medienservice

<http://orcid.org/0000-0002-2775-6751>



Agenda

Teil 1 13.15-15.15 Uhr

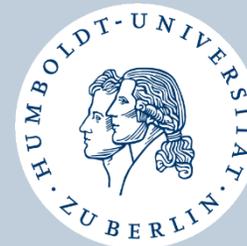
Grundbegriffe des Forschungsdatenmanagements

- Kaffeepause - 

Teil 2 15.45-17.45 Uhr

Effektive Datenmanagementplanung





Grundbegriffe des Forschungsdatenmanagements

Was sollte man kennen?



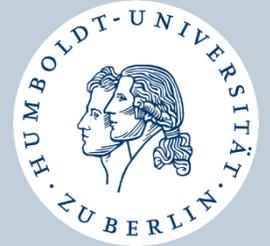


Open + X ?

- **Open Science:** Umfasst Open Access, Open Data, OER
- **Open Access:** „[...] wissenschaftliche Literatur und wissenschaftliche Materialien für alle Nutzerinnen und Nutzer frei zugänglich zu machen: kostenlos und möglichst frei von technischen und rechtlichen Barrieren“*
- **Open Data:** Zum Beispiel öffentliche Statistiken, Geodaten, Genome u. a. Bezieht sich primär auf nicht-textuelle Daten (siehe zum Beispiel <http://daten.berlin.de>)
- **Open Educational Resources (OER):** frei zugängliche und nachnutzbare Lern-/Lehrmaterialien
- **Citizen Science:** Bürger unterstützen Datensammlung; z. B. Sichtbarkeit von Sternen, Vögel zählen (siehe www.citizen-science-germany.de)

*Simukovic et al. 2014 <http://nbn-resolving.de/urn:nbn:de:kobv:11-100224755>





Entwicklung des Forschungsdatenmanagements in Deutschland

1998/2013 Deutsche Forschungsgemeinschaft (DFG)

[Sicherung guter wissenschaftlicher Praxis](#)

2009 Deutsche Forschungsgemeinschaft

[Empfehlungen zur gesicherten Aufbewahrung und Bereitstellung digitaler Forschungsprimärdaten](#)

2010 Allianz der deutschen Wissenschaftsorganisationen

[Grundsätze zum Umgang mit Forschungsdaten](#)

2014 Hochschulrektorenkonferenz (HRK)

[Management von Forschungsdaten – eine zentrale strategische Herausforderung für Hochschulleitungen](#)

2015 Deutsche Forschungsgemeinschaft

[Leitlinien zum Umgang mit Forschungsdaten](#)



Was sind Forschungsdaten?

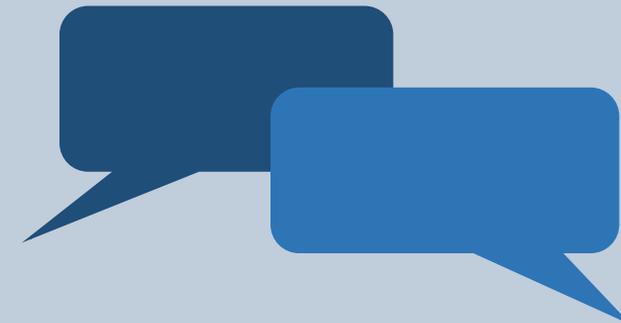
Diskutieren Sie mit Ihrem Nachbarn über die Arten von Forschungsdaten, die Sie produzieren sowie welche Vor- und Nachteile Sie beim Forschungsdatenmanagement sehen.

Notieren Sie wichtige Punkte auf einem Zettel:

- Konkrete Beispiele und Arten von Forschungsdaten in Ihrem Fach/ihrer Einrichtung
- (persönliche) Vorteile eines effektiven Forschungsdatenmanagements
- Probleme/Vorbehalte gegenüber der Veröffentlichung von Forschungsdaten

Zeit: 10 Minuten

Anschließend Sammlung der Ergebnisse





Was sind Forschungsdaten?

Forschungsdaten

- Rohdaten und deren Verarbeitungsstufen
- Simulationsdaten
- Beobachtungsdaten
- Interviewdaten
- Modelle
- Software
- Persönliche Daten
- Physische Objekte
- Abgeleitete Daten

Vorteile

- Nachnutzung
- Prüfbarkeit/
Nachvollziehbarkeit
- Wiederauffindbarkeit/
Organisationshilfe
- Neue Ideen, Kollaborationen
- Kosteneffizienz
- Vergleichbarkeit/
Interoperabilität
- Verpflichtungen nachkommen
- Vorteile bei Förderprogramme
- Qualität

Nachteile

- Erhöhter Dokumentations- und Bearbeitungsaufwand
- Zu viele Daten
- Speicherbedarf
- Ressourcenbedarf
- Fehlende Innovation
- Fehlende Standards
- „Ideenklau“
- Risiko der Urheberrechtsverletzung
- Nachprüfbarkeit
- Qualität



Was sind Forschungsdaten?

- Keine feste Definition
- Allgemein: „alle digital vorliegenden Daten, die während des Forschungsprozesses entstehen oder ihr Ergebnis sind“
([Kindling/Schirnbacher 2013](#))



Quelle: [Aust et al. 2016](#)

Forschungsdaten aus der Biologie

EMBL-EBI Services Research Training About us

Protein Data Bank in Europe
Bringing Structure to Biology

Search
 Examples: hemoglobin, BRCA1_HUMAN
 Search EMDB

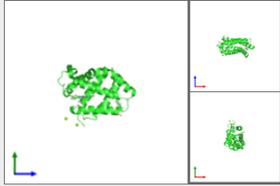
Feedback

PDBe > 2bk9

Drosophila Melanogaster globin
Source organism: *Drosophila melanogaster*

Primary publication:
Bishistidyl heme hexacoordination, a key structural property in *Drosophila melanogaster* hemoglobin.
de Sanctis D, Dewilde S, Vonrhein C, Pesce A, Moens L, Ascenzi P, Hankeln T, Burmester T, Ponassi M, Nardini M, Bolognesi M
J. Biol. Chem. **280** 27222-9 (2005)
PMID: 15917230

X-ray diffraction
1.2Å resolution
Released: 20 May 2005
Model geometry 
Fit model/data 



Quick links

- 2bk9 overview
- Citations
- Structure analysis
- Function and Biology
- Ligands and Environments
- Experiments and Validation
- View
- Downloads
- 3D Visualisation

Citations

2 review citations

The respiratory basis of locomotion in *Drosophila*.
Lehmann et al. (2010) [1 more](#)

Function and Biology Details

Biochemical function: [heme binding](#)

Biological process: [oxygen transport](#)

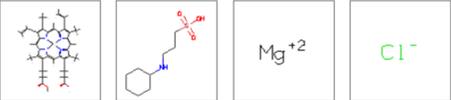
Cellular component: [cytoplasm](#)

Sequence domains:

- [Globin](#)
- [Globin-like](#)

Ligands and Environments

4 bound ligands:



[1 x HEM](#) [1 x CXS](#) [2 x MG](#) [1 x CL](#)

No modified residues

Experiments and Validation Details



Structure analysis Details

Assembly composition: monomeric ([preferred](#))

Entry contents: 1 distinct polypeptide molecule

Macromolecule:

- Globin 1, isoform B

Chain: A [Molecule details >](#)

Quelle: de Sanctis et. al. 2005. *Drosophila Melanogaster* globin. Protein Data Bank in Europe. <https://www.ebi.ac.uk/pdbe/entry/pdb/2bk9>

Forschungsdaten aus den Geowissenschaften



27	Size:	16922 data points														
28	?															
29	Date/Time	Latitude	Longitude	Ice cov [%]	Floes [%]	New ice [%]	Old ice [%]	Floe size [m]	Distance [km]	Icebergs [#kr]	Aves [#km**]	Penguins [#]	Albatrosses [1]	Petrels [#km'	Seals [#km**	Whales [#km**2]
78	1992-10-02T12	-567.037	-503.400						589							
79	1992-10-02T12	-567.157	-502.800						587							
80	1992-10-02T12	-567.275	-502.300						586							
81	1992-10-02T12	-567.392	-501.700						585							
82	1992-10-02T13	-567.505	-501.200	0					583	1.98	0.00	0.00	1.98			
83	1992-10-02T13	-567.642	-500.600	0					582	0.00	0.00	0.00	0.00			
84	1992-10-02T13	-567.761	-500.200	0					580							
85	1992-10-02T13	-567.688	-500.100						581						0.00	
86	1992-10-02T13	-567.695	-500.000						581						0.00	
87	1992-10-02T13	-567.693	-500.000						581							
88	1992-10-02T14	-567.673	-500.000						581							
89	1992-10-02T14	-567.673	-500.000						581							
90	1992-10-02T14	-567.673	-500.000						582							
91	1992-10-02T14	-567.673	-500.000						582							
92	1992-10-02T14	-567.673	-500.000						582							
93	1992-10-02T14	-567.673	-500.000						580							
94	1992-10-02T15	-567.673	-500.000						579							
95	1992-10-02T15	-567.673	-500.000	0					577	0.00	0.00	0.00	0.00			
96	1992-10-02T15	-567.673	-500.000	0					575	0.96	0.00	0.00	0.96			
97	1992-10-02T15	-567.673	-500.000	0					574	0.92	0.00	0.92				
98	1992-10-02T15	-567.673	-500.000	0					572	0.94	0.00	0.00	0.94	0.00		
99	1992-10-02T15	-567.673	-500.000	0					570						0.00	
100	1992-10-02T16	-567.673	-500.000						571							
101	1992-10-02T16	-567.673	-500.000						571						0.00	
102	1992-10-02T16	-567.673	-500.000						571							
103	1992-10-02T16	-567.673	-500.000						571							
104	1992-10-02T16	-568.653	-495.800						571							
105	1992-10-02T16	-568.642	-495.900						571							
106	1992-10-02T17	-568.645	-495.900						571							
107	1992-10-02T17	-568.616	-495.900						571							
108	1992-10-02T17	-568.621	-495.900						571							
109	1992-10-02T17	-568.624	-495.900						571							

