

Forschungsdatenmanagement

Hochschule für den öffentlichen Dienst in Bayern
Fachbereich Archiv- und Bibliothekswesen
Ausbildungsjahr 2022/2025

Dr. Yves Vincent Grossmann
grossmann@mpdl.mpg.de
Max Planck Digital Library

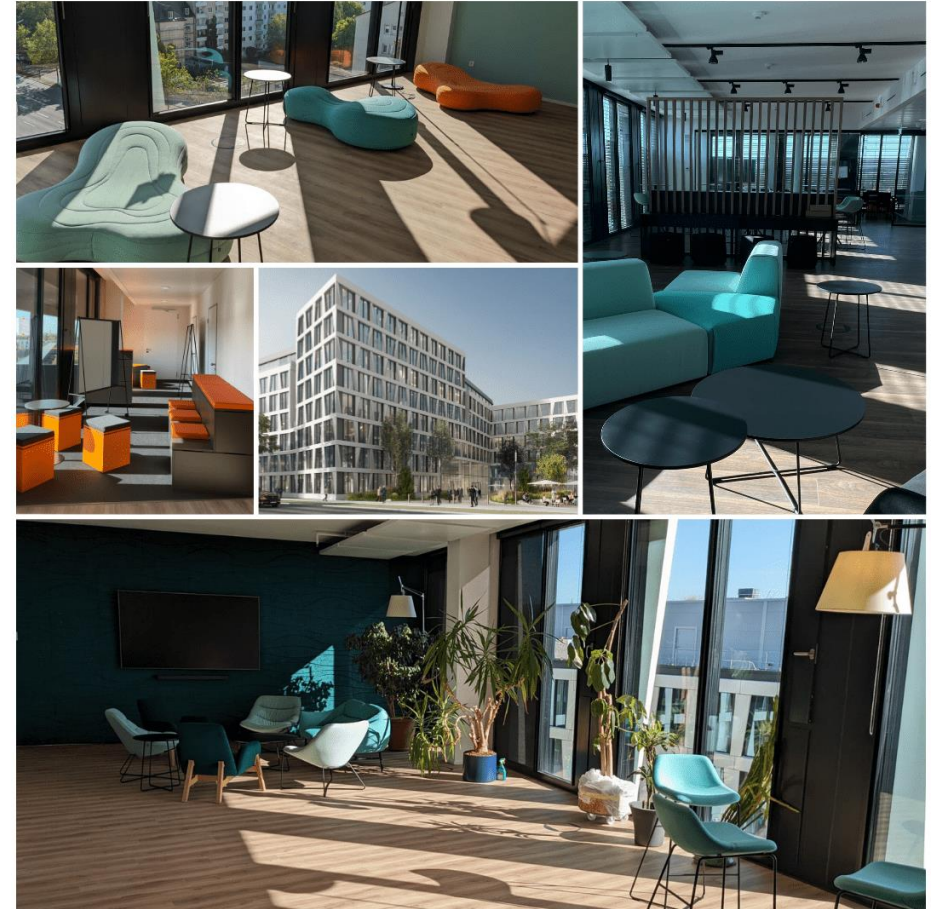
Willkommen

Dozent

- Yves Vincent Grossmann
- <https://orcid.org/0000-0002-2880-8947>
- Referent für Forschungsdatenmanagement 2020-2024 an der Max Planck Digital Library
- ab März 2024
 - Leitung der Bibliothek Naturwissenschaften mit den beiden Außenstellen Mathematik und Informatik sowie
 - Referent für Forschungsdatenmanagement mit dem Schwerpunkt Natur- und Lebenswissenschaften
 - an der Universitätsbibliothek Johann Christian Senckenberg der Johann Wolfgang Goethe-Universität Frankfurt am Main
- Kontakt: grossmann@mpdl.mpg.de bzw. y.grossmann@ub.uni-frankfurt.de

Max Planck Digital Library

- Landsberger Straße 346, 80687 München (Laim)
- Informationsservices seit 2007
- Seit 2012 eigenständige Serviceeinheit
- rund 80 Mitarbeiter_innen aus Software-Entwicklung, Bibliothek, Wissenschaftsmanagement und Verwaltung
- www.mpdl.mpg.de
- Abteilung Collections für das Thema Forschungsdaten



Max Planck Digital Library

- MPDL versteht sich als eine der vier zentralen IT-Service-Einrichtungen der Max-Planck-Gesellschaft
 - MPDL, MPCDF, IKT und GWDG
- MPDL bildet gemeinsam mit den Institutsbibliotheken das Bibliothekswesen der Max-Planck-Gesellschaft
 - Subsidiaritätsprinzip

Max-Planck-Gesellschaft

- 84 Forschungsinstitute
 - auch Italien, Niederlande und USA (Florida)
- Max Planck Digital Library und lokale Bibliotheken an den Instituten
- Drei Rechenzentren
- Generalverwaltung in München
- Rund 24.000 Mitarbeiter
- Budget von 2 Milliarden € im Jahr 2020
- 31 Nobelpreisträger



FDM-Services der MPDL



RDM Support

Support in the area of research data management



Workshops

Recurring workshops on research data management



Labfolder

An ELN solution by scientists for scientists, licensed for the Max Planck Society



RDMO for MPG

A service organize the data management and the writing of data management plans



Keeper

Data sharing within a project, collaborative working and long-term archiving



Edmond

The open research data repository of the Max Planck Society



DOI-Service

Permanent identifiers for scientific output



Software Licensing Service

An institutionalized basic supply of software and online services

<https://rdm.mpdl.mpg.de/mpdl-services/>

Ziele des Kurses

- Verständnis von Forschungsdaten
- Ziele und Anwendungen von Forschungsdatenmanagement
- Überblick zu aktuellen Entwicklungen und Diskussionen
- ...

Struktur des Kurses

Teil 1: Forschungsdaten

Teil 2: Forschungsdateninfrastruktur

Teil 3: Forschungsdatenmanagement in Bibliotheken

Teil 4: Open Science und wissenschaftliche Software

Sitzungstermine

- 24. Januar 2024, 14:00-15:35 Uhr, via BBB
- 7. Februar 2024, 14:00-15:35 Uhr, via BBB
- 22. Februar 2024, 8:30-12:00 Uhr, vor Ort

Teil 1

Forschungsdaten

Inhalt von Teil 1

1. Beispiele von Forschungsdaten
2. Definitionsversuche
3. Klassifikationen
4. Rechtliche Rahmen
5. Teilen und Probleme

Wofür Forschungsdatenmanagement?

Es ist sinnvoll, weil:

- Wiederverwendung von Daten, zumindest der eigenen, aber auch anderer
- Wissenschaftliches Ansehen
- die Akzeptanz von Daten als eigenständige Publikation wird in Zukunft wahrscheinlich zunehmen
- „passt irgendwie“ hat noch nie einen Nobelpreis gewonnen
- ...

Wofür Forschungsdatenmanagement?

Es ist eine Verpflichtung:

- Standard für „Gute wissenschaftliche Praxis“ von Institutionen
- Forschungsdatenpolicies von Institutionen
- Vorgaben von Förderinstitutionen
- Anforderung von Verlagen
- ...

Einige FDM-Themen

Rechtliche
Aspekt

FAIR-Daten-
Grundsätze

Dauerhafter
Identifikator

Schulungsmög-
lichkeiten

Handhabung
von Dateien

Dateiformate

Metadaten

Normen

Elektronische
Laborjournale

Sicherung und
Archivierung

Daten-
Repositories

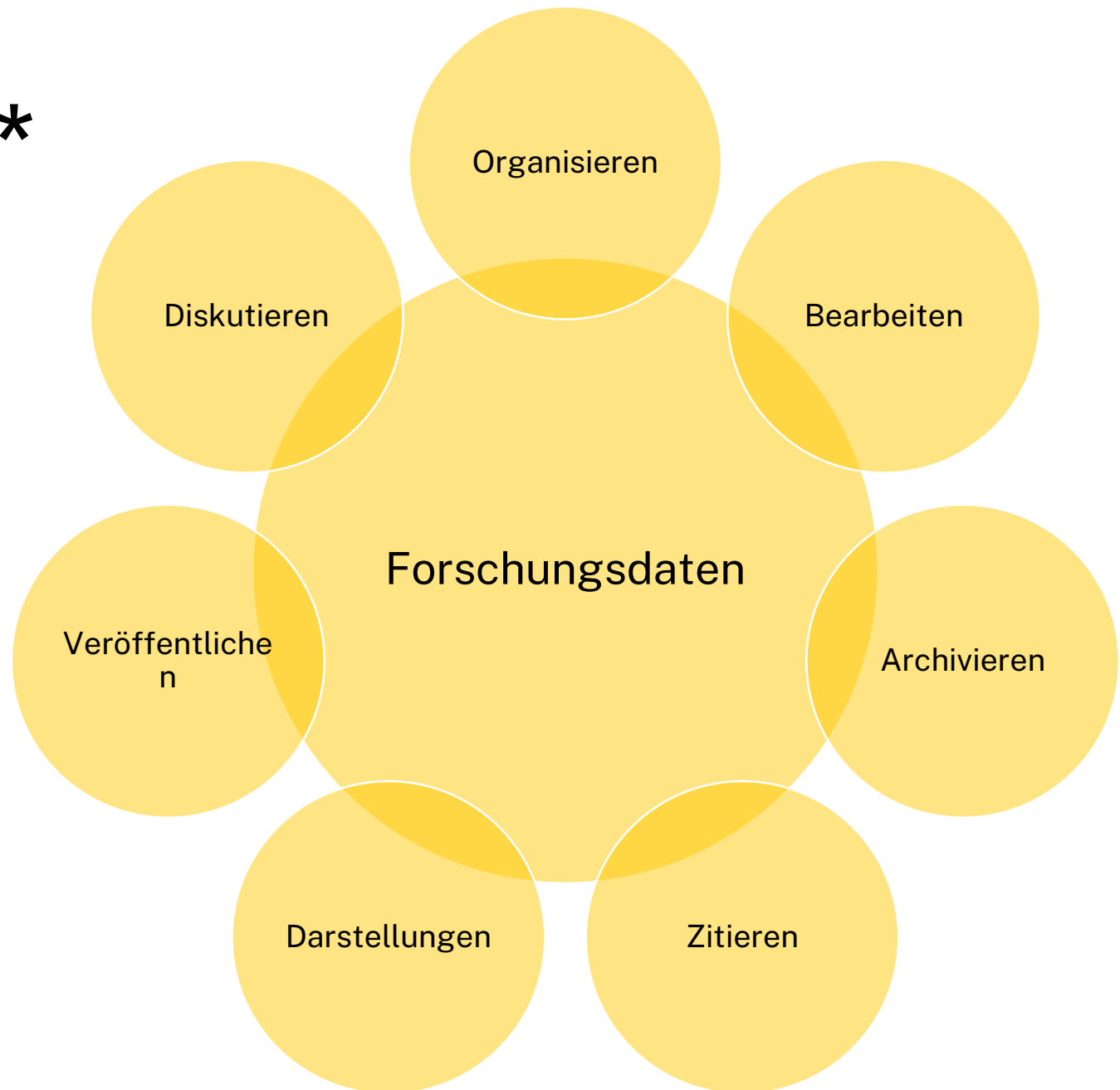
NFDI und EOSC

Open Data

Daten-Lizenzen

Forschungssoft-
ware

Datenumgang*



Beispiele für Forschungsdaten

Beispiele für Forschungsdaten

- Gravitationswellen
- LIGO (<https://ligo.org>)

LIGO LIGO Open Science Center
LIGO is operated by California Institute of Technology and Massachusetts Institute of Technology and supported by the U.S. National Science Foundation.

Getting Started
Tutorials
Data
Events
Bulk Data
Timelines
My Sources
Software
GPS → UTC
About LIGO
Data Analysis
Projects
Acknowledgement

Welcome to the LIGO Open Science Center

About LIGO
Get Started with LIGO data
Join the E-mail list for updates
For general information on LIGO, please visit ligo.org
If you have LSC credentials, you may go to the [development site](#)

Discoveries from the LIGO detectors!

released 2017 June 1:
Event of January 4, 2017: GW170104: total mass 50

released 2016 June 15:
Event of December 26, 2015: GW151226: total mass 22

released 2016 June 15:
Candidate event of October 12, 2015: LVT151012: total mass 37

released 2016 Feb 11:
Event of September 14, 2015: GW150914: total mass 65

The LIGO Laboratory's Data Management Plan describes the scope and timing of LIGO data releases.

Jupyter notebook
See the new tutorial on signal processing with LIGO data, as a Jupyter (Python) notebook.
[Tutorial on Binary Black Hole Signals in LIGO Open Data](#)

Beispiele für Forschungsdaten

- Gensequenzen
- GenBank
(<https://www.ncbi.nlm.nih.gov/genbank/>)

The screenshot shows the NCBI GenBank entry for the Homo sapiens class II AP endonuclease (APE) gene, partial CDS. The entry includes the following information:

- LOCUS:** HUMAPEB 3019 bp DNA linear PRI 31-DEC-1994
- DEFINITION:** Homo sapiens class II AP endonuclease (APE) gene, partial CDS.
- ACCESSION:** M99703
- VERSION:** M99703.1 GI:178748
- KEYWORDS:** 3' diesterase associated activity; DNA repair protein; DNA-binding; class II AP endonuclease; nuclear localized protein.
- SOURCE:** Homo sapiens (human)
- ORGANISM:** Homo sapiens
- REFERENCE:** Harrison, L., Ascione, G., Menninger, J.C., Ward, D.C. and Demple, B. Human apurinic endonuclease gene (APE): structure and genomic mapping (chromosome 14q11.2-12) Hum. Mol. Genet. 1 (9), 677-680 (1992)
- PUBMED:** 1284593
- FEATURES:** Location/Qualifiers
 - source:** 1..3019 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606"
 - intron:** 596..778 /gene="APE" /number=1
 - gene:** join(779..904,1115..1302,1869..2061) /gene="APE"
 - mRNA:** join(779..904,1115..1302,1869..2061) /gene="APE"
 - exon:** 779..904 /gene="APE" /number=2
 - CDS:** join(847..904,1115..1302,1869..2061)

The right sidebar contains additional tools and information:

- Change region shown:** [Dropdown menu]
- Customize view:** [Dropdown menu]
- Analyze this sequence:** Run BLAST, Pick Primers, Highlight Sequence Features, Find in this Sequence
- Articles about the APEX1 gene:** Expression and prognostic significance of APE1/Ref-1 and NPM1 [Am J Clin Pathol. 2014], Dysregulation of human apurinic/aprimidinic endonuclease 1 (APE1) [Br J Ophthalmol. 2014], PRP19 transforms into a sensor of RPA-ssDNA after DNA damage and drives ATI [Mol Cell. 2014]
- Reference sequence information:** RefSeq alternative splicing, See 6 reference mRNA sequence splice variants for the APEX1 gene.
- More about the APEX1 gene:** Apurinic/aprimidinic (AP) sites occur frequently in DNA molecules by spontaneous hydrolysis, by DNA damaging agents or by DNA glycosylases. Also Known As: APE, APE1, APEN, APEX, ...

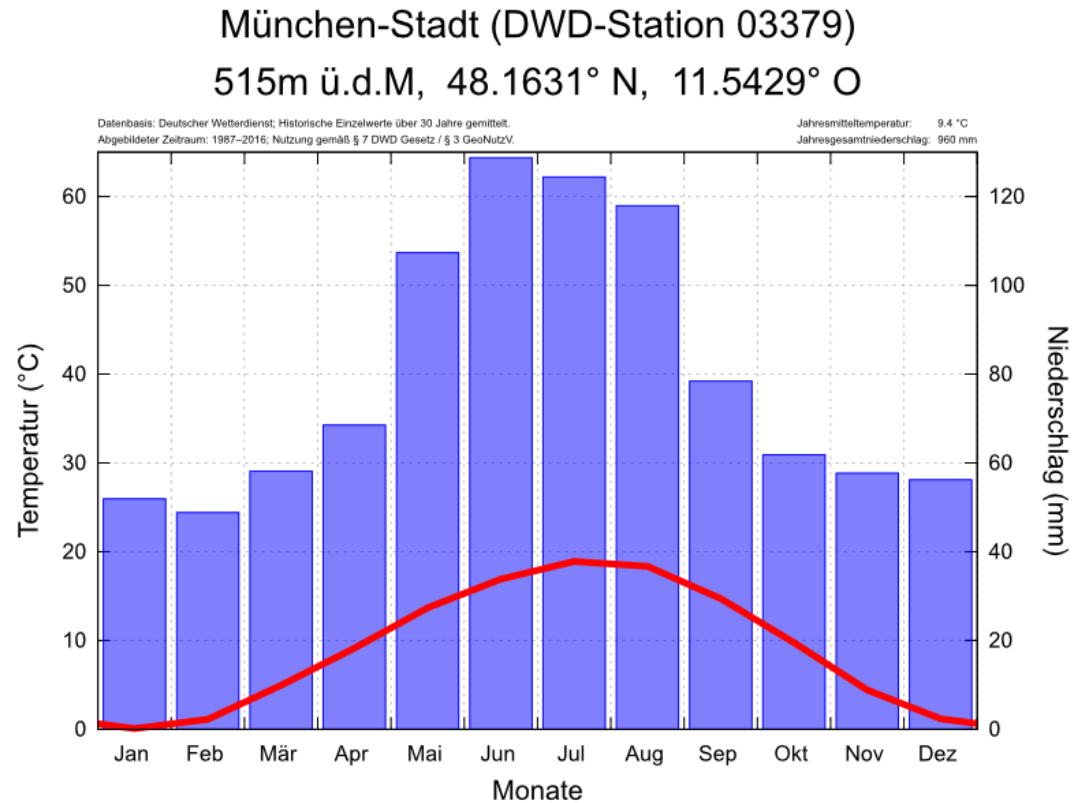
Beispiele für Forschungsdaten

- Tierbewegungen
- Movebank (<https://www.movebank.org>)

The screenshot displays the Movebank website interface. On the left, there is a 'User Login' section with fields for 'Username' and 'Password', and a 'Log in' button. Below it are links for 'Create new account' and 'Request new password'. The main content area is divided into a search panel and a map. The search panel has a search bar with 'stork' entered, a 'Search' button, and a list of search results sorted by 'Animal Identifier'. The map on the right shows Europe with various countries labeled (Deutschland, France, Italia, España, etc.) and several green circular markers representing tracking data points. The Movebank logo and navigation menu are visible at the top.

Beispiele für Forschungsdaten

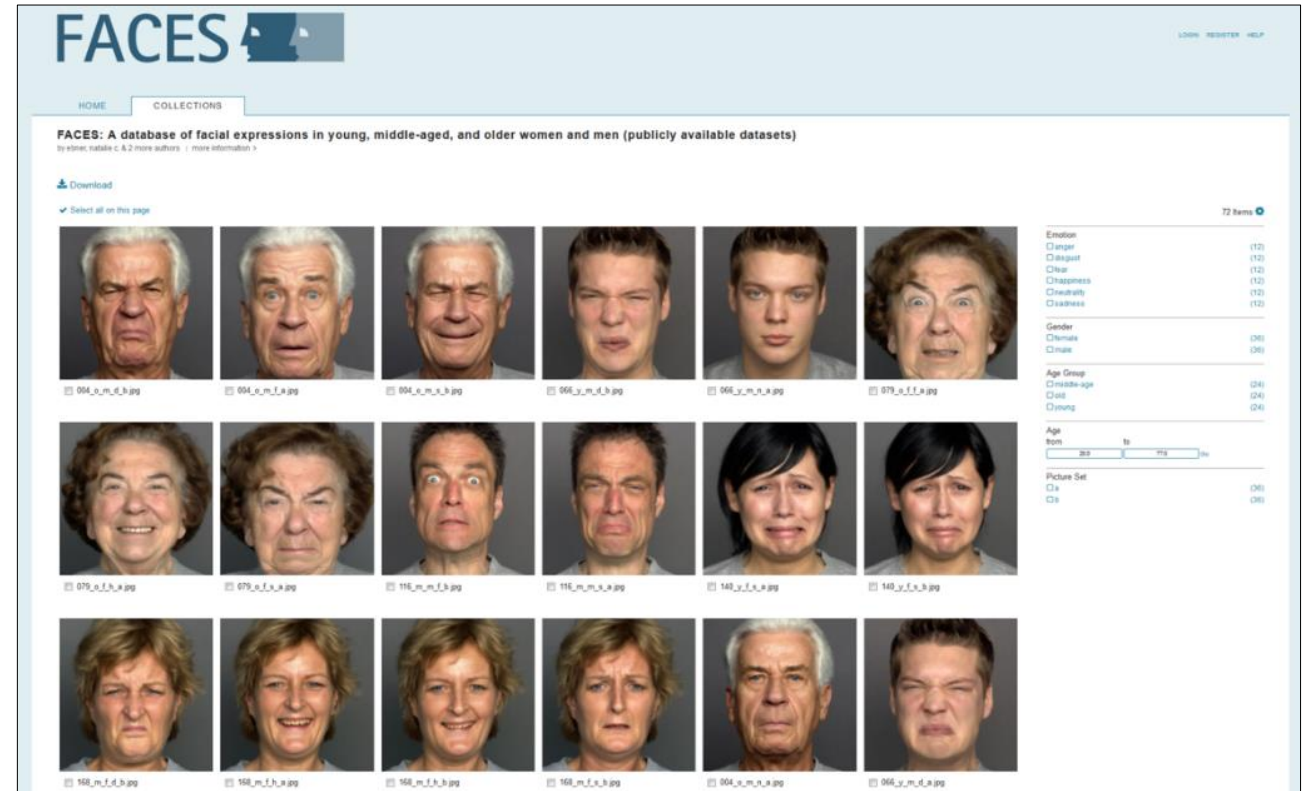
- Klimadaten
- DKRZ World Data Centre for Climate: <https://cera-www.dkrz.de/WDC/Ui/cerasearch/>



Der-blaue-elefant, Klimadiagramm München-Stadt DWD 03379 Jahre 1987-2016, CC BY SA 4.0.

Beispiele für Forschungsdaten

- Gesichter und Mimik
- **FACES**
(<https://faces.mpdl.mpg.de>)



The screenshot displays the FACES database website. The main content area shows a grid of 18 facial images arranged in three rows and six columns. Each image is labeled with a file name below it, such as 004_u_m_1.jpg, 004_u_m_2.jpg, 004_u_m_3.jpg, 006_y_m_1.jpg, 006_y_m_2.jpg, 006_y_m_3.jpg, 073_u_f_1.jpg, 073_u_f_2.jpg, 073_u_f_3.jpg, 116_m_m_1.jpg, 116_m_m_2.jpg, 116_m_m_3.jpg, 142_y_f_1.jpg, 142_y_f_2.jpg, 142_y_f_3.jpg, 168_m_f_1.jpg, 168_m_f_2.jpg, 168_m_f_3.jpg, 004_u_m_4.jpg, and 006_y_m_4.jpg. The sidebar on the right contains filters for Emotion, Gender, Age Group, and Picture Set, with counts for each category. The top navigation bar includes 'HOME' and 'COLLECTIONS' tabs, and the 'FACES' logo is prominently displayed.

Beispiele für Forschungsdaten

- Daten für Psychologie
- PsychData (<https://rdc-psychology.org>)
- Verschiedene Zugangsszenarien, je nach Sensibilität der Daten (z.B. Patientendaten)



zpid[®]
Max-Planck-Institut für Psychologie

Contact

RDC at ZPID

Das FDZ am ZPID | Daten teilen | Daten nutzen | Die FDZ-Wissensbasis | FAQ | Über das FDZ am ZPID | DE

Entwicklungspsychologie

Weinert et al. (2015). Forschungsdaten der Münchner Longitudinalstudie zur Genese individueller Kompetenzen (LOGIK): Entwicklung der Intelligenz und des logischen Denkvermögens.

Bibliografische Informationen

Ersteller: Max-Planck-Institut für psychologische Forschung; Weinert, Franz; Schneider, Wolfgang