Quelle: Smetacek et al. 1997. Ice and animal observations during POLARSTERN cruise ANT-X/6. PANGAEA. <http://doi.org/10.1594/PANGAEA.88838>

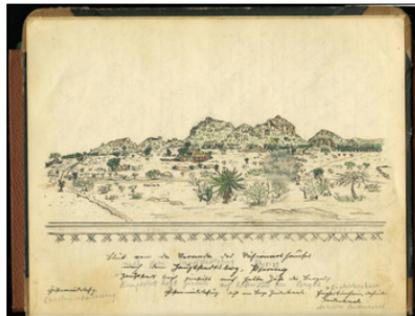


Forschungsdaten aus den Afrikawissenschaften



View of main mountain Phiring from missionary houses, Christian settlement and local kraals close to mountain

[Previous result](#) | [View all results](#) | [Next result](#)



Resource tools

File information	File size	Options
Original JPG File 5430 × 4125 pixels (22.4 MP) 46 cm × 34.9 cm @ 300 PPI	2.3 MB	Download
Low resolution print 2000 × 1519 pixels (3.04 MP) 16.9 cm × 12.9 cm @ 300 PPI	919 KB	Download
Screen 1053 × 800 pixels (0.84 MP) 8.9 cm × 6.8 cm @ 300 PPI	293 KB	Download
Preview Full screen preview	293 KB	View

> Share

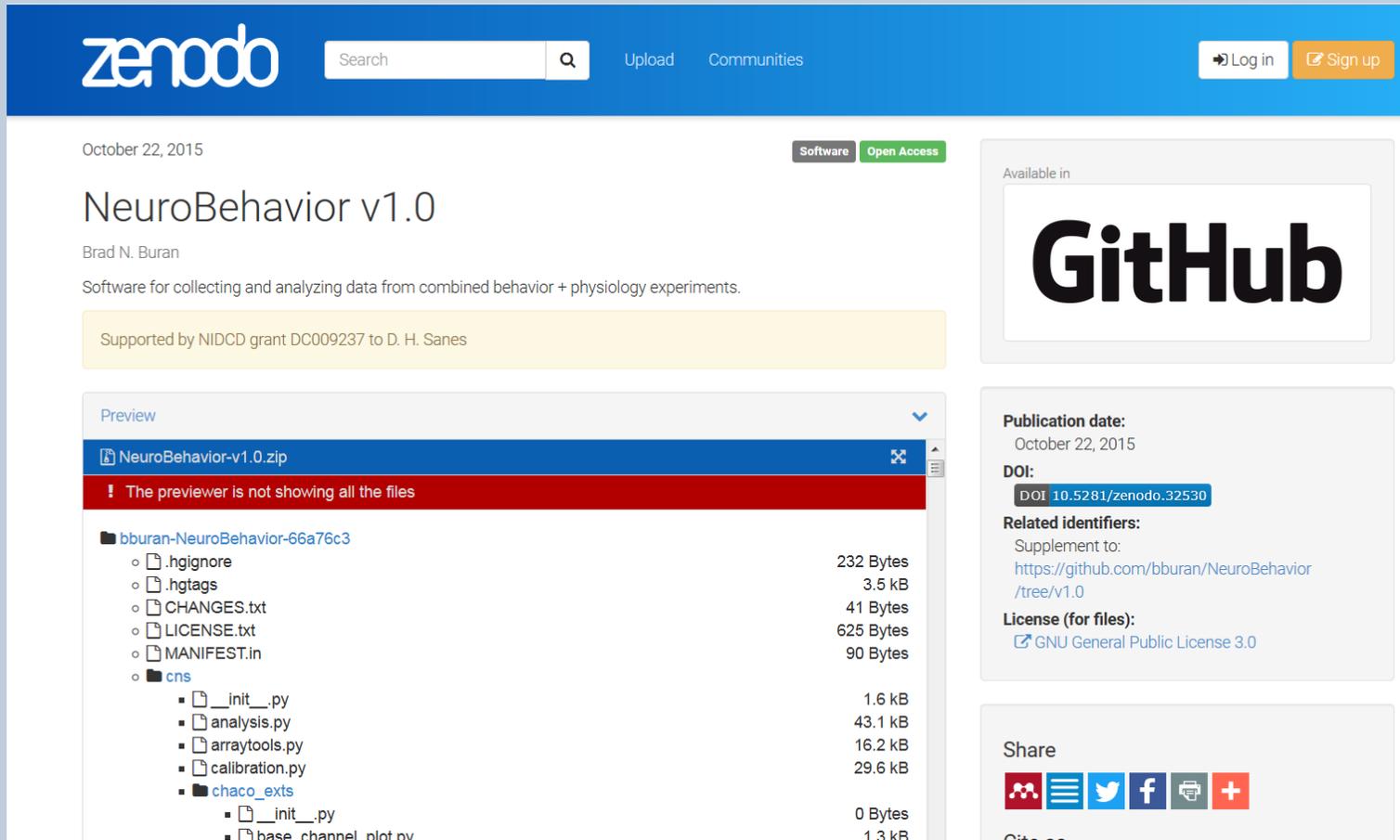
Resource details

Resource ID 1385	Access Open	Category Photos, UNISA Archive (Photos)	Source UNISA Archive	Source No. HOFFMANN DRAWINGS DIARY 5_UNISA	Location (Northern/Southern Transvaal?), South Africa	Recording date October 1897
Appeared in Das Tagebuch des Missionars Carl Hoffmann 18. Juli 1894-2. Dezember 1910, Band 5: 15. Oktober 1897-11. Juli 1901, S. 6.	Language of Source German	Type of Source Drawing	Inscription Blick von der Veranda des Missionarshauses auf den Hauptstadtberg Phiring; Hauptstadt liegt jenseits auf halber Höhe des Berges; Christenniederlassung; dicht am Berge Heidenkraale; Euphorbienhain, dahinter Heidenkraale.	Description View of main mountain Phiring from missionary houses, Christian settlement and local kraals close to mountain	Usage terms Creative Commons Namensnennung 4.0 International Lizenz	Created by Hoffmann License url  https://creativecommons.org/licenses/by/4.0/

Quelle: Joubert et. al. 2015. The Hoffmann Collection of Cultural Knowledge. Medien-Repositorym der Humboldt-Universität zu Berlin. <http://doi.org/10.17172/MR/22>



Forschungssoftware



The screenshot shows the Zenodo repository page for the software 'NeuroBehavior v1.0'. The page includes a search bar, navigation links for 'Upload' and 'Communities', and buttons for 'Log in' and 'Sign up'. The main content area displays the software title, author 'Brad N. Buran', and a description: 'Software for collecting and analyzing data from combined behavior + physiology experiments.' A yellow banner indicates support by NIDCD grant DC009237 to D. H. Sanes. A 'Preview' section shows a file tree for 'bburan-NeuroBehavior-66a76c3' with files like '.hgignore', '.hgtags', 'CHANGES.txt', 'LICENSE.txt', 'MANIFEST.in', and a 'cns' directory containing Python scripts. A red warning bar states 'The previewer is not showing all the files'. On the right, a 'GitHub' logo is shown under 'Available in', and a 'Publication date' of October 22, 2015, with a DOI of 10.5281/zenodo.32530. Related identifiers include a GitHub repository link. The license is identified as GNU General Public License 3.0. A 'Share' section contains social media icons for GitHub, Twitter, Facebook, and others.

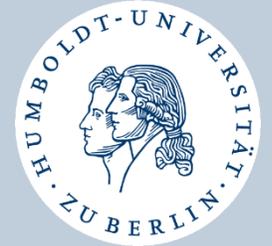
Quelle: Brad N. Buran. 2015. NeuroBehavior v1.0. Zenodo. <http://doi.org/10.5281/zenodo.32530>

Forschungsdaten als „Paket“

- Bereitstellung von Zusatzmaterial, wie zum Beispiel:
 - Codebüchern
 - zugehörigen Fragebögen
 - Software (inkl. Syntax)
 - anderem Begleitmaterial
- Dokumentation der Forschungsdaten in einer separaten Textdatei (zum Beispiel Readme)



Foto: Creative Commons CC0



Was ist Forschungsdatenmanagement?

- „alle Aktivitäten, die mit der Aufbereitung, Speicherung, Archivierung und Veröffentlichung von Forschungsdaten verbunden sind“
([Simukovic/Kindling/Schirmbacher 2013](#))
- Gründe:
 - Transparenz
 - Reproduzierbarkeit
 - Validierung
 - Wissenschaftliche Anerkennung und Reputation
 - Rechtliche und ethische Anforderungen

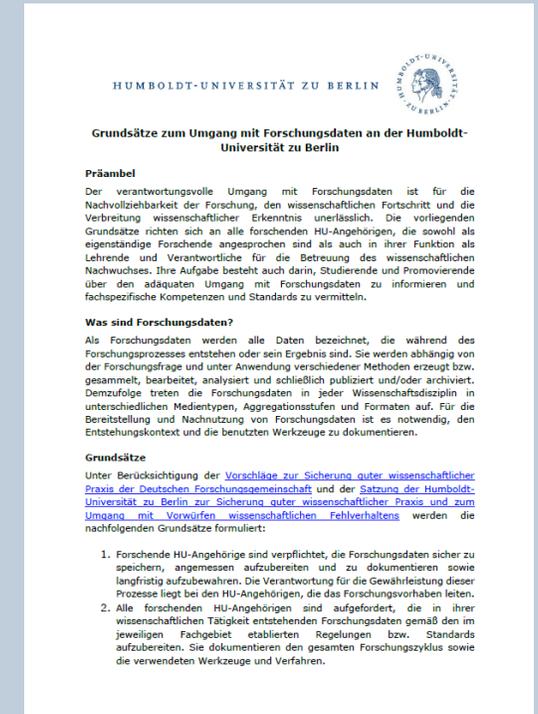
Der Zugang zu Forschungsdaten und die Erstellung eines Datenmanagementplans werden zunehmend von Forschungsförderern und Verlagen verlangt.





Was ist eine Forschungsdaten-Policy?