Mitwirkende: Weinert, Franz; Schneider, Wolfgang; Stefanek, Jan

Förderung: Max-Planck-Gesellschaft; VW-Stiftung; Jacobs Foundation

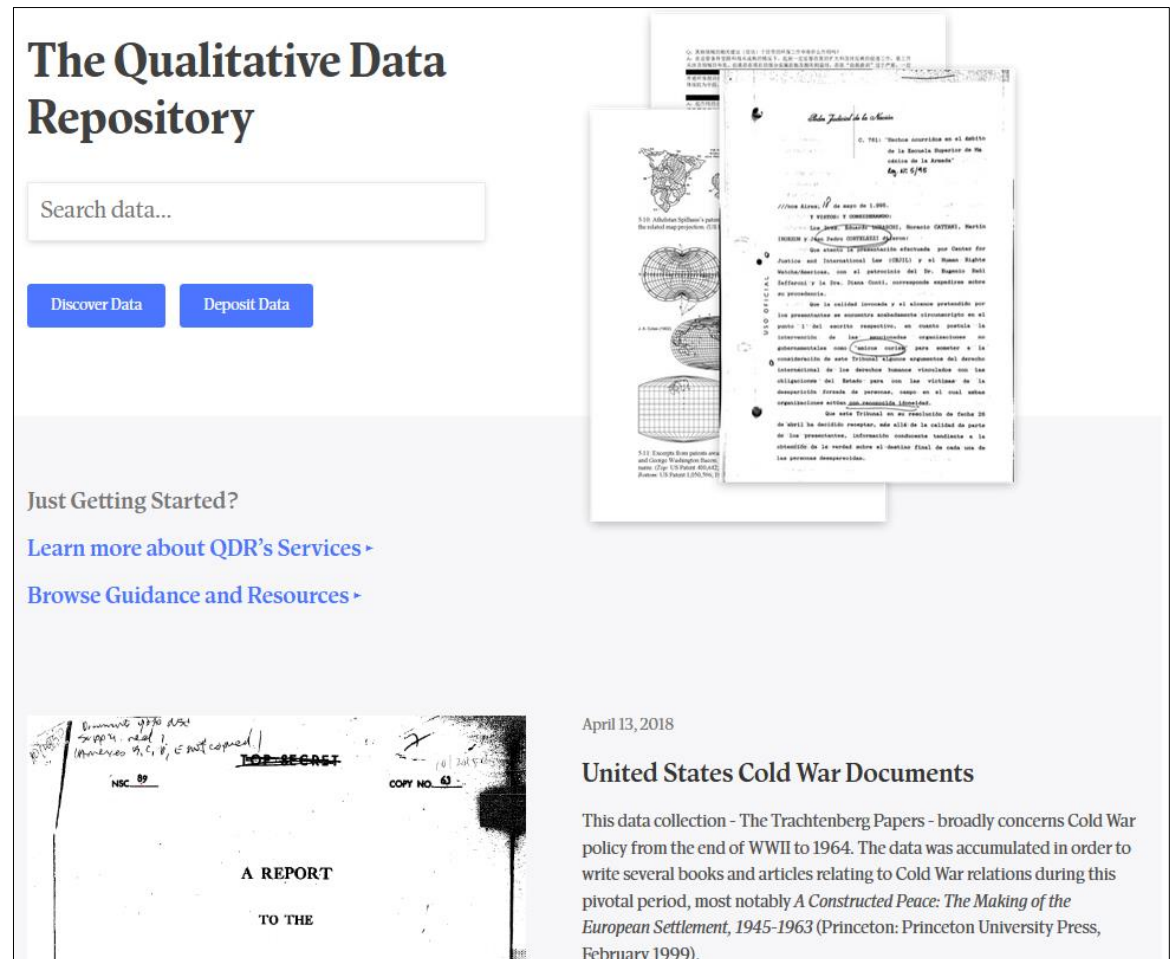
Titel: Forschungsdaten der Münchner Longitudinalstudie zur Genese individueller Kompetenzen (LOGIK): Entwicklung der Intelligenz und des logischen Denkvermögens

Jahr der Publikation: 2015

<https://doi.org/10.5160/psychdata.wtfz05lo22>

Beispiele für Forschungsdaten

- Qualitative Daten
- United States Cold War Documents (<https://qdr.syr.edu>)



The screenshot shows the homepage of The Qualitative Data Repository (QDR). The header features the title "The Qualitative Data Repository" in a large, bold font. Below the title is a search bar with the placeholder text "Search data...". Two buttons, "Discover Data" and "Deposit Data", are positioned below the search bar. To the right of the search bar, there is a preview of a historical document with a map and text. Below the search bar, there is a section titled "Just Getting Started?" with two links: "Learn more about QDR's Services >" and "Browse Guidance and Resources >". At the bottom left, there is a preview of a document titled "A REPORT TO THE" with a "TOP SECRET" stamp and a "COPY NO. 62" stamp. The date "April 13, 2018" is displayed in the bottom right corner, along with the title "United States Cold War Documents". A paragraph of text describes the data collection, mentioning the Trachtenberg Papers and the book *A Constructed Peace: The Making of the European Settlement, 1945-1963*.

The Qualitative Data Repository

Search data...

Discover Data Deposit Data

Just Getting Started?

[Learn more about QDR's Services >](#)

[Browse Guidance and Resources >](#)

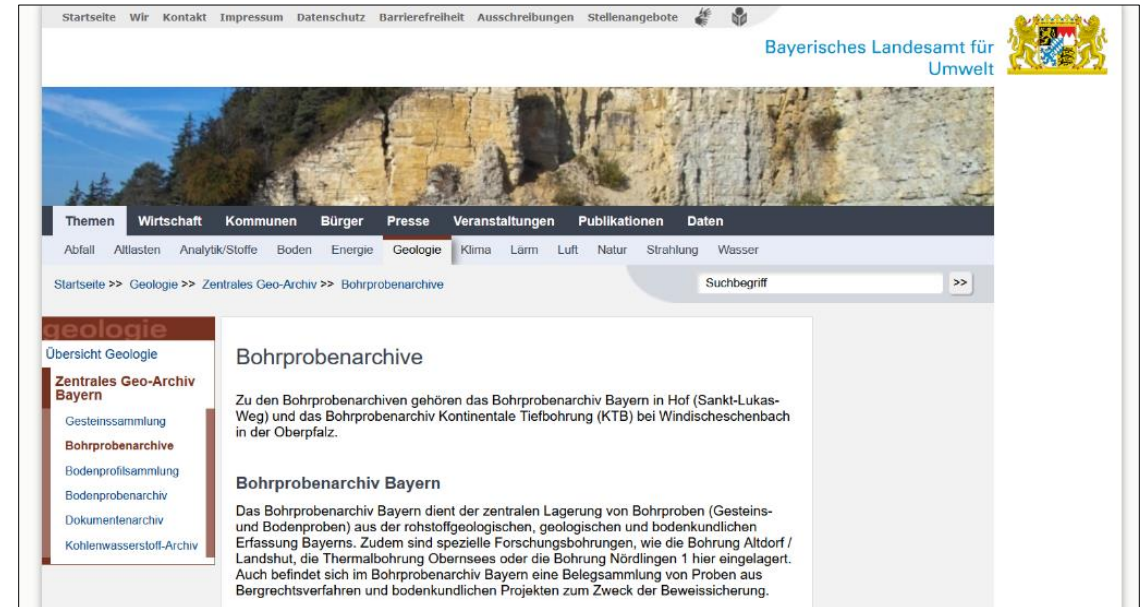
April 13, 2018

United States Cold War Documents

This data collection - The Trachtenberg Papers - broadly concerns Cold War policy from the end of WWII to 1964. The data was accumulated in order to write several books and articles relating to Cold War relations during this pivotal period, most notably *A Constructed Peace: The Making of the European Settlement, 1945-1963* (Princeton: Princeton University Press, February 1999).

Worum geht es nicht

- Physische Daten
 - Wissenschaftliche Publikationen
 - Daten über Wissenschaft
-
- z.B. [Bohrprobenarchive des Bayerischen Landesamts für Umwelt](#)



Worum geht es nicht

- Physische Daten
- Wissenschaftliche Publikationen
- Daten über Wissenschaft



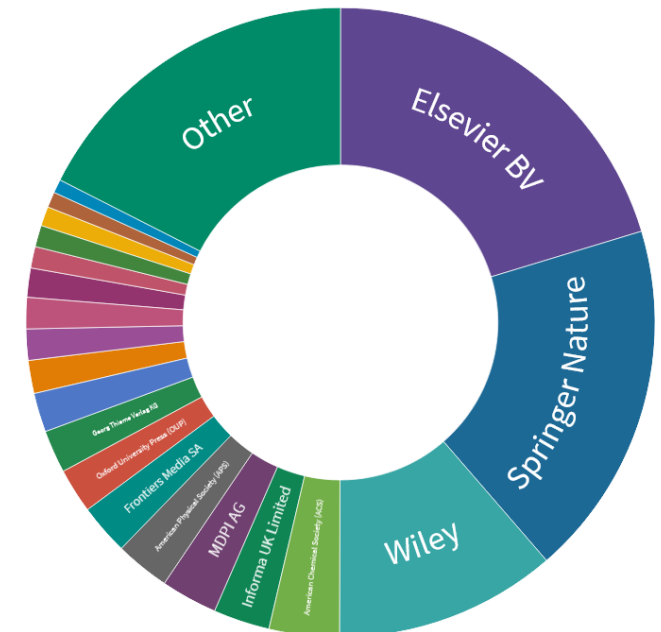
[“Paper Weaving”](#) by FeatheredTar is licensed under CC BY 2.0

Worum geht es nicht

- Physische Daten
- Wissenschaftliche Publikationen
- Daten über Wissenschaft

Top 20 publishers based on corresponding author share

Filter:
Germany



<https://esac-initiative.org/market-watch/>

Forschungsdaten-Quiz

Frage:

Was sehen Sie?

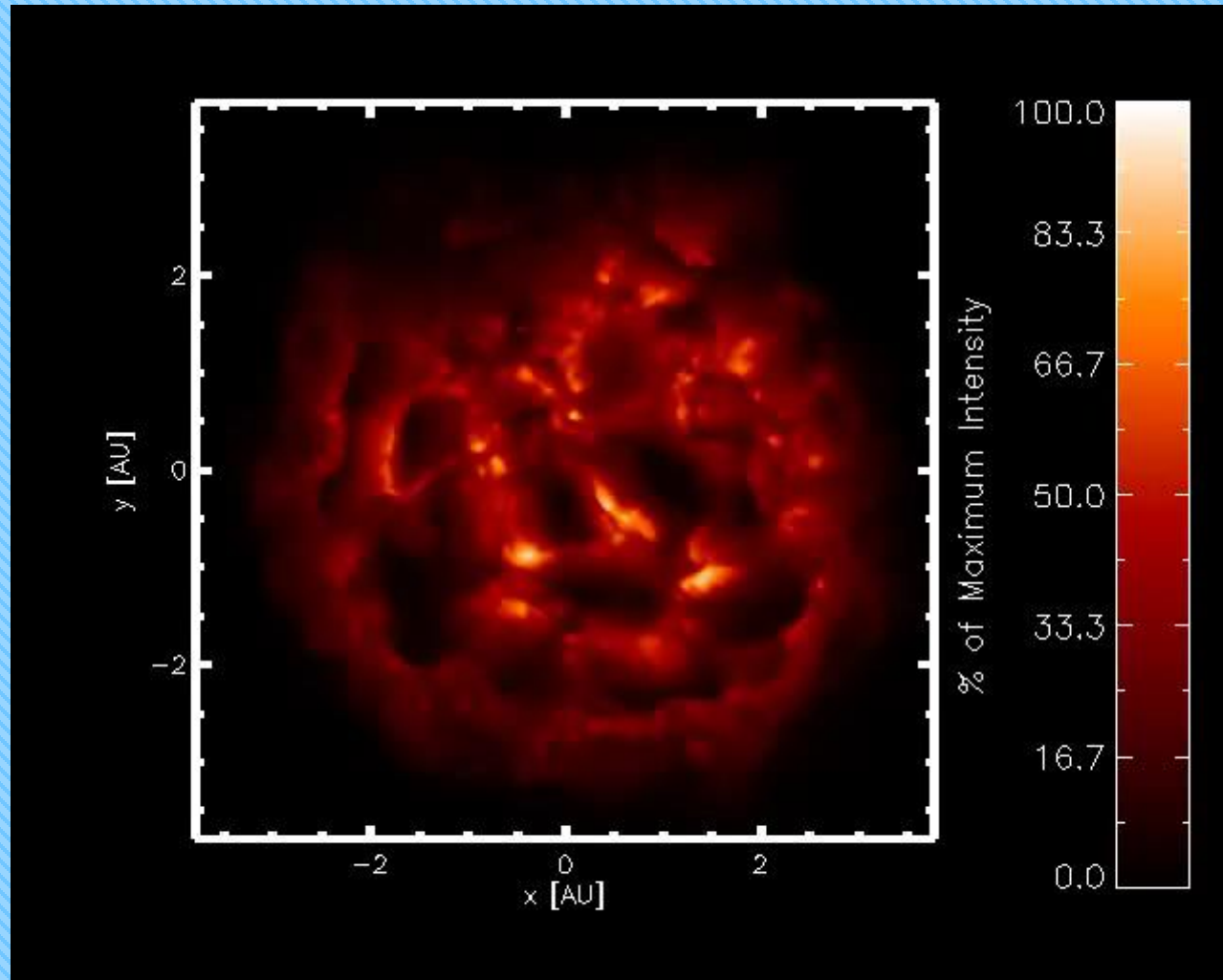
Methode:

Ideen reinrufen

Dauer:

3 min

Kurzes Video-Quiz



Was verbirgt sich hinter dieser Videodatei eines Forschungsdatensatzes?

A. Chiavassa, R. Kudritzki, B. Davies, B. Freytag, & S. E. de Mink. (2022, March 16): Photo-center displacements for RSGs as seen by Gaia (Version v1), Max Planck Institute for Astrophysics, CC BY 4.0, Zenodo, <https://doi.org/10.5281/zenodo.6363011>.