- Liefert grundlegenden Handlungsrahmen für den Umgang mit Forschungsdaten
- Zentrale Elemente einer Policy:
 - Zielgruppe
 - Definition von Forschungsdaten
 - Verpflichtungen
 - Aufforderungen
 - Empfehlungen
 - Beteiligte Infrastruktur der Einrichtung
 - Verbindlichkeiten
 - Verweis auf Open Access



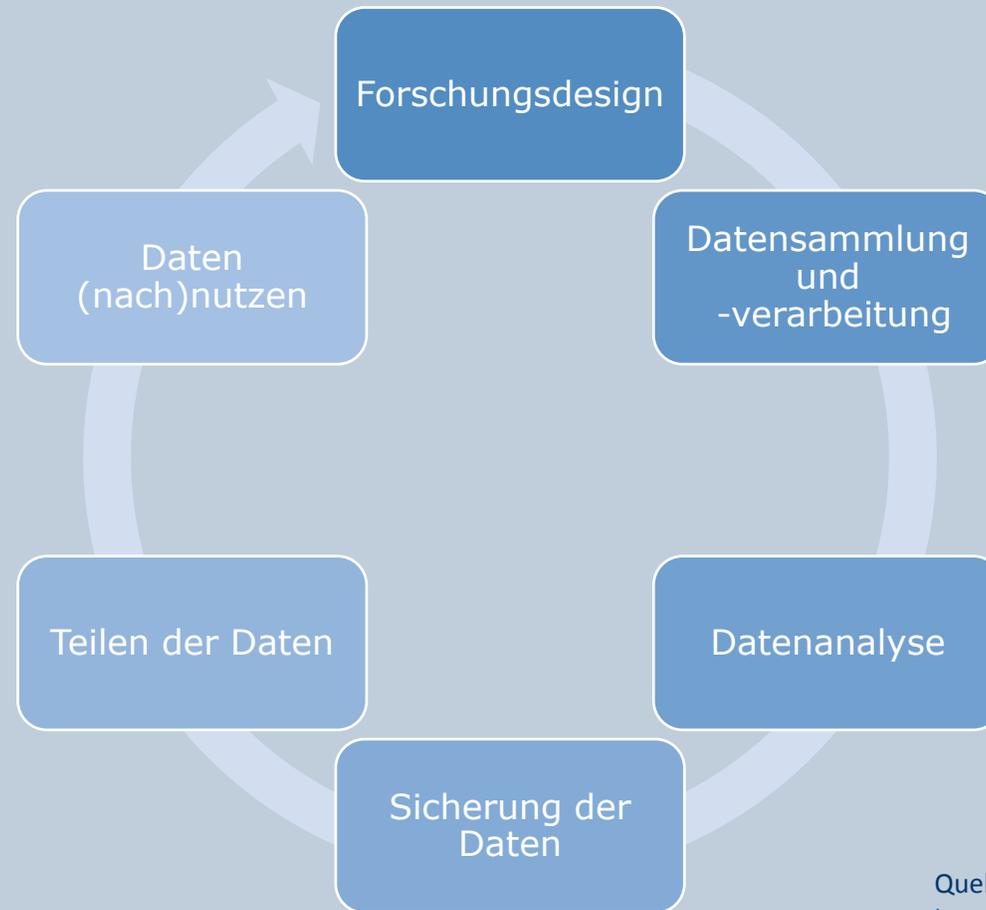


Fachspezifische Policies

Fachspezifische Entwicklungen, u. a.:

- [Richtlinien zum Umgang mit Forschungsdaten in der Biodiversitätsforschung](#)
- [Empfehlungen zu datentechnischen Standards und Tools bei der Erhebung von Sprachkorpora \(Handreichung der Linguistik\)](#)
- [Der Umgang mit Forschungsdaten im Fach Psychologie: Konkretisierung der DFG-Leitlinien \(Deutsche Gesellschaft für Psychologie\)](#)

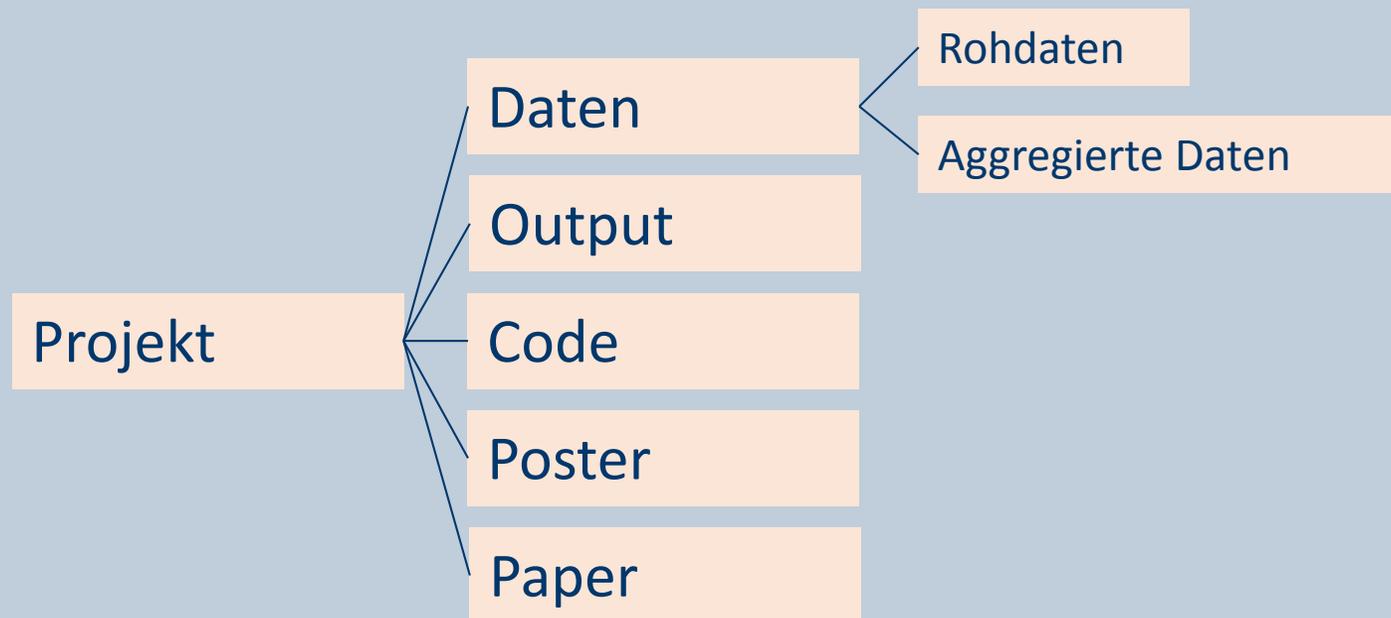
Forschungsdaten-Lebenszyklus



Quelle: UK Data Archive: Research Data Lifecycle.
<http://www.data-archive.ac.uk/create-manage/life-cycle>

Datensammlung und -verarbeitung

- Klare Ordnerstruktur → Bedeutung sollte klar ersichtlich sein; auch für andere Forschende
- Hierarchische Strukturen erleichtern das Auffinden:



Dateibenennung

- Konventionen zur Dateibenennung
 - Dokumentieren Sie alle Namenskonventionen oder genutzten Abkürzungen
 - z. B. [Projekt]_[Interview]_[Ort]_[Personen-ID]_[YYYYMMDD].mp4
 - Nutzen Sie Datums-/Zeitstempel oder eine separate ID (v1.0.0) für jede Version
 - Das Datum sollte zu Beginn oder am Ende des Dateinamens stehen, um die Sortierung zu erleichtern
 - Vermeiden Sie Sonderzeichen { } [] < > () * % # ` ; " , : ? ! & @ \$ ~
 - Nutzen Sie den Unterstrich (_) oder Großschreibung des ersten Buchstabens um Namen zu separieren
- Kontrolle der Dateiversion
 - Legen Sie obsolete Dateiversionen nach einem Backup separat ab
 - Hilfreich ist auch Versionskontroll-Software (z.B. Git, Subversion)



Sicherung der Daten

- Einrichtung von Backups
 - Mindestens 3 Kopien einer Datei
 - Auf mindestens 2 unterschiedlichen Medien
 - Wovon mindestens eins dezentral ist
- Testen Sie die Datenwiederherstellung zu Beginn sowie in regelmäßigen Abständen
- Schützen Sie Ihre (sensiblen) Daten
 - Hardware (bspw. separater abschließbarer Raum)
 - Dateiverschlüsselung
 - Sicherheit der Passwörter
 - Mindestens zwei Personen sollten Zugang zu Ihren Daten haben

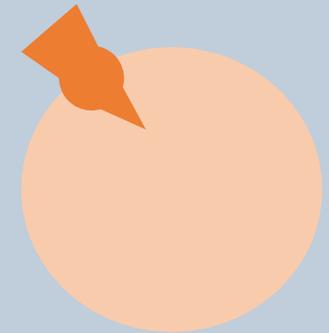
Übung: Welche Dateiformate nutzen Sie?



Bitte notieren Sie alle Dateiformate, die Sie im Laufe eines Forschungsprojekts oder in Ihrer täglichen Arbeit nutzen (1 Post-it pro Dateiformat).

Zeit: 5 Minuten

Anschließend Sammlung der Ergebnisse



Dateiformate

Für die Langzeitarchivierung sollten die Dateien:

- Unverschlüsselt, nicht komprimiert, nicht firmeneigen/patentiert sein und einen offenen, dokumentierten Standard nutzen

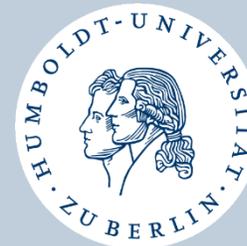
Dateityp	Empfehlung	Zu vermeiden
Tabellendaten	CSV, TSV, SPSS portable	Excel
Text	TXT, ODT, HTML, RTF; PDF/A nur wenn Layout wichtig ist	Word, PowerPoint
Multimedia	Container: MP4, Ogg Codec: Theora, Dirac, FLAC	QuickTime, H264
Bild	TIFF, JPEG2000, PNG	GIF, JPG
Strukturierte Daten	XML, RDF, JSON	RDBMS

Dateien zusätzlich zum Originalsoftwareformat auch in einem empfohlenen Format ablegen!



Langzeitarchivierung

- Langzeitarchivierung bedeutet für die Ewigkeit (> 10 Jahre)
- Speichern Sie Ihre Dateien spätestens zum Ende des Projekts auch in einem archivierbaren Dateiformat ab
- Langzeitarchivierung kostet in der Regel Ressourcen (Geld, Personal) und ist mit zusätzlichem Aufwand verbunden (Dokumentation, Metadaten)
- Nicht alle Dateien müssen langzeitarchiviert werden! Wählen Sie Meilenstein-Dateien oder Dateien, die einer Publikation zugrunde liegen



Exkurs: Personenbezogene Daten

Was ist hier zusätzlich zu beachten?

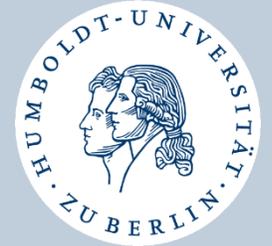


Datenschutz/personenbezogene Daten

- Erfüllung von rechtlichen Datenschutzvorgaben
- Schutz der Identität, vor allem bei sensiblen/vertraulichen Themen
- Schutz der Teilnehmer vor kommerzieller Nutzung der Angaben
- Ethisch-moralische Verpflichtung
- Eine Archivierung personenbezogener Daten ist nur mit Einwilligung der Teilnehmer und Teilnehmerinnen möglich!



Foto: Creative Commons CC0



Informierte Einwilligung

- Teilnehmer und Teilnehmerinnen müssen darüber informiert werden, was mit ihren Daten passiert
- Sie müssen einer Teilnahme auf Basis dieser Informationen zustimmen
- Datenanalyse und -archivierung benötigen separate Einwilligungen
- Vorlagen und Checklisten bietet unter anderem <https://www.forschungsdaten-bildung.de/einwilligung>
- Häufiges Problem bei der Archivierung:
 - Keine Einwilligung vorhanden
 - strikte Formulierungen in Einwilligungserklärungen

(Über-)Anonymisierung

- Entfernen aller identifizierenden Informationen/Details
- Ersetzen sensibler Informationen mit einer Beschreibung, die Bezug zu originärem Kontext hat
- Anonymität vor Vollständigkeit/Information!
- Alternativ: Zugriffskontrolle
→ So offen wie möglich und so geschlossen wie nötig



Foto: Creative Commons CC0



Wege des Datenzugangs

- **Offen:** frei zugänglich für alle (*Public Use File*)
- **Eingeschränkt:**
 - Nur für angemeldete Nutzer (auch mittels *Ferndatenzugriff*)
 - Nur für die Wissenschaft (*Campus Files* oder *Scientific Use File*) oder auf Anfrage (Prüfung erfolgt durch WissenschaftlerIn)
- **Geschlossen:** Zugriff nur vor Ort an sicherem Rechner (*On-Site Use* bzw. *Gastwissenschaftlerplätze*)

Falls eine Überanonymisierung droht, sollte man eine Einschränkung des Zugriffs in Erwägung ziehen.



Eingeschränkter Zugriff

Project on Human Development in Chicago Neighborhoods (PHDCN): Master File, Wave 2, 1997-2000 (ICPSR 13608)

Alternate Title: [PHDCN Master, 1997-2000](#)

Principal Investigator(s): [Earls, Felton J.](#), Harvard Medical School; [Brooks-Gunn, Jeanne](#), Scientific Director. Columbia University. Teacher's College. Center for the Study of Children and Families; [Raudenbush, Stephen W.](#), Scientific Director. University of Michigan. School of Education and Survey Research Center; [Sampson, Robert J.](#), Scientific Director. Harvard University. Department of Sociology

Summary: [🔗](#)

The Project on Human Development in Chicago Neighborhoods (PHDCN) was a large-scale, interdisciplinary study of how families, schools, and neighborhoods affect child and adolescent development. One component of the PHDCN was the Longitudinal Cohort Study, which was a series of coordinated longitudinal studies that followed over 6,000 randomly selected children, adolescents, and young adults, and their primary caregivers over time to examine the changing circumstances of their lives, as well as the personal characteristics, that might lead them ... [\(more info\)](#)

Series: [Project on Human Development in Chicago Neighborhoods \(PHDCN\) Series](#)

Access Notes

- **One or more files in this data collection have special restrictions**; consult the [restrictions note](#) to learn more. You can [apply online for access to the restricted-use data](#). A login is required to apply.

A downloadable version of data for this study is available however, certain identifying information in the downloadable version may have been masked or edited to protect respondent privacy. Additional data not included in the downloadable version are available in a restricted version of this data collection. For more information about the differences between the downloadable data and the restricted data for this study, please refer to the codebook notes section of the PDF codebook. Users interested in obtaining restricted data must complete and sign a Restricted Data Use Agreement, describe the research project and data protection plan, and obtain IRB approval or notice of exemption for their research.

Any public-use data files in this collection are available for access by the general public. Access does not require affiliation with an ICPSR member institution.

Dataset(s)

Dataset

Documentation: [Codebook.pdf](#) [Questionnaire.pdf](#)

Download: No downloadable data files available.

Study Description

Citation [🔗](#)

Earls, Felton J., Jeanne Brooks-Gunn, Stephen W. Raudenbush, and Robert J. Sampson. Project on Human Development in Chicago Neighborhoods (PHDCN): Master File, Wave 2, 1997-2000. ICPSR13608-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2005-12-06. <http://doi.org/10.3886/ICPSR13608.v1>

Persistent URL: <http://doi.org/10.3886/ICPSR13608.v1>

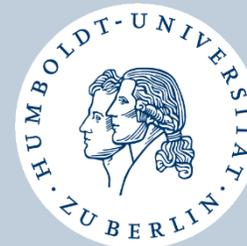
Export Citation:

- [RIS](#) (generic format for RefWorks, EndNote, etc.)
- [EndNote XML](#) (EndNote X4.0.1 or higher)



Zugang zu personenbezogenen Daten nur mit separatem Vertrag und Begründung

<http://doi.org/10.3886/ICPSR13608.v1>



Fragen?





Was sind Repositorien?

- Datenbanken, in denen Objekte archiviert, dokumentiert und publiziert werden können
 - Spezialisierte Repositorien für Texte, Forschungsdaten oder Filme
 - Disziplinspezifische vs. generische vs. institutionelle Repositorien
- Bekannte Datenrepositorien im deutschsprachigen Raum:
 - Pangaea (Geowissenschaften)
 - GESIS (Sozialwissenschaften)
 - PsychData (Psychologie)
 - Zenodo (fachübergreifend)

Registry of Research Data Repositories



Stand 02.06.2017:
1.864 verzeichnete
Forschungsdaten-
repositorien

The screenshot shows the re3data.org website interface. At the top, there is a search bar and navigation links for Search, Browse, Suggest, Resources, and Contact. A DataCite logo is also present. On the left side, there is a 'Filter' menu with various categories like Subjects, Content Types, Countries, etc. The main content area displays search results for 'C. Elegans Gene Expression' and 'BRENDA'. Each result includes a title, subject tags, content type, country, and a brief description. The 'C. Elegans Gene Expression' result is for a dataset from Canada, and the 'BRENDA' result is for a database from Germany.

<http://www.re3data.org>



Wie wählt man ein Repository aus?



Hat das Repository ein Zertifikat (z. B. Data Seal of Approval)?



Werden persistente Identifikatoren vergeben (z. B. DOI, handle, URN)?



Wie ist der Zugang zu den Daten (offen, eingeschränkt, unzugänglich)?



Werden Nutzungs- und Lizenzbedingungen der Daten durch das Repository genannt?

Institutionelles Repositorium



Technische Universität Berlin

Home Search Browse Publish

DepositOnce > Home

Deposit Once

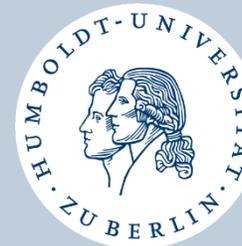
Repository for Research Data and Publications

Search DepositOnce

© 2017 Servicezentrum Forschungsdaten und Publikationen | About DepositOnce | Policy | Contact | Imprint | Privacy Statement

<https://depositonce.tu-berlin.de>





Fachspezifisches Repositorium

NOMAD

The NOMAD Laboratory A European Centre of Excellence

HOME PROJECT INDUSTRY OUTREACH TEAM EXTERNALS NEWS PRESS KIT CONTACT US

REPO HOME REPO TEAM WHY SHARING? DOIs TERMS FAQ **UPLOAD YOUR FILES** **SEARCH AND DOWNLOAD** REPO CONTACT

NOMAD REPOSITORY

The NOMAD Repository was established to host, organize, and share materials data.

The NOMAD (Novel Materials Discovery) Repository was established to host, organize, and share materials data.

NOMAD copes with the increasing demand and requirement of storing scientific data and making them available for longer periods. This rule of good scientific practice is set by many funding agencies worldwide. **NOMAD keeps scientific data for at least 10 years for free.** NOMAD also facilitates research groups to share and exchange their results, inside a single group or between two or more, and to recall what was actually done some years ago.

The NOMAD Repository enables the confirmatory analysis of materials data, their reuse, and repurposing. **NOMAD makes scientific data citable** as one can request digital objective identifiers (DOI's).

Have a look at [youtube](#) to see our [movies on the concept](#) and a [basic tutorial](#).

Upload of data is possible without any barrier. Results are accepted in their raw format as produced by the underlying code. The only condition is that the

Repository News

Currently, the NOMAD Repository contains **3,407,035** entries.