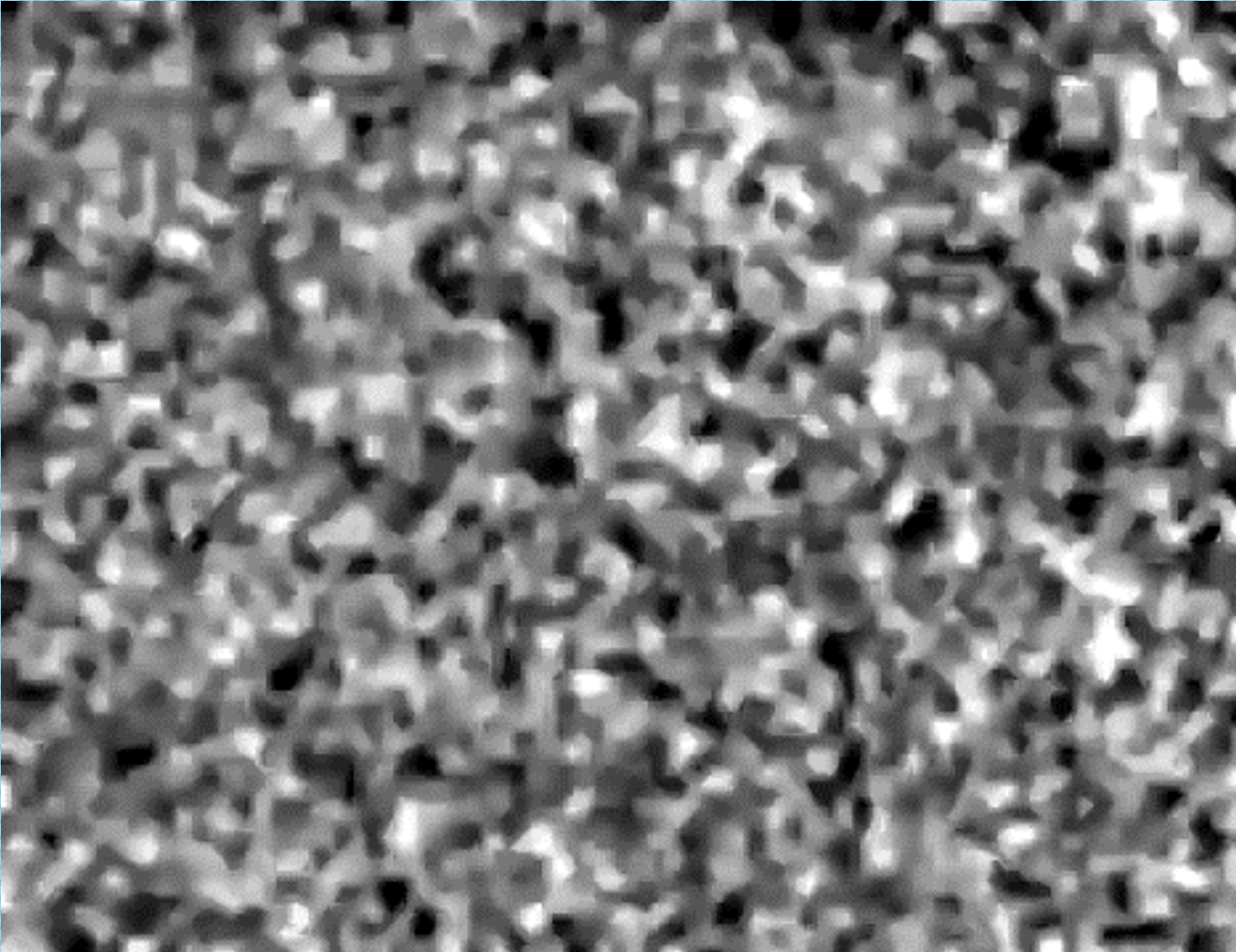
Kurzes Audio-Quiz

Was versteckt sich hinter dieser Audio-Datei eines Forschungsdatensatzes?



Auflösung: Cecilia Durojaye, 2020, „Recodings – Nigerian Talking drum“, file D1s,Edmond, <https://doi.org/10.17617/3.IUMMLP>, CC BY 4.0.

Kurzes Video-Quiz

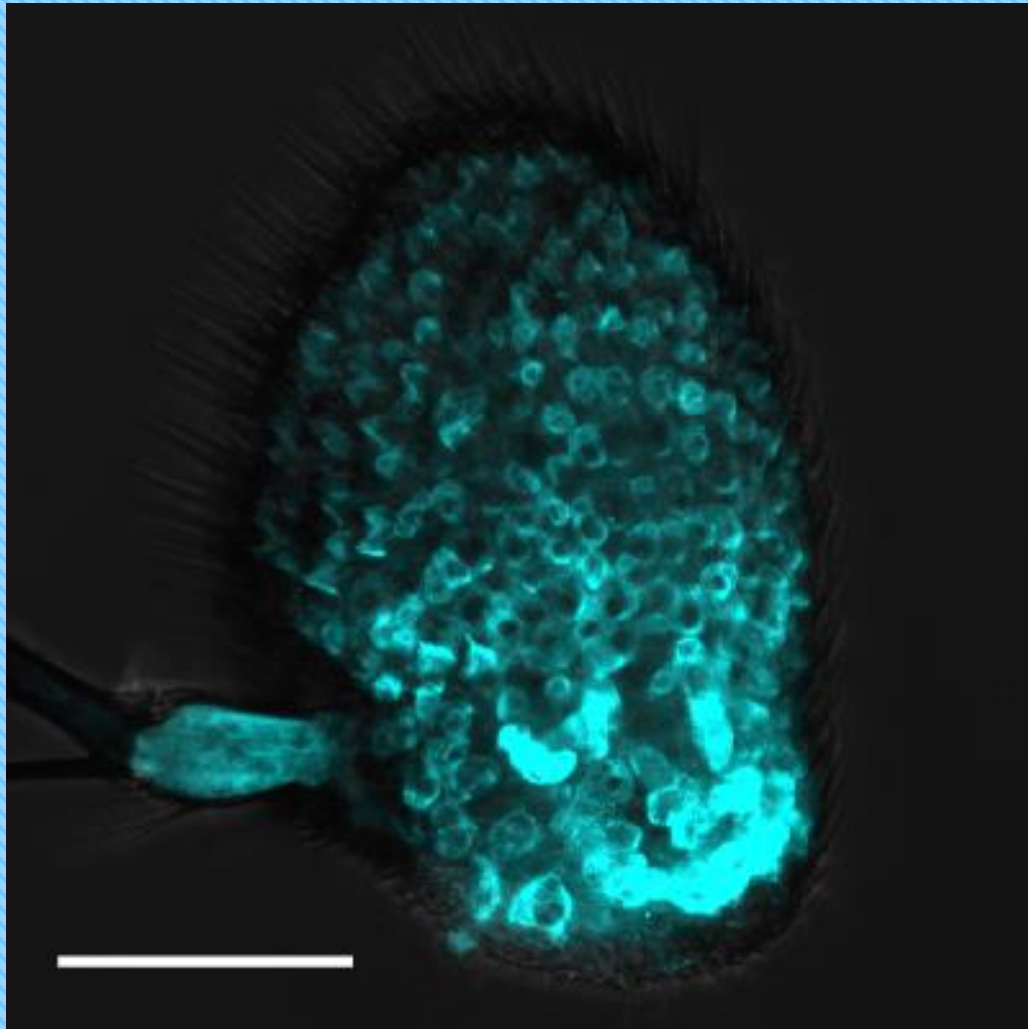


Was verbirgt sich hinter dieser Videodatei eines Forschungsdatensatzes?

Kaumudi Prabhakara (Max Planck Institute for Dynamics and Self-Organization), Spiralwaves, 2015, CC BY 4.0,

<https://www.youtube.com/watch?v=DmRZn073Uus>.

Kurzes Bild-Quiz



Was verbirgt sich hinter dieser Bilddatei eines Forschungsdatensatzes?

Prelic, Sinisa, 2021, "Functional interaction between *Drosophila* olfactory sensory neurons and their support cells", Max Planck Institute for Chemical Ecology, <https://doi.org/10.17617/3.7m>, Edmond, V1, CC BY 4.0, 20191021 image 4 ASE5-Gal4 GFP.tif.

Definitionsversuche

Definitionsversuche: 1

„Zu Forschungsdaten zählen u. a. Messdaten, Laborwerte, audiovisuelle Informationen, Texte, Surveydaten oder Beobachtungsdaten, methodische Testverfahren sowie Fragebögen.“

Fragenkatalog der DFG, Stand 10. Juni 2021, S. 1,

https://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/forschungsdaten/forschungsdaten_checkliste_de.pdf

Definitionsversuche: 2

Forschungsdaten sind *„Daten, die im Zuge wissenschaftlicher Vorhaben z.B. durch Digitalisierung, Quellenforschungen, Experimente, Messungen, Erhebungen oder Befragungen entstehen.“*

Allianz AG „Forschungsdaten“, Forschungsdatenmanagement – Eine Handreichung, 2018, S. 4, https://gfzpublic.gfz-potsdam.de/rest/items/item_3055893_5/component/file_3055894/content

Definitionsversuche: 3

„Forschungsdaten umfassen alle Daten, die im wissenschaftlichen Arbeitsprozess entstehen und in digitaler Form verarbeitet werden.“

Leitlinie zum Umgang mit Forschungsdaten in der Leibniz-Gemeinschaft, 2018, S. 1, https://www.leibniz-gemeinschaft.de/fileadmin/user_upload/Bilder_und_Downloads/Forschung/Open_Science/Leitlinie_Forschungsdaten_2018.pdf

Klassifikation von Forschungsdaten

Klassifikationen: Nach dem Charakter der Daten

Qualitative Daten

- Texte (Quellen, Transkripte)
- Bilder (Scans, Fotos, Mikroskopiedaten, Teleskopdaten, Satellitendaten)
- Multimedia (Audiodaten, Videodaten, 3D-Daten, 4D-Daten)

Quantitative Daten

- Zahlenreihen (Umfragedaten, Experimentaldaten, Sensormessreihen, Gensequenzen, Geodaten)
- Born-digitals (Simulationsdaten, Algorithmen, Websites)

Klassifikationen: Nach dem Charakter der Daten

Text (TXT, ODT, PDF, ...)

Glossed text

See p. 12. 13 in:

van der Berg, Margot C. & Smith, Norval S.H. 2013. Early Shona. In: Mchale, Suzanne Maria & Mair, Philippe & Eisenstadt, Martin & Huter, Magana (eds.) *The Survey of African and Creole Languages. Vol. 5: English-based and Dutch-based languages*, 3-13. Oxford: Oxford University Press.

The following lines, taken from the play in van Dyk's (c1765) language manual, present a detailed account of a plantation funeral that is generally considered to be true to life (van Trier-Guicherit 1991).

Zangwe Officer: *Tippa kiser nyshi ut go masira negere*
close coffin make 1Pl. go master black
Black Overseer: 'Close the coffin, let's go, Gentlemen.

awosa wakke awonwa ope de bawon
2PL six person lift DE:SG woman
you six persons, lift up the woman.

ba pty de bawon kange negere alle
Q place DET.PL woman with black all
Where is the woman with all the blacks?

awosa wakke na bippi kry now nese
2PL walk LOC heap wail man must
You must walk as a group, the wailers must.

wabbe na fesi awosa no regezi war
walk LOC face 2Pl. NEG forget one
walk at the front. Don't forget a

santi kango-kango
thing come-go
thing! Come, come!

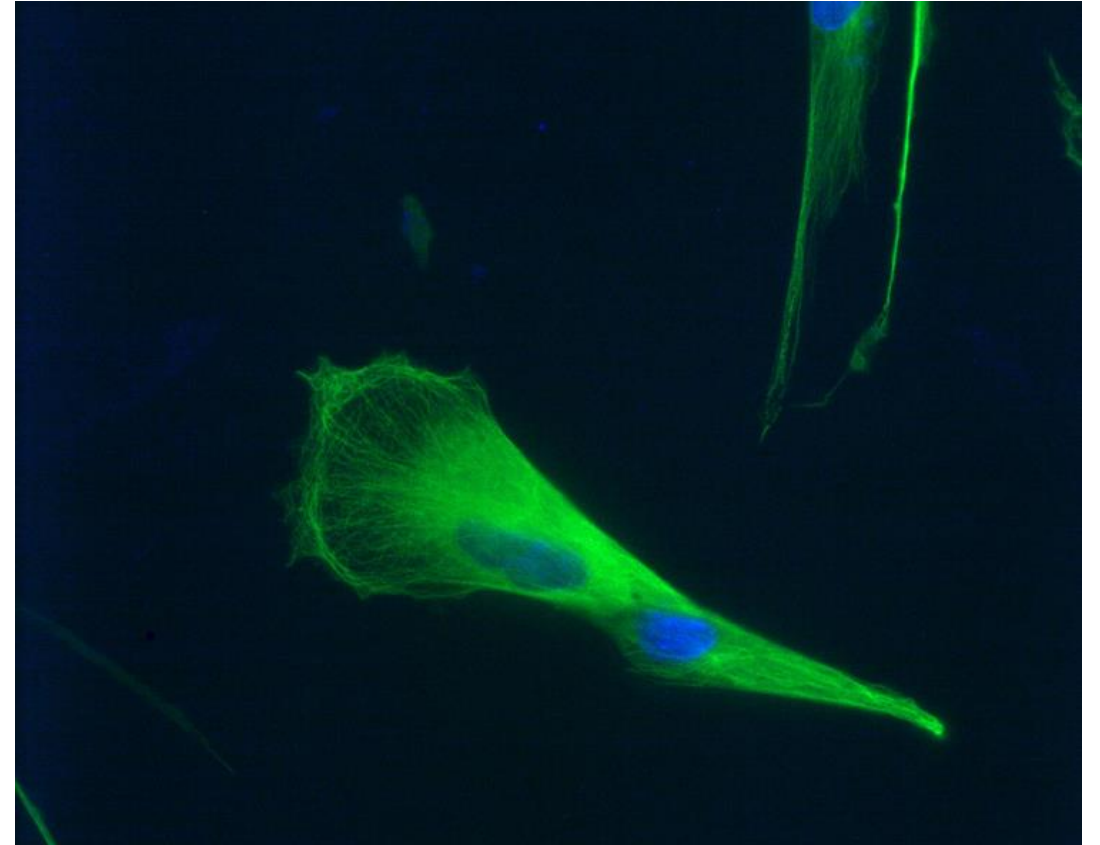
Zangwe: *Uu so ut jure dible awonwa go ni jure de*
that so 1Pl. carry dead person go 1SG play
Singers: 'That's how we carry the dead away. I celebrate, We

ut zurei fu ju alle de ju go lawi en, na
1Pl. regret for 2Pl. all that 3SG go expire so LOC
are very sorry for you. It is you who died like this.