New codes supported: **BigDFT, CP2K, CPMD, DMol³, Elk, FLEUR, GPAW, MOLCAS, NWChem, octopus, ONETEP, ORCA, SIESTA, and TURBOMOLE**

Upload from *MedeA* possible ... [more](#)

<http://nomad-repository.eu>



Metadaten für Forschungsdaten

Eine Übersicht zu (disziplinspezifischen) Metadatenstandards finden Sie unter <http://www.dcc.ac.uk/resources/metadata-standards> oder <http://rd-alliance.github.io/metadata-directory/subjects>

- Beispiele:
 - Dublin Core
 - ISO 19115 (Geowissenschaften)
 - Darwin Core (Biodiversität)
 - Data Documentation Initiative (Sozialwissenschaften)



Beispiel für Metadaten

Titel	Der Einfluss der Religion auf die Integration von Flüchtlingen aus Ex-Jugoslawien
URI	http://dx.doi.org/10.7802/72
Primärforscher	Bradaric, Lejla; Universität zu Köln, Institut für Soziologie und Sozialpsychologie
Publikationsjahr	2014
Verfügbarkeit	Freier Zugang (mit Registrierung)
Fachgebiet	Migration Religionssoziologie
Kategorie	Religion und Weltanschauung Person, Persönlichkeit, Rolle
Inhalt	In der qualitativ ausgerichteten explorativen Studie ist am Beispiel von Flüchtlingen aus Ex-Jugoslawien (Orthodoxe, Katholiken, Muslime), die zum Erhebungszeitraum bereits seit 13-19 Jahren in Deutschland gelebt haben, die Integration in Deutschland in Anlehnung an die Integrationsdimensionen "kognitiv, strukturell, sozial und identifikativ" (Integrationsmodell von Hartmut Esser) untersucht worden. Es sind Befragte mit möglichst vielen gleichen demographischen Merkmalsausprägungen wie Geschlecht, Alter, Einwanderungszeitraum, herkunftsspezifischer Kulturraum und Wohnort in Deutschland ausgewählt worden, um eine Vergleichbarkeit der Antworten bei der Stichprobe (n=26) zu ermöglichen. Die Religionszugehörigkeit ist dahingegen unterschiedlich, um somit den Einfluss der Religion auf den Integrationsverlauf genauer zu beleuchten. Die Befragten konnten auswählen, in welcher Sprache das Interview mit Ihnen durchgeführt werden soll. Der zweisprachige Fragebogen ist den vier Integrationsphasen entsprechend mittels 96 Fragen so konzipiert worden, dass sich die Integration der Befragten in die unterschiedlichen Dimensionen (kognitiv, strukturell, sozial und identifikativ) messen lässt. Neben der Abfrage demographischer Daten und dem Integrationsverlauf beinhaltet die Untersuchung zudem Fragen zu Ressourcen wie z.B. dem Bildungsstand der Eltern oder Aufenthaltsstatus. Im anschließenden Untersuchungsschritt sind die Befragten anhand der ausgewerteten Merkmalskombinationen folgenden unterschiedlichen Integrationstypen zugeordnet worden: 1. Der Assimilierte, 2. Der Binationale, 3. Der Integrierte und mit Herkunft identifizierende, 4. Der Segmentierte, 5. Der Marginalisierte, 6. Der Kosmopolit, 7. Der Sonderfall.
Untersuchungsgebiet	Aachen / DE-NW-AAH
Grundgesamtheit	26 Befragte, männlich, zw. 1975 bis 1981 geboren, im Alter von 11 bis 17 in den Jahren 1990 bis 1996 als Kriegsflüchtlinge aus Ex-Jugoslawien nach Deutschland eingewandert
Auswahlverfahren	Durch die Bosnische, Kroatische und Serbische Migrantenorganisation in Aachen Zugang zu Befragten gefunden; Befragte anhand der gewünschten demographischen Merkmale gewählt
Erhebungsverfahren	PAPI (Papierfragebogen) - Persönliches Interview
Erhebungszeitraum	2009-02; 2009-06
Lizenzen	CC BY-SA 4.0

Quelle: Bradaric, Lejla (2014): Der Einfluss der Religion auf die Integration von Flüchtlingen aus Ex-Jugoslawien. Version: 1. GESIS Datenarchiv. Datensatz. <http://doi.org/10.7802/72>



Schlagwörter vergeben

- Thesauri und Klassifikationen sind Dokumentationssprachen, die zur inhaltlichen Beschreibung von Objekten (z. B. Forschungsdaten) verwendet werden
- Die Auffindbarkeit der Daten wird dadurch erleichtert
- Für sehr viele Fachbereiche gibt es bereits eigene, spezialisierte Klassifikationen und Thesauri
- **Übersicht:** Basel Register of Thesauri, Ontologies & Classifications
<http://www.bartoc.org/>

Creative Commons (CC) Lizenzen



CC0 (Gemeinfreiheit/Public Domain)



CC BY (Namensnennung)



CC BY-ND (Namensnennung - Keine Bearbeitung)



CC BY-NC (Namensnennung - Nicht kommerziell)



CC BY-SA (Namensnennung - Weitergabe unter gleichen Bedingungen)



CC BY-NC-SA (Namensnennung - Nicht-kommerziell - Weitergabe unter gleichen Bedingungen)



CC BY-NC-ND (Namensnennung - Nicht-kommerziell - Keine Bearbeitung)

Weitere Informationen: <http://forschungslizenzen.de>

Quelle: <https://creativecommons.org/licenses/>

Übung: Lizenzen

Welche Lizenz lässt sich kombinieren?



Zeit: 5 Minuten

Quelle: <https://creativecommons.org/licenses/>

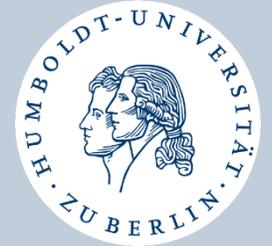
Übung: Lizenzen

Welche Lizenz lässt sich kombinieren?



Zeit: 5 Minuten

Quelle: <https://creativecommons.org/licenses/>



10 Dinge, die man über ORCID wissen sollte

1. Steht für Open Researcher and Contributor ID
2. (alpha-)numerischer 16-stelliger Code
3. Eindeutige wissenschaftliche Identität (auch über Namenswechsel oder Tippfehler hinweg)
4. Wird von Zeitschriften, Forschungsförderern und Institutionen als Normdatei genutzt
5. Wird von WissenschaftlerIn gepflegt
6. Hält länger als eine E-Mail-Adresse
7. ORCID-Erstellung dauert ca. 30 Sekunden
8. Wird von gemeinnütziger Initiative betrieben
9. Wächst kontinuierlich (02.06.17: 3.448.951 ORCIDs)
10. Verbindung zu Web of Science, zenodo, DataCite u.a.

Verlage und Forschungsdaten

- Science
- Springer Nature
- Elsevier
- Public Library of Science (PLOS)
- ...



Datenpublikation im Rahmen eines Zeitschriftenartikels



- Pilotprojekt von *Zeitschrift für Soziologie* sowie *Soziale Welt*
- Autorinnen und Autoren verpflichten sich zur öffentlichen Zugänglichmachung ihrer für den Artikel genutzten Daten und Analyseskripte
- Ziel: Möglichkeit der Replikation
- Verfügbarmachung über Datenrepositorium

Artikel: <https://doi.org/10.1515/zfsoz-2015-1025> und zugehörige Datenpublikation (rechts): <http://doi.org/10.7802/1350>

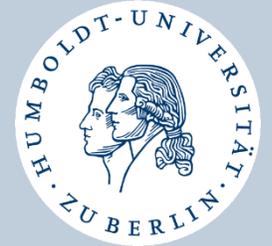
Titel	Code/Syntax: Wer profitiert von den Korrekturmöglichkeiten in der Sekundarstufe? Der Einfluss von Bildungsherkunft und Migrationshintergrund im Bildungsverlauf
URI	http://dx.doi.org/10.7802/1350
Primärforscher	Kurz, Karin; Institut für Soziologie der Universität Göttingen Böhner-Taute, Eileen; Institut für Soziologie der Universität Göttingen
Publikationsjahr	2016
Verfügbarkeit	Eingeschränkter Zugang
Name der Zeitschrift	Zeitschrift für Soziologie
Jahrgang	45
Heftnummer	6
Fachgebiet	Sozialwissenschaften Forschung
Kategorie	Erziehung, Schulwesen
Inhalt	Sekundäranalyse der Daten des Sozio-oekonomischen Panels (SOEP). Im vorliegenden Beitrag werden nachträgliche Wechsel von Jugendlichen zu Schulformen, die die Hochschulreife ermöglichen, mittels SOEP-Daten untersucht. Dabei werden auch mögliche Unterschiede zwischen den Bundesländern Bayern, Baden-Württemberg, Niedersachsen und Nordrhein-Westfalen, die unterschiedlich offene Schulstrukturen aufweisen, geprüft. Die zentrale Frage ist, ob solche Aufstiegsprozesse zu einer Veränderung von Bildungsungleichheiten, die zu Beginn der Sekundarstufe I bestehen, beitragen und ob es darin Unterschiede zwischen den untersuchten Bundesländern gibt. Zur Untersuchung der Bildungsungleichheiten wird zwischen einer konditionalen und un konditionalen Analyseperspektive unterschieden.
Untersuchungsgebiet	Deutschland / DE
Grundgesamtheit	Die Grundgesamtheit bildet die wohnberechtigte Bevölkerung in der Bundesrepublik Deutschland einschließlich West-Berlin ab dem Jahr 1984 und die deutsche Wohnbevölkerung in der DDR seit Juni 1990.
Auswahlverfahren	Retrospektive Befragung
Erhebungsverfahren	Eigenständig auszufüllender Fragebogen
Erhebungszeitraum	2000; 2011
Lizenzen	CC BY-NC-SA 4.0
Quelle	Sozio-oekonomisches Panel (SOEP)
Referenzpublikation	Kurz, Karin und Eileen Böhner-Taute, 2016: Wer profitiert von den Korrekturmöglichkeiten in der Sekundarstufe? Der Einfluss von Bildungsherkunft und Migrationshintergrund im Bildungsverlauf. Zeitschrift für Soziologie. 45 (6). DOI 10.1515/zfsoz-2015-1025

Dateien der Ressource

Dateien	Größe	Format	Download-Option
datenaufbereitung_la.do	16.59Kb	text/x-stata-syntax	Bitte geben Sie den Verwendungszweck an <input type="text"/>

[Herunterladen](#)





Data Journals

Ein Data Paper dokumentiert und beschreibt Forschungsdaten, um die Verbreitung und Nachnutzung zu erleichtern. Es informiert über die Datenerhebung, Charakteristiken und Funktionen sowie potenzielle Nachnutzungsmöglichkeiten.

- Beispiele:
 - Scientific Data <http://www.nature.com/sdata>
 - Data in Brief <http://www.journals.elsevier.com/data-in-brief>
 - Data <http://www.mdpi.com/journal/data>
 - Data Science Journal <http://datascience.codata.org>
 - ...

Übersicht: <https://www.cms.hu-berlin.de/de/dl/dataman/teilen/dokumentation/datajournal>

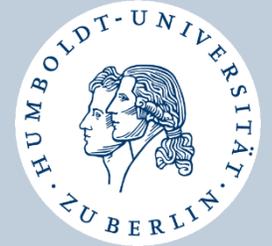




Forschungsdaten finden

- B2FIND Datensuche: <http://b2find.eudat.eu>
- Creative Commons CC Search: <https://search.creativecommons.org>
- DataCite Metadata Search: <https://search.datacite.org>
- DataONE Datenkatalog: <https://search.dataone.org/#data>
- DataSearch von Elsevier: <https://datasearch.elsevier.com>
- DataSearch von GESIS: <http://datasearch.gesis.org>
- Research Data Discovery Service von JISC:
<http://ckan.data.alpha.jisc.ac.uk/dataset>





DataSearch von Elsevier

DataSearch Beta

Search for research data across domains and types, from many domain-specific, cross-domain and institutional data repositories.

Find research data

Or Try: cosmic microwave radiation, frog phylogeny or protein structure prediction

G+ f t in ✉

DataSearch is in development. We welcome your feedback and ideas to help us improve.

[Give Feedback](#)

ELSEVIER [FAQ](#) [About](#) [Terms of Use](#) [Privacy Policy](#) Copyright © 2016 Elsevier B.V.

Sources

- ScienceDirect (4447335)
- arXiv (564619)
- PubMed Central (105029)
- PetDB (84194)
- NeuroElectro (22245)
- Dryad (14627)
- Harvard Dataverse (10147)
- ICPSR (9424)
- ThermoML at NIST TRC (7777)
- Mendeley Data (71)

Stand 21.04.2017:
über 5 Millionen
Einträge

<https://datasearch.elsevier.com>



Zusammenfassung Forschungsdaten managen



Lebenszyklus umschreibt den Prozess von der Datenentstehung bis hin zur Veröffentlichung

Finden Sie die **Balance** zwischen Perfektion und Realisierbarkeit

Nutzen Sie die verschiedene **Beratung**möglichkeiten Ihrer Einrichtung!

Datemanagement bietet

- ✓ Vertrauenswürdige Daten
- ✓ Reproduzierbare Daten
- ✓ Umgang mit großen Datenmengen
- ✓ Erzeugung von langlebigen Daten



Hilfreiche Links

Wiki Forschungsdaten: <http://www.forschungsdaten.org>

Informationswebseite Forschungsdaten aus Baden-Württemberg:
<https://www.forschungsdaten.info>

E-Learning Module: <http://www.researchdatamanagement.ch>

MANTRA Online Course: <http://datalib.edina.ac.uk/mantra/>

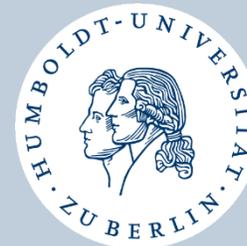
Digital Curation Centre: <http://www.dcc.ac.uk>

Helmholtz Open Science Webinare: <http://os.helmholtz.de/bewusstseinschaerfen/workshops/webinare-zu-forschungsdaten/>



Kaffeepause bis 15:45





Effektive Datenmanagementplanung

Was gibt es zu beachten?



Was ist ein Datenmanagementplan?

- Alle Informationen, die die Sammlung, Aufbereitung, Speicherung, Archivierung und Veröffentlichung von Forschungsdaten im Rahmen eines Forschungsprojekts hinreichend beschreiben und dokumentieren
- „[...] Analyse des Workflows von der Erzeugung der Daten bis zu deren Nutzung“*
- Umfang kann zwischen wenigen Absätzen und mehreren Seiten variieren

* J. Ludwig, H. Enke (Hrsg.): Leitfaden zum Forschungsdaten-Management. Handreichungen aus dem WissGrid-Projekt. Verlag Werner Hülsbusch: Glückstadt, 2013.

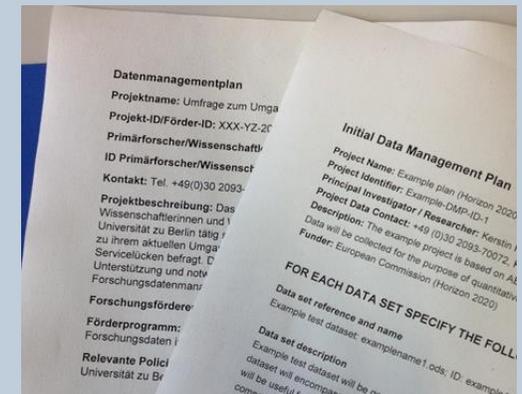
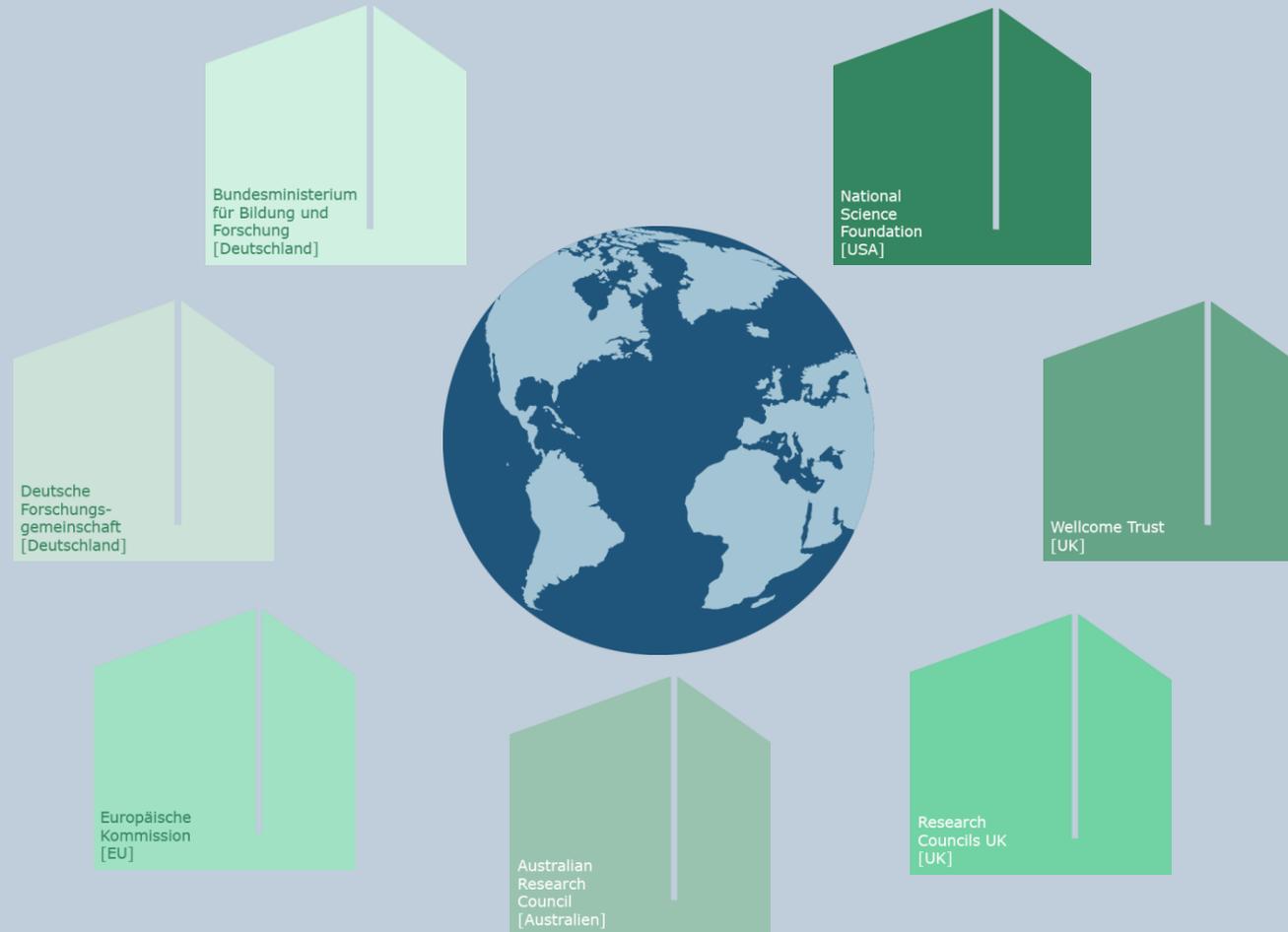
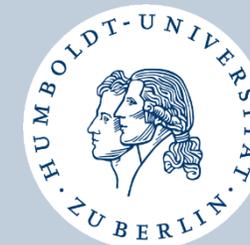


Foto: Creative Commons CC0

Wer verlangt einen Datenmanagementplan?