Klassifikationen: Nach dem Charakter der Daten

Text (TXT, ODT, PDF, ...)

Bilder (TIFF, JPEG, ...)



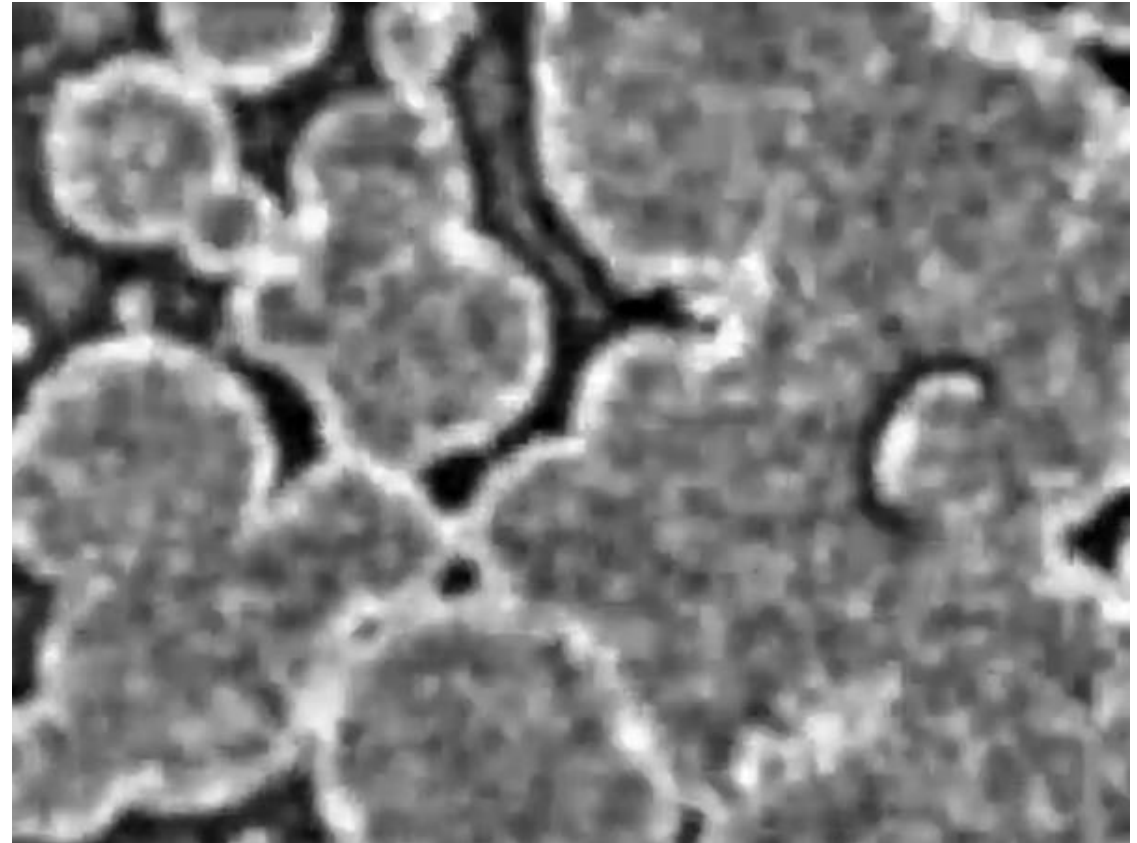
Ries, Albert (Max Planck Institute for Biochemistry), IMR90 fibroblasts, 2014, CC BY 4.0, <https://doi.org/10.17617/3.ND7FXM>.

Klassifikationen: Nach dem Charakter der Daten

Text (TXT, ODT, PDF, ...)

Bilder (TIFF, JPEG, ...)

Multimedia (MP4, Dicom, ...)



Kaumudi Prabhakara (Max Planck Institute for Dynamics and Self-Organization),
Spiralwaves, 2015, CC BY 4.0, <https://www.youtube.com/watch?v=DmRZn073Uus>.

Klassifikationen: Nach dem Charakter der Daten

Text (TXT, ODT, PDF, ...)

Bilder (TIFF, JPEG, ...)

Multimedia (MP4, Dicom, ...)

Zahlenreihen (CSV, XLSX, ...)

```
HCCT-2010
Schmücke station
MAAP @ interstitial inlet
MPIC Mainz
Contact: Johannes Schneider, johannes.schneider@mpic.de

Data status: Final (1.3.2011)
(unreliable data points have been removed)

Date and Time
Black_Carbon_ng/m3

11.09.2010 18:30:05      144
11.09.2010 18:31:05      117
11.09.2010 18:32:05      157
11.09.2010 18:33:05      165
11.09.2010 18:34:05      198
11.09.2010 18:35:05      226
11.09.2010 18:36:05      230
11.09.2010 18:37:05      216
11.09.2010 18:38:05      232
11.09.2010 18:39:05      232
11.09.2010 18:40:05      249
11.09.2010 18:41:05      241
11.09.2010 18:42:05      449
11.09.2010 18:43:05      329
11.09.2010 18:44:05      252
11.09.2010 18:45:05      215
11.09.2010 18:46:05      119
11.09.2010 18:47:05      173
11.09.2010 18:48:05      215
11.09.2010 18:49:05      235
11.09.2010 18:50:05      256
11.09.2010 18:51:05      230
11.09.2010 18:52:05      252
11.09.2010 18:53:05      238
11.09.2010 18:54:05      244
11.09.2010 18:55:05      215
11.09.2010 18:56:05      196
11.09.2010 18:57:05      162
11.09.2010 18:58:05      188
11.09.2010 18:59:05      216
11.09.2010 19:00:05      197
11.09.2010 19:01:05      213
11.09.2010 19:02:05      243
```

Schneider, Johannes (2017): HCCT2010, CC BY 4.0,
<https://dx.doi.org/10.17617/3.i>

Klassifikationen: Nach dem Charakter der Daten

Text (TXT, ODT, PDF, ...)

Bilder (TIFF, JPEG, ...)

Multimedia (MP4, Dicom, ...)

Zahlenreihen (CSV, XLSX, ...)

Born-digitals (netCDF, grib, HTML, ...)

```
netcdf tos_O1_2001-2002 (
dimensions:
  lon = 180 ;
  lat = 170 ;
  time = UNLIMITED ; // (24 currently)
  bnds = 2 ;
variables:
  double lon(lon) ;
    lon:standard_name = "longitude" ;
    lon:long_name = "longitude" ;
    lon:units = "degrees_east" ;
    lon:axis = "X" ;
    lon:bnds = "lon_bnds" ;
    lon:original_units = "degrees_east" ;
  double lon_bnds(lon, bnds) ;
  double lat(lat) ;
    lat:standard_name = "latitude" ;
    lat:long_name = "latitude" ;
    lat:units = "degrees_north" ;
    lat:axis = "Y" ;
    lat:bnds = "lat_bnds" ;
    lat:original_units = "degrees_north" ;
  double lat_bnds(lat, bnds) ;
  double time(time) ;
    time:standard_name = "time" ;
    time:long_name = "time" ;
    time:units = "days since 2001-1-1" ;
    time:axis = "T" ;
    time:calendar = "360_day" ;
    time:bnds = "time_bnds" ;
    time:original_units = "seconds since 2001-1-1" ;
  double time_bnds(time, bnds) ;
  float tos(time, lat, lon) ;
    tos:standard_name = "sea_surface_temperature" ;
    tos:long_name = "Sea Surface Temperature" ;
    tos:units = "K" ;
    tos:cell_methods = "time: mean (interval: 30 minutes)" ;
    tos:fill_value = 1.e+20F ;
    tos:missing_value = 1.e+20F ;
    tos:original_name = "sststst" ;
    tos:original_units = "degC" ;
    tos:history = " At 16:37:23 on 01/11/2005: CMOR altered the data in the following ways: added 2.73150E+02 to yield output units: Cyclical dimension was output
at a different lon" ;
// global attributes:
  :title = "IPSL model output prepared for IPCC Fourth Assessment SRES A2 experiment" ;
  :institution = "IPSL (Institut Pierre Simon Laplace, Paris, France)" ;
  :source = "IPSL-CM4_v1 (2003) : atmosphere : IPSL (IPSL-CM4_IPCC, 96x71x19) ; ocean ORCA2 (ips1_cm4_v1_0, 2x2131) ; sea ice LIM (ips1_cm4_v" ;
  :contact = "sebastien.denvil, sebastien.denvil@ipsl.jussieu.fr" ;
  :project_id = "IPCC Fourth Assessment" ;
```

Sea surface temperatures collected by PCMDI for use by the [IPCC](https://www.unidata.ucar.edu/software/netcdf/examples/files.html), 2015, <https://www.unidata.ucar.edu/software/netcdf/examples/files.html>

Klassifikation: Nach der Quelle der Daten bzw. der Art der Forschung

Observational [Beobachtung]: data captured in real-time, usually irreplaceable. For example, sensor data, survey data, sample data, neuroimages.

Experimental [Experiment]: Data from lab equipment, often reproducible, but can be expensive. For example, gene sequences, chromatograms, toroid magnetic field data.

Simulation [Simulation]: data generated from test models where model and metadata are more important than output data. For example, climate models, economic models.

Derived or compiled [Ableitung]: data is reproducible but expensive. For example, text and data mining, compiled database, 3D models.

Reference or canonical [Sammlung]: a (static or organic) conglomeration or collection of smaller (peer-reviewed) datasets, most probably published and curated. For example, gene sequence databanks, chemical structures, or spatial data portals.

Klassifikation: Nach dem Grad der Aggregation bzw. Verdichtung

Rohdaten: Als Rohdaten soll die Gesamtmenge aller Daten bezeichnet werden, die empirischer Wissenschaft als Forschungsgrundlage zur Verfügung stehen.

Primärdaten: Primärdaten seien diejenigen Daten, die als Teilmenge der Rohdaten tatsächlich zur Forschung herangezogen werden.

Sekundärdaten: Sekundärdaten seien Daten, die in Prozessschritten aus Primärdaten gewonnen wurden.

Gruppenarbeit

Frage:

Mit welchen Forschungsdaten haben Sie bereits gearbeitet oder sind damit in Kontakt gekommen?

Beispiele:

Arbeit in einer Bibliothek, Besichtigungen in anderen Institutionen, Beratung, Alltag etc.

Methode:

Brainwriting Pool (<https://pad.gwdg.de/s/Tj886w4kB>)

Dauer:

5 Minuten

Rechtliche Rahmen

Rechtliche Vorgaben

- Standard für „Gute wissenschaftliche Praxis“ von Institutionen
- Forschungsdatenpolicies von Institutionen
- Vorgaben von Förderinstitutionen
- Anforderung von Verlagen
- ...

DFG-Verhaltenskodex „Sicherung guter wissenschaftlicher Praxis“

Leitlinie 13: Herstellung von öffentlichem Zugang zu Forschungsergebnissen

„[...] Wissenschaftlerinnen und Wissenschaftler entscheiden in eigener Verantwortung – unter Berücksichtigung der Gepflogenheiten des betroffenen Fachgebiets –, ob, wie und wo sie ihre Ergebnisse öffentlich zugänglich machen. Ist eine Entscheidung, Ergebnisse öffentlich zugänglich zu machen, erfolgt, beschreiben Wissenschaftlerinnen und Wissenschaftler diese vollständig und nachvollziehbar. Dazu gehört es auch, soweit dies möglich und zumutbar ist, die den Ergebnissen zugrunde liegenden Forschungsdaten, Materialien und Informationen, die angewandten Methoden sowie die eingesetzte Software verfügbar zu machen und Arbeitsabläufe umfänglich darzulegen. [...]“ (<http://doi.org/10.5281/zenodo.6472827>, S. 18f.

Max-Planck-Gesellschaft: Gute Wissenschaftliche Praxis

„Vorhalten von Forschungsdaten

Forschungsdaten sind vielfältig und entsprechend sind Art und Umfang der Sicherung und Aufbewahrung durch Forschungsgegenstand und Forschungsmethode bestimmt. Es wird von der Institutsleitung erwartet, die im betreffenden Fachgebiet üblichen Aufbewahrungsmedien bereitzustellen wie auch zu garantieren, dass sowohl analog als auch digital abgespeicherte Information gesichert ist und zugänglich bleibt.“

(<https://www.mpg.de/197494/rulesScientificPractice.pdf>, S. 13)

Europäische Kommission

ERC Work Programme 2021:

”Finally, as from 2021 it is no longer possible for applicants to opt out of the submission of Research Data Management plans.”

([European Commission Decision C\(2021\) 930 of 22/02/2021](#), S. 4)

Marie Skłodowska-Curie Actions:

“[...] data management plan submitted at mid-term and an update towards the end of the project if needed” ([European Commission Decision C\(2021\) 930 of 22/02/2021](#), S. 82)

Horizon Europe

“The beneficiaries must manage the digital research data generated in the action ('data') responsibly, in line with the FAIR principles and by taking all of the following actions:

- **establish a data management plan** ('DMP') (and regularly update it)
- as soon as possible and within the deadlines set out in the DMP, **deposit** the **data** in a trusted repository [...].
- as **soon as possible** and within the deadlines set out in the DMP, ensure **open** access – via the repository – to the deposited data [plus CC0, CC BY or equivalent].“

Horizon Europe and Euratom: General Model Grant Agreement, Version 1.0, 1st June 2021, S. 109,
https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/common/agr-contr/general-mga_horizon-euratom_en.pdf

Teilen und Probleme

Teilen von Forschungsdaten

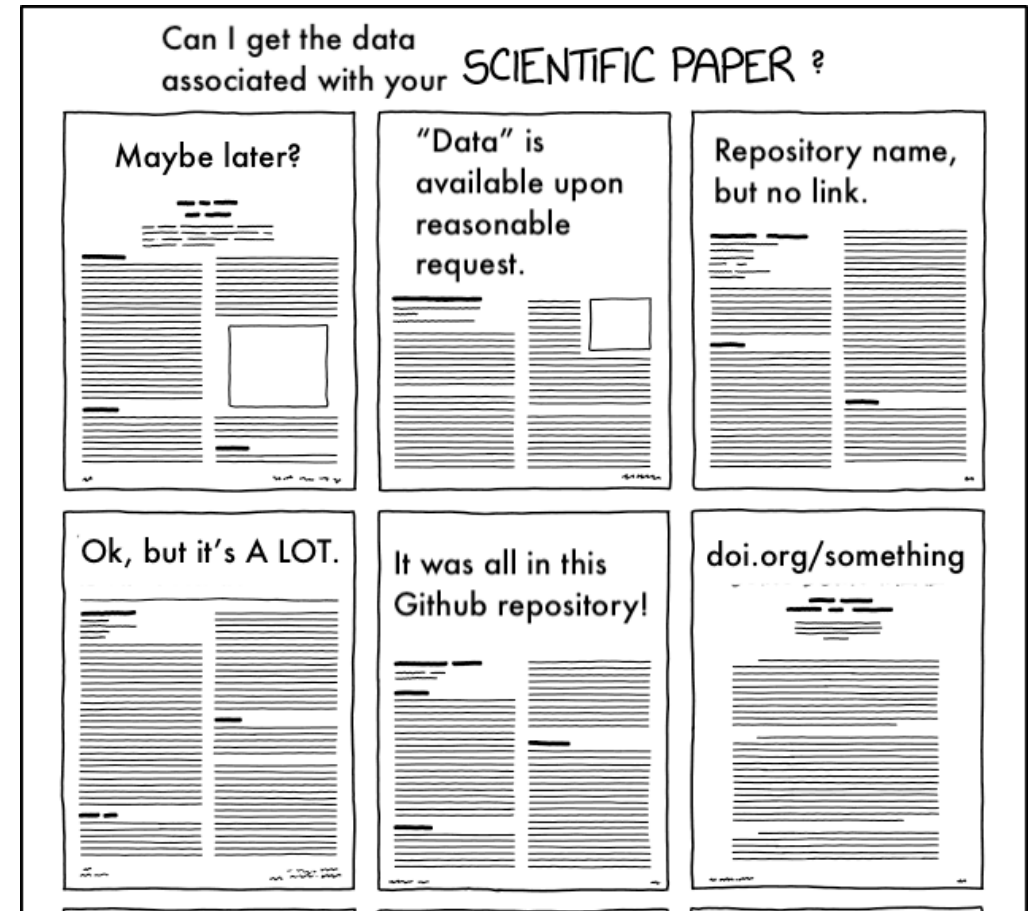
- Community-Standards
- erwünscht von Institutionen
- erwünscht von Geldgebern
- größere Sichtbarkeit der eigenen Forschung
- Datenpublikation als eigenes Genre
- ...

Probleme bei Zugängen

Eigentlich sollten die Forschungsdaten für eine Veröffentlichung zugänglich sein

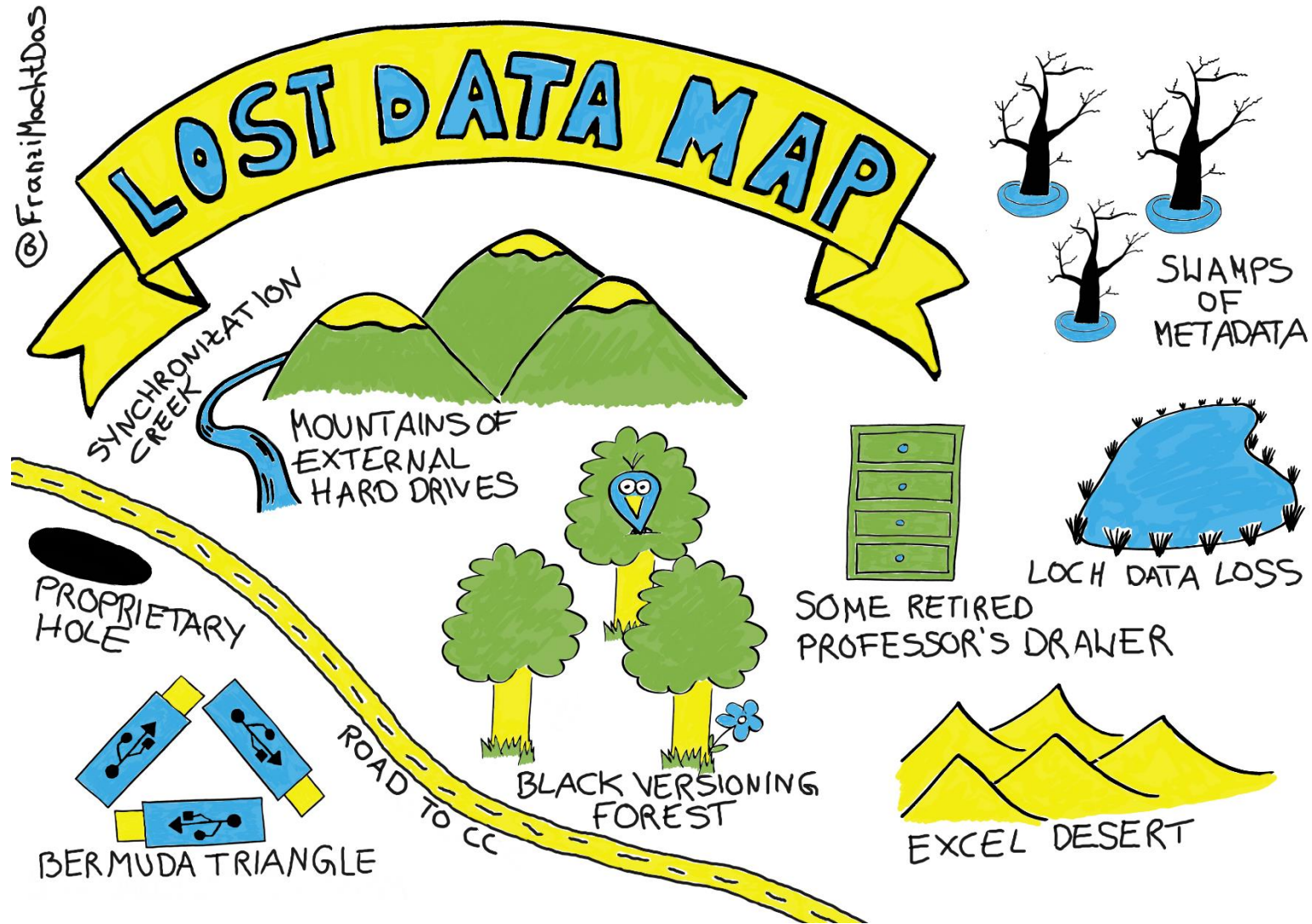
vs.

„available upon reasonable request“



<https://twitter.com/JohnBorghini/status/1388873617002299394/photo/1>, CC BY-NC 2.5.

Probleme beim Management

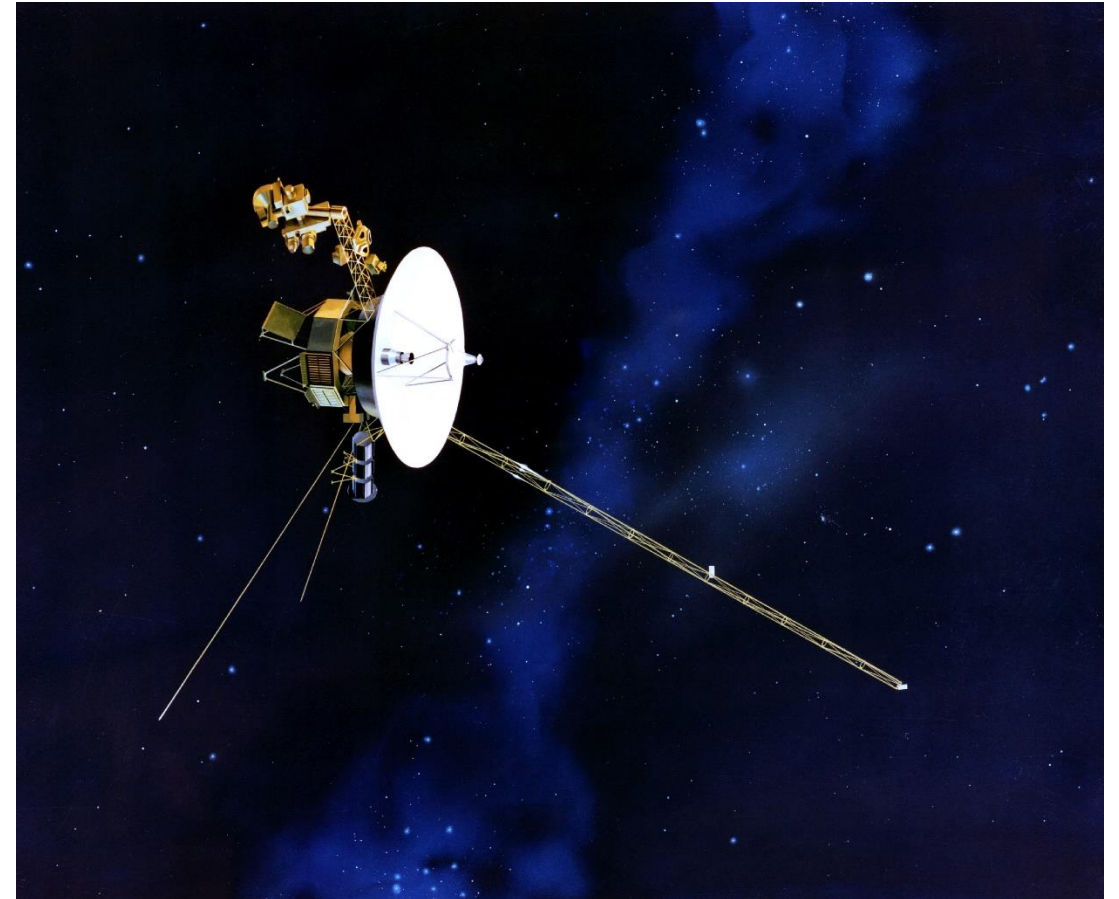


Probleme bei Forschungsdaten

„Fehlerhafte Daten von Voyager 1: Zähe Suche nach technischer Dokumentation

Seit zwei Monaten sendet Voyager 1 falsche Daten. Die Suche nach der Ursache wird dadurch erschwert, dass technische Dokumente zu der Sonde verschollen sind.“

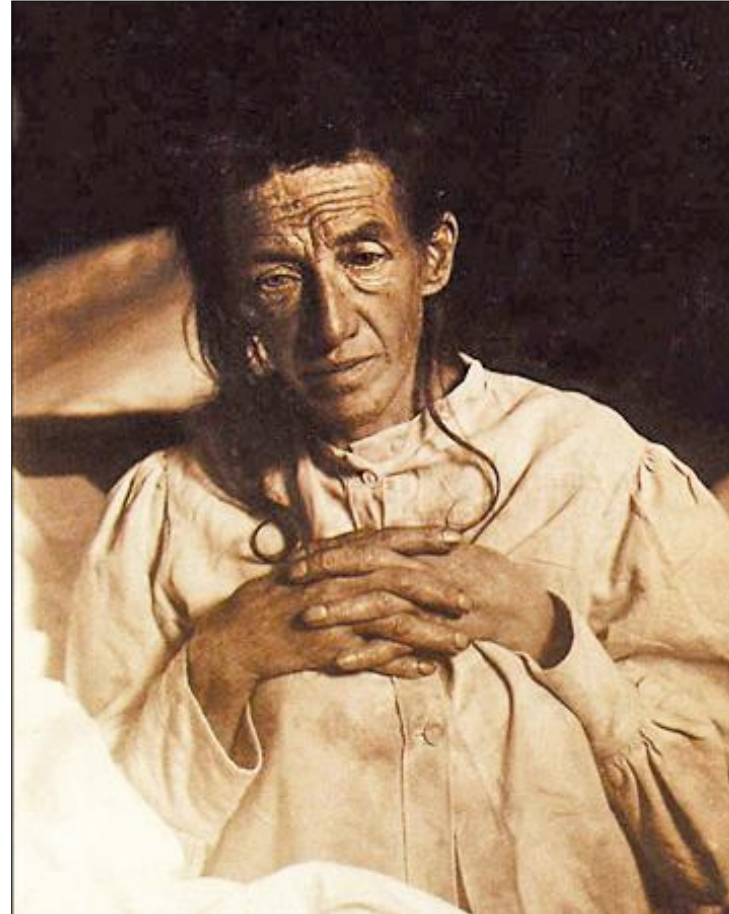
(<https://www.heise.de/news/Fehlerhafte-Daten-von-Voyager-1-Zaehe-Suche-nach-technischer-Dokumentation-7170603.html>)



NASA/JPL, Artist's concept of Voyager in flight, D=2194, gemeinfrei.