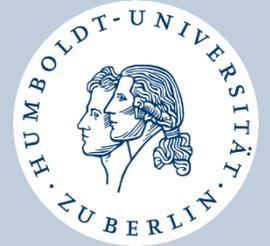


Förderanforderungen



Förderer	Plan gefordert?	Abgabe bei Antrag?	Inhalt	Updates?
Horizon 2020 (ab 2017)	Datenmanagementplan	Nein, erster Plan innerhalb der ersten 6 Projektmonate	Inhalte des Horizon 2020 Template	Update, falls signifikante Änderungen auftreten sowie zum Projektende
Deutsche Forschungsgemeinschaft	Angaben zum Umgang mit Forschungsdaten	Ja	Inhalte der Leitlinie zum Umgang mit Forschungsdaten	Nein
Bundesministerium für Bildung- und Forschung (BMBF)	Plan manchmal erforderlich, abhängig vom Programm	Falls notwendig, ja	Inhalt hängt vom jeweiligen Programm ab	Nein
BMBF Bildungsforschung	Datenmanagementplan	Ja	Inhalte der Checkliste	Ja

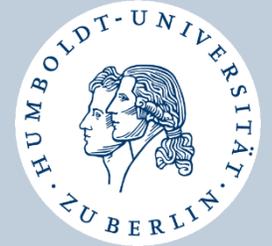




Horizon 2020

- Ab 2017 alle Programme (zuvor 7 Pilot-Programme)
- Jeder Zeit Möglichkeit des “Opt-Out”, jedoch sollten auch diese Projekte einen DMP auf freiwilliger Basis abgeben
 - Mögliche Gründe für Opt-Out: geistige Eigentumsrechte, Datenschutz, Gefährdung des Projektziels
- Data Pilot bezieht sich primär auf Forschungsdaten, die Grundlage einer Publikation sind
- Andere Daten können auf freiwilliger Basis ebenfalls geteilt werden
- Anfallende Kosten für die Zugänglichmachung der Daten können mit beantragt werden

Förderanforderung Europäische Kommission (Horizon 2020)



- Einen Datenmanagementplan gemäß der bereitgestellten Vorlage einzureichen: 1. Version innerhalb der ersten 6 Projektmonate, eine aktualisierte Version in der Mitte der Projektlaufzeit und eine letzte Version in der Abschlussphase des Projekts
- Die entstehenden Forschungsdaten in geeigneten institutionellen, projekt- oder fachspezifischen Repositorien öffentlich zugänglich zu machen
- Die breite, möglichst uneingeschränkte Nachnutzung der Forschungsdaten zu unterstützen, indem offene Lizenzen (z. B. CC BY, CC0) angewendet werden
- Zusätzliche Kontextinformationen zu verwendeten Werkzeugen und Instrumenten wie Software-Code, Algorithmen oder Analyseprotokolle zur Verfügung zu stellen

Quelle: European Commission: H2020 Programme. Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020. Version 3.2, 21 March 2017. Online verfügbar: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf





FAIR Guiding Principles

- **F**indable
Forschungsdaten sollten auffindbar sein.
- **A**ccessible
Forschungsdaten sollten zugänglich sein.
- **I**nteroperable
Forschungsdaten sollten interoperabel sein.
- **R**e-usable
Forschungsdaten sollten nachnutzbar sein.

FAIR Guiding Principles:
<https://www.force11.org/group/fairgroup/fairprinciples>



Förderanforderung Deutsche Forschungsgemeinschaft (DFG)



Die DFG erwartet die Einhaltung der Regeln guter wissenschaftlicher Praxis und:

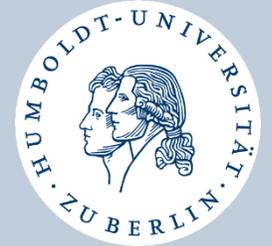
"Werden in dem geplanten Projekt systematisch Forschungsdaten erhoben, die für eine Nachnutzung geeignet sind, sollte der Antrag ein entsprechendes Konzept für die Überführung der Forschungsdaten in vorhandene Datenbanken oder Repositorien enthalten. Häufig ist es in diesem Fall sinnvoll, bereits in der Planungsphase Kontakt zu den Ansprechpersonen der entsprechenden Infrastrukturen aufzunehmen, um existierende Standards nutzen zu können oder um gegebenenfalls anfallende Kosten für diese Schritte in den Antrag integrieren zu können."

Die gegebenenfalls zusätzlich anfallenden Kosten sollten folglich im Antrag benannt werden.

Quelle: Deutsche Forschungsgemeinschaft: Förderbedingungen der DFG. Einen Antrag stellen. Online verfügbar: http://www.dfg.de/foerderung/antragstellung_begutachtung_entscheidung/antragstellende/antragstellung/index.html



DFG Leitlinien zum Umgang mit Forschungsdaten



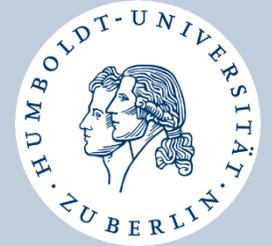
Antrag sollte enthalten:

- welche Forschungsdaten im Verlauf eines wissenschaftlichen Forschungsvorhabens entstehen, erzeugt oder ausgewertet werden
- fachspezifisch angemessene Konzepte und Überlegungen für die Qualitätssicherung, für den Umgang mit und die langfristige Sicherung der Forschungsdaten
- Informationen zu Datentypen, falls vorhanden zu disziplinspezifischen Standards und zur Wahl geeigneter Repositorien, sofern diese vorhanden sind
- Angaben zu ggf. betroffenen Rechten Dritter sowie erste Planungen zum zeitlichen Rahmen der Datenveröffentlichung

Quelle: Deutsche Forschungsgemeinschaft: Leitlinien zum Umgang mit Forschungsdaten. 2015. Online verfügbar: http://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten.pdf



Anforderungen des Bundesministeriums für Bildung und Forschung (BMBF)



Aktuelle Bekanntmachungen beinhalten:

„Der Zugang zu den wissenschaftlichen Erkenntnissen und Daten ist eine wesentliche Grundlage für Forschung, Entwicklung und Innovation. Die langfristige Sicherung und Bereitstellung der Forschungsdaten leistet einen Beitrag zur Nachvollziehbarkeit und Qualität wissenschaftlicher Arbeiten. Deshalb sollen Forschungsergebnisse, die im Rahmen dieser Förderrichtlinie entstehen, als Open-Access-Veröffentlichung publiziert (siehe auch Nummer 6) und Forschungsdaten (digital; unter Wahrung der Rechte Dritter, insbesondere Datenschutz, Urheberrecht) zur Nachnutzung bereitgestellt werden.“

Quelle: Bundesministerium für Bildung und Forschung: Bekanntmachung. Richtlinie zur Förderung transnationaler Forschungsprojekte zur Entwicklung von innovativen Therapieverfahren für seltene Erkrankungen innerhalb des ERA-NET "E-Rare". Bundesanzeiger vom 17.01.2017. Online verfügbar: <https://www.bmbf.de/foerderungen/bekanntmachung-1299.html>



BMBF: Konzept zur Qualitätssicherung, zum Umgang und zur langfristigen Sicherung von Forschungsdaten



Konzept sollte enthalten:

- Datentypen
- Standards
- gewähltes Repository oder sonstige Form der langfristigen Sicherung (Wie Zugang? Wann? Welche Verarbeitungsstufe?)
- Archivierung der Forschungsdaten für mindestens 10 Jahre in der eigenen Einrichtung oder einer fachlich einschlägigen, überregionalen Infrastruktur

Kosten der Datenaufbereitung sollten ebenfalls genannt werden.

Quelle: Deutsches Luft- und Raumfahrtzentrum, Bundesministerium für Bildung und Forschung: Leitfaden für die Erstellung von Projektskizzen zur „Richtlinie zur Förderung von Forschungsverbänden zur Kinder- und Jugendgesundheit“ Im Rahmen der „Förderinitiative Gesund – ein Leben lang“ des Bundesministeriums für Bildung und Forschung. 2015. Online verfügbar: http://www.dlr.de/pt/Portaldata/45/Resources/Dokumente/Leitfaden/Leitfaden_KiJu_Skizze_final.pdf



Förderanforderung des BMBF im Bereich Bildungsforschung

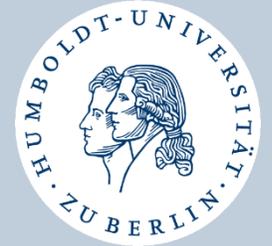


„Die Antragsteller/-innen verpflichten sich, die im Rahmen des Projektes gewonnenen Daten nach Abschluss des Projekts in weitergabefähiger Form einer geeigneten Einrichtung (z. B. GESIS - Leibniz-Institut für Sozialwissenschaften oder einem Forschungsdatenzentrum) zur Verfügung zu stellen mit dem Ziel, langfristige Datensicherung, Sekundärauswertungen oder eine Nachnutzung zu ermöglichen. Dort werden die Daten archiviert, dokumentiert und auf Anfrage der wissenschaftlichen Community zur Verfügung gestellt.“

„Beantragt werden können Mittel für [...] das Management der selbst generierten Daten, für ggf. anfallende Gebühren zur Nutzung von Sekundärdaten, für Investitionen und weitere Positionen.“

Quelle: Bundesministerium für Bildung und Forschung: Bekanntmachung des Bundesministeriums für Bildung und Forschung von Richtlinien zur Förderung von Forschung im Bereich „Sprachliche Bildung und Mehrsprachigkeit“. 2012. Online verfügbar: <http://www.bmbf.de/foerderungen/20319.php>





Weitere Informationen und Templates

Horizon 2020

European Commission (2016): Horizon 2020 Programme. Guidelines on FAIR Data Management in Horizon 2020. Version 3.0, 26 July 2016.

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

Aktualisierungsdatum: 25.10.2016.

DFG

Deutsche Forschungsgemeinschaft (2015): Leitlinien zum Umgang mit Forschungsdaten. Online verfügbar:

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

Aktualisierungsdatum: 25.10.2016.

BMBF

Verbund Forschungsdaten Bildung (2015): Checkliste zur Erstellung eines Datenmanagementplans für die Bildungsforschung, 1.0. Verfügbar unter:

<http://www.forschungsdaten-bildung.de/dmp>

Aktualisierungsdatum: 25.10.2016.



Datenmanagementplan (DMP)

Ein DMP besteht unter anderem aus:

- Administrativen Informationen (Projektname, Datenurheber, weitere Mitwirkende, Kontakt, Förderprogramm usw.)
- Projekt- und Datensatzbeschreibung
- Angaben zu Metadaten und Standards
- Daten teilen
- Archivierung und Sicherung der Daten
- Verantwortlichkeiten
- Kosten

Wichtig: DMPs sollten kurz und knapp gehalten werden.
So informativ wie möglich, so detailliert wie nötig!