Probleme bei Forschungsdaten

- „Manipulierte“ Alzheimer-Daten könnten die Forschung seit 2006 in die Irre geführt haben
- <https://www.telegraph.co.uk/news/2022/07/21/manipulated-alzheimers-data-may-have-misled-research-16-years/>



Auguste Deter, Alois Alzheimer Patient im November 1901, erster beschriebener Patient mit Alzheimer-Krankheit, https://commons.wikimedia.org/wiki/File:Auguste_D_aus_Marktbreit.jpg, gemeinfrei.

Allgemeine Literaturempfehlungen

- Putnings, Markus, Heike Neuroth, und Janna Neumann (Hrsg.) 2021: Praxishandbuch Forschungsdatenmanagement, <https://doi.org/10.1515/9783110657807>.
- Büttner, Stephan; Enke, Harry; Helbig, Kerstin; Hobohm, Hans-Christoph; Kindling, Maxi; Ludwig, Jens; Neumann, Janna; Pampel, Heinz; Schwarz, Karin (2016): Lehrbuch Forschungsdatenmanagement; https://handbuch.tib.eu/w/Lehrbuch_Forschungsdatenmanagement.

Linkempfehlungen

- <https://www.forschungsdaten.info> (Universität Konstanz)
- <https://www.forschungsdaten.org> (freie Community)
- <https://rdm.mpdl.mpg.de> (MPDL-Service)

Aktivierung zur letzten Sitzung

1. Was sind mögliche Beispiele von Forschungsdaten?
2. Wie kann man Forschungsdaten klassifizieren?
3. Welche rechtliche Rahmen können die Arbeit mit Forschungsdaten vorgeben?
4. Welche Probleme bei der Arbeit mit Forschungsdaten können auftreten?

Teil 2

Forschungsdateninfrastruktur

Inhalt von Teil 2

1. Forschungsdateninfrastruktur
2. Forschungsdatenrepositorien
3. Beispiele von Forschungsdatenrepositorien
4. Datenmanagementpläne
5. FAIR-Prinzipien
6. Metadaten und Lizenzen
7. Sortieren und Ordnen

Forschungsdateninfrastruktur

Infrastruktur?

- Vergleich Verkehrsinfrastruktur
 - Straßen, Brücken
 - Schienen, Wasserwege
 - Tankstellennetz, Verkehrsfunk, GPS
 - Straßenverkehrsordnung, Straßenwacht, ADAC
 - Räumdienst, Mautsystem, KFZ-Steuer
 - Rad- und Autoindustrie, Verkehrsministerium, ÖPNV



Bundesarchiv, Bild 183-F0307-0001-046, Straßenverkehr in Leipzig, 1967, CC BY SA 3.0.

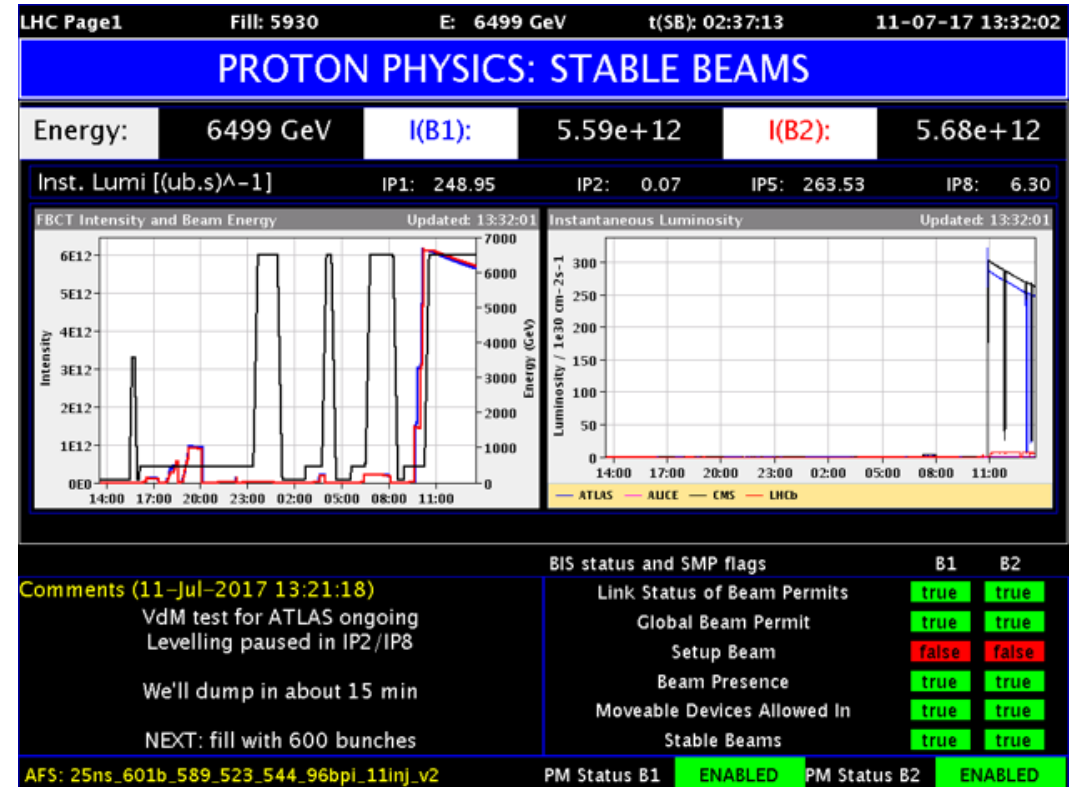
Forschungsdateninfrastruktur

- Wiss. (Groß-)Anlagen
- Metadaten, Standards, Identifikatoren
- Repositorien, Registries, Datenjournale
- Rechtsgrundlagen

- Nationale Forschungsdateninfrastruktur
- European Open Science Cloud
- ...

Wissenschaftliche Großanlagen

- Large Hadron Collider
- Teilchenbeschleuniger am Europäischen Kernforschungszentrum CERN
- <https://op-webtools.web.cern.ch/vistar/vistars.php>
- „No beam before March 2024“



Wissenschaftliche Großanlagen

Polarstern

- Forschungsschiff des Alfred-Wegener-Institut für Polar- und Meeresforschung
- <https://www.awi.de/expedition/schiffe/polarstern.html>
- Data Flow Framework:
<https://www.awi.de/en/about-us/service/computing-centre/data-flow-framework.html>



Janek Uin: MOSAIC-Expedition, Polarwinter 2019/2020, CC BY 4.0.

Wissenschaftliche Großanlagen

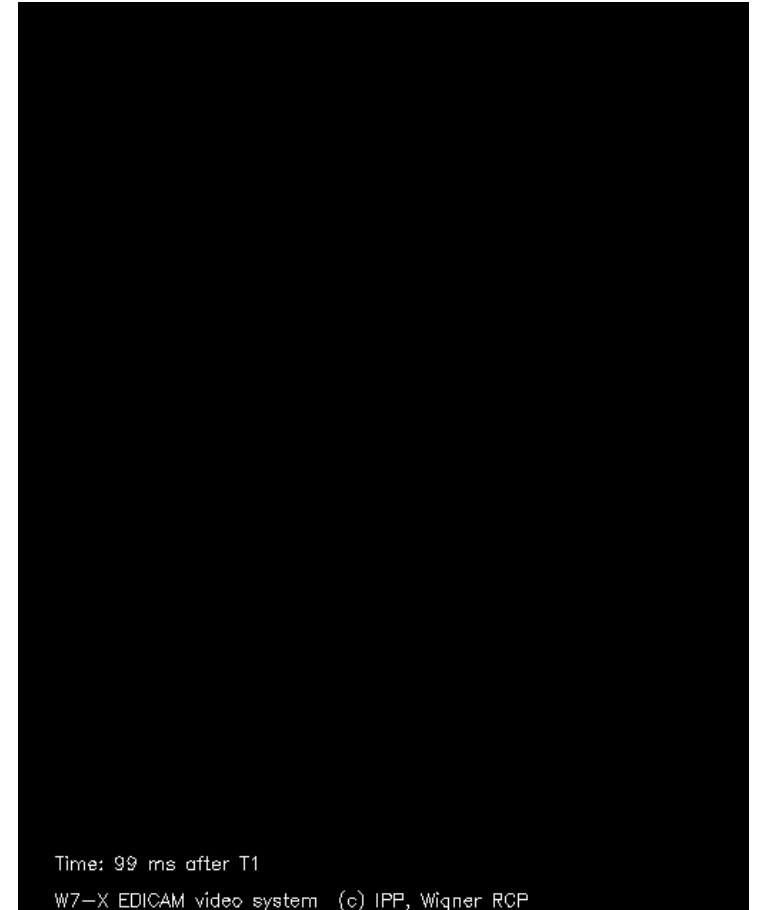
- Antennen des Atacama Large Millimeter/submillimeter Array (ALMA)
- auf dem Chajnantor Plateau in den chilenischen Anden
- <https://www.eso.org/public/teles-instr/alma/>



ESO/C. Malin, The Atacama Large Millimeter/submillimeter Array (ALMA) by night, under the Magellanic Clouds, 2013, CC BY 4.0, <https://www.eso.org/public/images/ann13016a/>.

Wissenschaftliche Großanlagen

- Wendelstein 7-X
- einer der weltweit größten Fusionsanlage vom Typ „Stellarator“
- Ziel: Kraftwerkseignung dieses Bautyps untersuchen
- <https://www.ipp.mpg.de/wendelstein7x>



IPP/Wigner RCP

Wissenschaftliche Großanlagen

- Erheben/Erzeugen/Verteilen Daten
- Erweitern Möglichkeiten
- Beschleunigen Prozesse
- Zwingen zur Kooperation
- ...

Forschungsdatenrepositorien

Einteilung von Forschungsdatenrepositorien

1. Fachlich

- [PsychData](#)
- [Pangaea](#)
- [G-Node](#)
- ...