Spezifika der Forschungsförderer



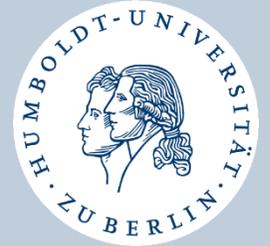
Anforderung/ Förderer	Horizon 2020 Data Pilot	DFG	BMBF*
Administrative Informationen	✓	✓	✓
Data(satz)beschreibung	✓	✓	✓
Metadaten/Standards	✓	✓	✓
Daten teilen	CC-BY/CC0, falls rechtlich möglich	✓	✓
Speicherung/Backup/ Archivierung	✓	✓	✓
Verantwortlichkeiten	✓	-	✓
Kosten	✓	✓	✓
Sonstiges	Datensicherheit; ethische Aspekte; relevante Policies/ Arbeitsabläufe; Aktualisierung	Informationen zur Qualitätssicherung; Zeitplan	* Betrifft nur bestimmte Programme





Administrative Informationen

- Projektname
- Datenurheber
- weitere Mitwirkende
- Kontaktinformationen
- Förderprogramm
- Evtl. Fördernummer/-id
- Abstract zum Forschungsvorhaben und dem Grund der Datensammlung (Projektbeschreibung)
- Angaben zu zugehörigen institutionellen oder fachspezifischen Policies



Datensatzbeschreibung

Angaben zu Forschungsdaten: Erhebung und Nachnutzung

- Welche Art von Daten werden erhoben? Wie ist die voraussichtliche Gesamtgröße?
- Gibt es bereits Daten, die nachgenutzt werden können?
- Was spricht gegen eine Nachnutzung vorhandener Daten?
- Welche Methoden, Geräte, Standards zur Datenerhebung werden eingesetzt?
- Ermöglichen die gewählten Dateiformate sowie die Software eine Langzeitverfügbarkeit und Nachnutzung durch andere?
- Wie ist die Verzeichnisstruktur und Bezeichnung der Ordner und Dateien geplant? Wie wird die Versionierung gehandhabt?
- Welche Maßnahmen zur Qualitätssicherung sind vorgesehen (bspw. Versuchswiederholung, Peer Review der Daten)?



Angaben zu Metadaten und Standards

Was wird in Zukunft benötigt um die Daten erneut zu analysieren und zu interpretieren?

- Metadaten
- Kontrollierte Vokabulare (speziell Schlagwortvergabe)
- Normdaten

Wichtig: Datenbeschreibung und -dokumentation benötigen Zeit – diese sollte eingeplant und durch personelle Ressourcen abgesichert sein (siehe Verantwortlichkeiten)

Verantwortlichkeiten



Projektleitung



Metadaten/Dokumentation



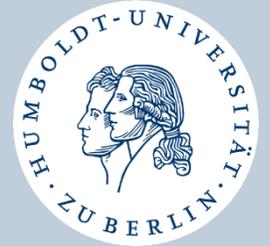
Erstellung/Update Datenmanagementplan



Backup

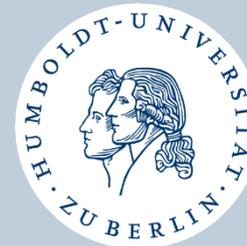


Koordination mit Projektpartnern/Externen



Kosten

- Einige Repositorien erwarten eine finanzielle Aufwandsentschädigung bei großen Datensätzen
- Manche Datenzentren verlangen Geld für die Bereitstellung von Daten, wenn man diese Nachnutzen möchte (bspw. für Zusendung einer DVD mit den Daten)
- Ebenso können Kosten bei der Langzeitarchivierung entstehen
- Diese Kosten können und sollten bei den Forschungsförderern mit beantragt werden



Fragen?





Praktische Hilfe

Tools zur Erstellung eines Datenmanagementplans:

- DMPonline
<https://dmponline.dcc.ac.uk>
 - DMPTool
<https://dmptool.org>
 - RDMO (Research Data Management Organiser)
<http://rdmorganiser.github.io>
- } DMP Roadmap

Anleitung und Musterpläne:

https://cms.hu-berlin.de/de/ueberblick/projekte/dataman/arbeiten/dmp_erstellen





DMP-Beispiele

Es gibt vermehrt öffentlich zugängliche Datenmanagementpläne:

- [DMPs im RIOjournal](#)
- [DMPs auf Zenodo](#)
- [Beispiel-DMPs auf der Webseite des Digital Curation Centre \(DCC\)](#)
- [Öffentliche DMPs des DMPTool](#)



Blick in DMPonline

DMP ONLINE Home About Roadmap Help

Welcome.

DMPonline has been developed by the Digital Curation Centre to help you write data management plans.

Screencast on how to use DMPonline

Sign in

Email address *

Password *

[Forgot your password?](#)

Remember me

Sign in

[Or, sign in with your institutional credentials](#) (UK users only)

Sign up

New to DMPonline? Sign up today.

<https://dmponline.dcc.ac.uk>





Praktische Umsetzung

Bitte versuchen Sie nun für ein fiktives oder reales Projekt einen Datenmanagementplan zu erstellen.

Wählen Sie bspw. für den Fachbereich Ozeanographie/ Meeresbiologie mit einer Datenmenge von insgesamt 5 TB

- ein geeignetes Repository,
- welche Lizenz vergeben werden soll,
- passende Standards;
- informieren Sie sich über mögliche Kosten

Versuchen Sie so viele Informationen wie möglich anzugeben.



Ihre Daten: Tierbeobachtungen in der Antarktis (Excel-Tabellen)



Foto: „Antarctica“ von Andreas Kambanis
<https://www.flickr.com/photos/andikam/12437173814/in/album-72157640778048215/> lizenziert unter Creative Commons CC BY 2.0
<https://creativecommons.org/licenses/by/2.0/>



Foto: Buckelwal, CC0



Foto: Königspinguine, CC0



Beispieldaten

27	Size:	16922 data points														
28	¶															
29	Date/Time	Latitude	Longitude	Ice cov [%]	Floes [%]	New ice [%]	Old ice [%]	Floe size [m]	Distance [km]	Icebergs [#kr]	Aves [#km**2]	Penguins [#k]	Albatrosses [f]	Petrels [#km²]	Seals [#km**2]	Whales [#km**2]
78	1992-10-02T12	-567.037	-503.400						589							
79	1992-10-02T12	-567.157	-502.800						587							
80	1992-10-02T12	-567.275	-502.300						586							
81	1992-10-02T12	-567.392	-501.700						585							
82	1992-10-02T13	-567.505	-501.200	0					583	1.98	0.00	0.00	1.98			
83	1992-10-02T13	-567.642	-500.600	0					582	0.00	0.00	0.00	0.00			
84	1992-10-02T13	-567.761	-500.200	0					580							
85	1992-10-02T13	-567.688	-500.100						581						0.00	
86	1992-10-02T13	-567.695	-500.000						581						0.00	
87	1992-10-02T13	-567.693	-500.000						581							
88	1992-10-02T14	-567.673	-500.000						581							
89	1992-10-02T14	-567.673	-500.000						581							
90	1992-10-02T14	-567.673	-500.000						582							
91	1992-10-02T14	-567.673	-500.000						582							
92	1992-10-02T14	-567.673	-500.000						582							
93	1992-10-02T14	-567.673	-500.000						580							
94	1992-10-02T15	-567.673	-500.000						579							
95	1992-10-02T15	-567.673	-500.000	0					577	0.00	0.00	0.00	0.00			
96	1992-10-02T15	-567.673	-500.000	0					575	0.96	0.00	0.00	0.96			
97	1992-10-02T15	-567.673	-500.000	0					574	0.92	0.00	0.92				
98	1992-10-02T15	-567.673	-500.000	0					572	0.94	0.00	0.00	0.94	0.00		
99	1992-10-02T15	-567.673	-500.000	0					570						0.00	
100	1992-10-02T16	-567.673	-500.000						571							
101	1992-10-02T16	-567.673	-500.000						571						0.00	
102	1992-10-02T16	-567.673	-500.000						571							
103	1992-10-02T16	-567.673	-500.000						571							
104	1992-10-02T16	-568.653	-495.800						571							
105	1992-10-02T16	-568.642	-495.900						571							
106	1992-10-02T17	-568.645	-495.900						571							
107	1992-10-02T17	-568.616	-495.900						571							
108	1992-10-02T17	-568.621	-495.900						571							
109	1992-10-02T17	-568.624	-495.900						571							

Quelle: V. Smetacek, H.J. de Baar, U. Bathmann, K. Lochte, R. van der Loeff, M. Michiel: Ice and animal observations during POLARSTERN cruise ANT-X/6. PANGAEA. 1997. <http://doi.org/10.1594/PANGAEA.88838>



Zusammentragen/Auflösung

Tabelle mit CC-BY-Lizenz:

Smetacek, V et al. (1997): Ice and animal observations during POLARSTERN cruise ANT-X/6.
<http://doi.org/10.1594/PANGAEA.88838>

Metadatenstandard: Entsprechend den Vorgaben von Pangaea. Alternativ bspw. Darwin Core oder ISO 19115

Formatierung: TAB-delimited TEXT-Dateien (ASCII) als ZIP-Archive oder Excel-Format (Pangaea-Vorgabe)

Kosten: 300 € je Datenupload zu einer Publikation

Klassifikation: Aquatic Sciences and Fisheries and Oceanic Abstracts Classification Codes

- 1362 Ornithology – Geographical distribution
- 1372 Mammalogy – Geographical distribution
- 2144 Regional studies, expeditions and data reports
- 2150 Ice





Herzlichen Dank für Ihr Interesse!

Kerstin Helbig, Koordinatorin Forschungsdatenmanagement
kerstin.helbig@cms.hu-berlin.de

Weitere Informationen zum Thema
Forschungsdatenmanagement:
<https://hu.berlin/dataman>



Folgen Sie uns bei Twitter [@DatawomenHUB](https://twitter.com/DatawomenHUB)