2. Institutionell

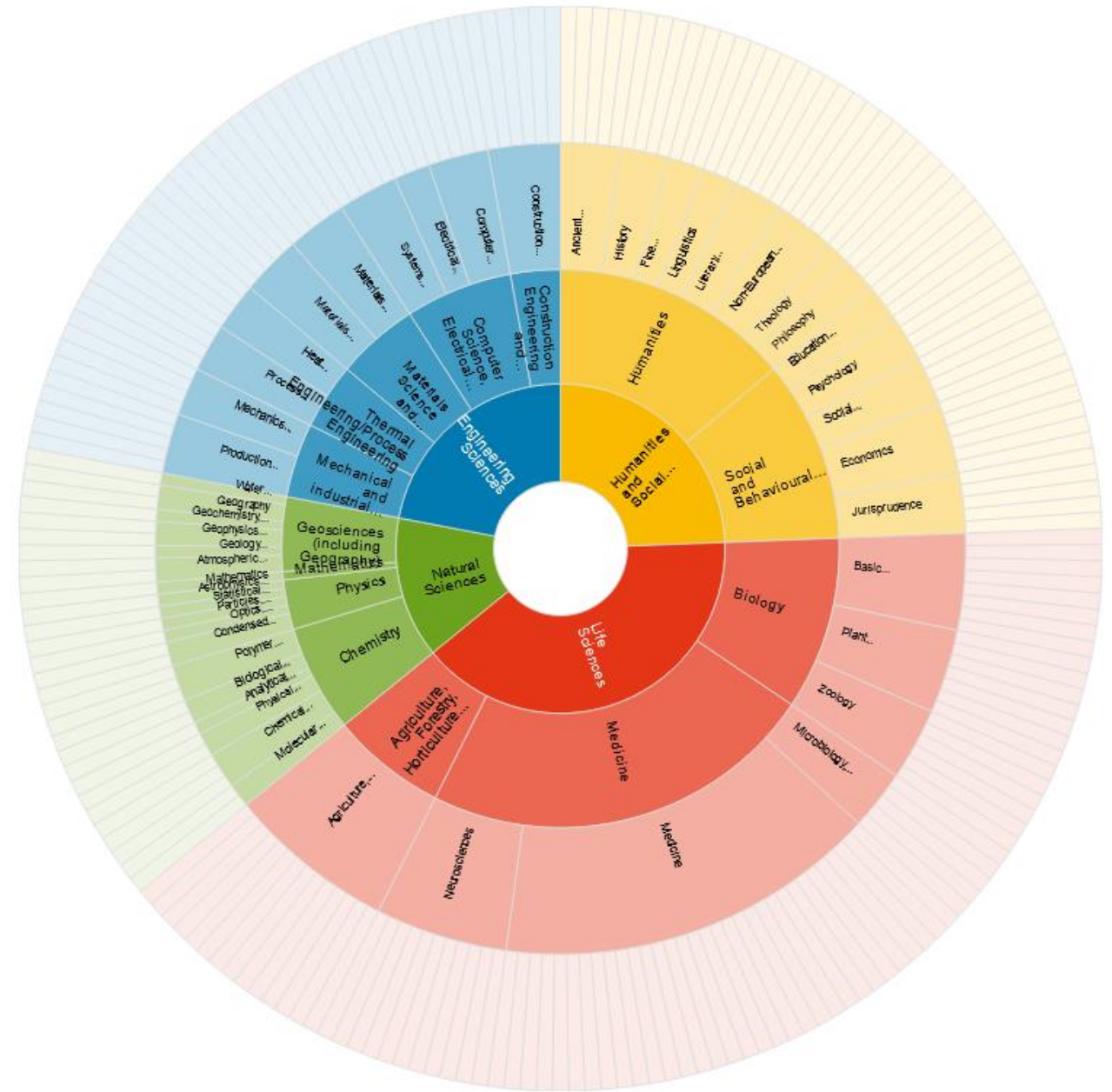
- [Open Data LMU](#)
- [Edinburgh DataShare](#)
- [Edmond](#)
- ...

3. Allgemein

- [Zenodo](#)
- [Dryad](#)
- [Radar](#)
- [Figshare](#)
- [Mendeley Data](#)
- ...

Forschungsdaten-repositorien

- Repositorien finden:
- re3data (<https://www.re3data.org>)
- Bild: Search by Subject, CC BY 4.0



Zertifizierungen

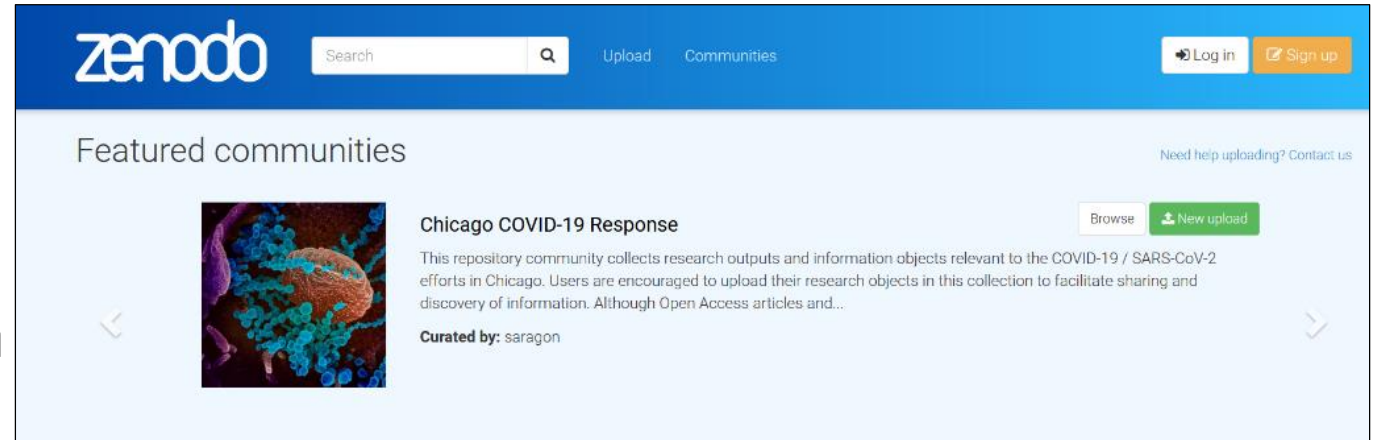
- CoreTrustSeal (<https://www.coretrustseal.org>)
- CLARIN certificate B (<https://www.clarin.eu>)
- DIN 31644 (<https://www.beuth.de/en/standard/din-31644/147058907>)
- DINI Certificate (<https://dini.de/dienste-projekte/dini-zertifikat/>)
- RatSWD (<https://www.konsortswd.de/ratswd/>)
- ...

Beispiele für Datenrepositorien

Repositories: Example 1

Zenodo

- <https://zenodo.org>
- vom CERN
- generisches Repositorium



CERN Data Centre & Invenio, CC BY 4.0, <https://zenodo.org>.

Repositories: Example 2

Discuss.Data

- <https://discuss-data.net>
- Repository für die post-sowjetische Region
- betrieben durch die Universität Bremen und die SUB Göttingen
- fachspezifisches Repository



Maidan Oral History Archive

I. Title of the Data Collection

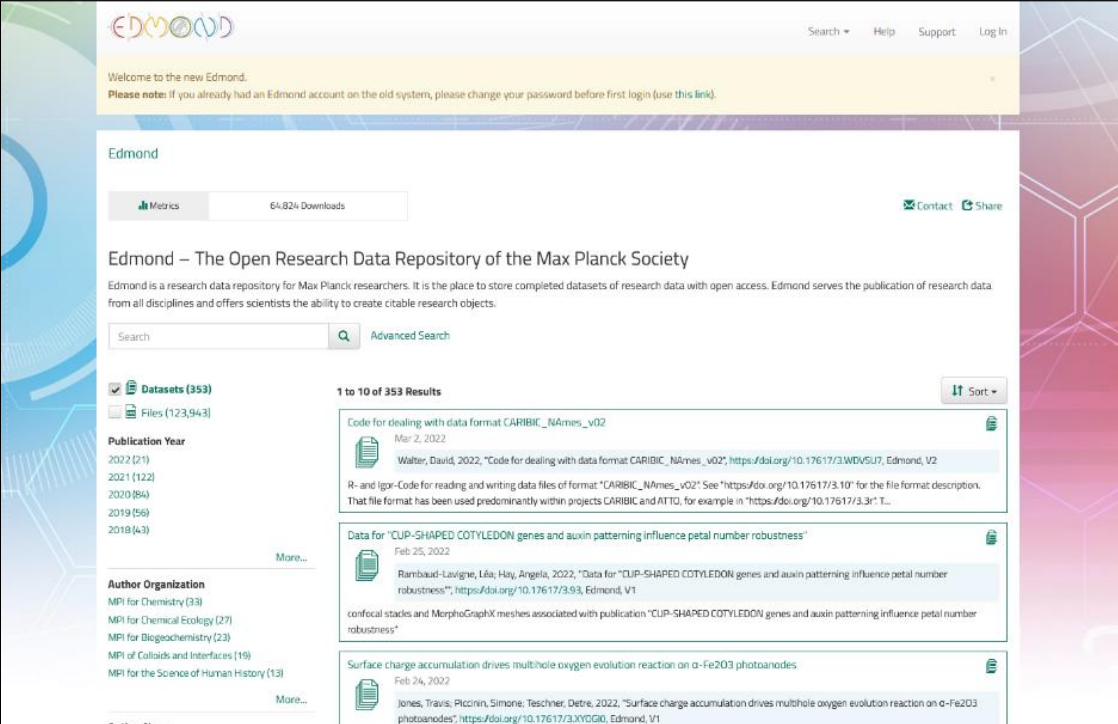
Maidan Oral History Archive

The Foundation to Preserve the History of Maidan (FPHM) (2021): Maidan Oral History Archive, v. 1.0, Discuss Data, <https://doi.org/10.48320/F5509D39-B2D4-409D-AD37-9E72570206AD>

Repositories: Example 3

Edmond

- <https://edmond.mpdl.mpg.de>
- durch die Max Planck Digital Library für die Max-Planck-Gesellschaft
- Institutionelles Repositorium

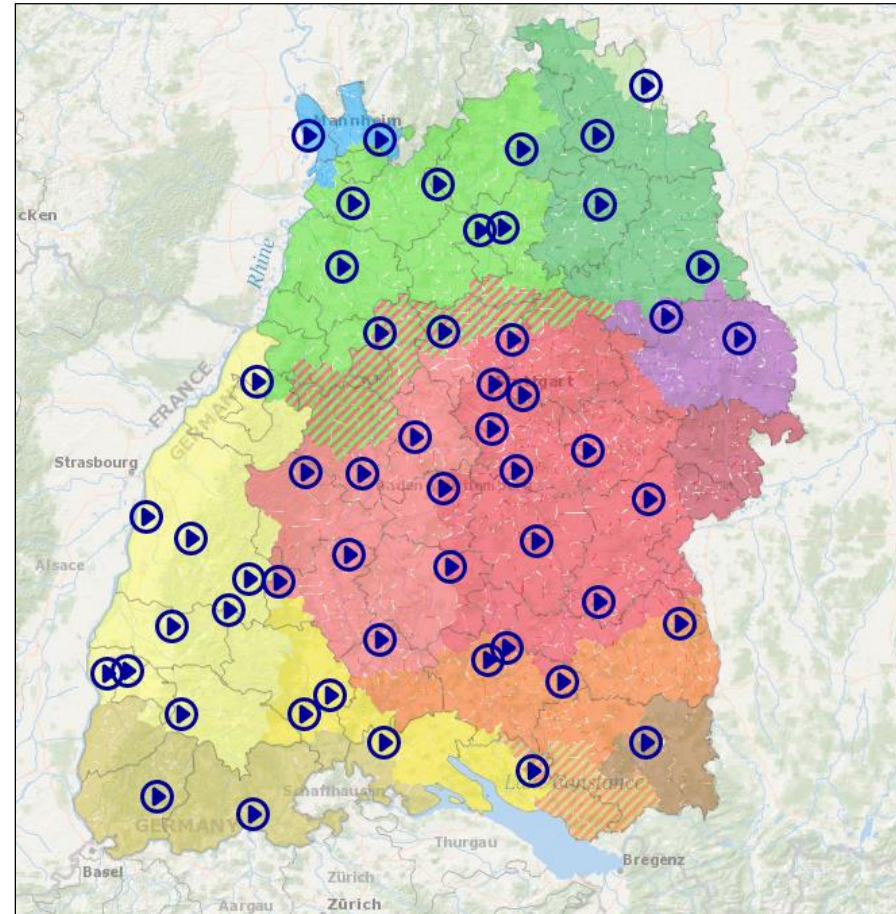


The screenshot displays the Edmond website interface. At the top, there is a navigation bar with the Edmond logo, a search bar, and links for Help, Support, and Log In. Below the navigation bar, a yellow banner contains a welcome message and a note about password changes for existing users. The main content area features a header with the Edmond logo and a metrics bar showing 64,824 Downloads. A search bar is present, and the page title is "Edmond – The Open Research Data Repository of the Max Planck Society". The main content area is divided into two columns. The left column contains filters for "Datasets (353)" and "Files (123,943)", with a "Publication Year" filter showing counts for 2022 (21), 2021 (122), 2020 (86), 2019 (56), and 2018 (43). The right column displays search results, with the first result being "Code for dealing with data format CARIBIC_NAmnes_v02" by David Walter, dated Mar 2, 2022. The second result is "Data for 'CUP-SHAPED COTYLEDON genes and auxin patterning influence petal number robustness'" by Rambaud-Lavigne, Léa; Hay, Angela, dated Feb 25, 2022. The third result is "Surface charge accumulation drives multihole oxygen evolution reaction on α -Fe₂O₃ photoanodes" by Jones, Travis; Piccirin, Simone; Teschner, Detre, dated Feb 24, 2022.

Repositories: Example 4

Sprachatlas Baden-Württemberg

- <https://escience-center.uni-tuebingen.de/escience/sprachatlas/#8/48.402/7.498>
- Eberhard Karls Universität Tübingen
- Audio-Daten mit Karte zu gesprochenen Sprachen in Baden-Württemberg



Recherche-Aufgabe

Frage:

Sind die Forschungsdatenrepositorien an Ihrer Institution zertifiziert?
Wenn ja, wie wird damit geworben?

Falls kein Datenrepositorium vorhanden, wählen Sie ein interessantes Repository aus [re3data](#) aus und recherchieren Sie dazu.

Methode:

Individuelle Online-Recherche

Dauer:

5 Minuten

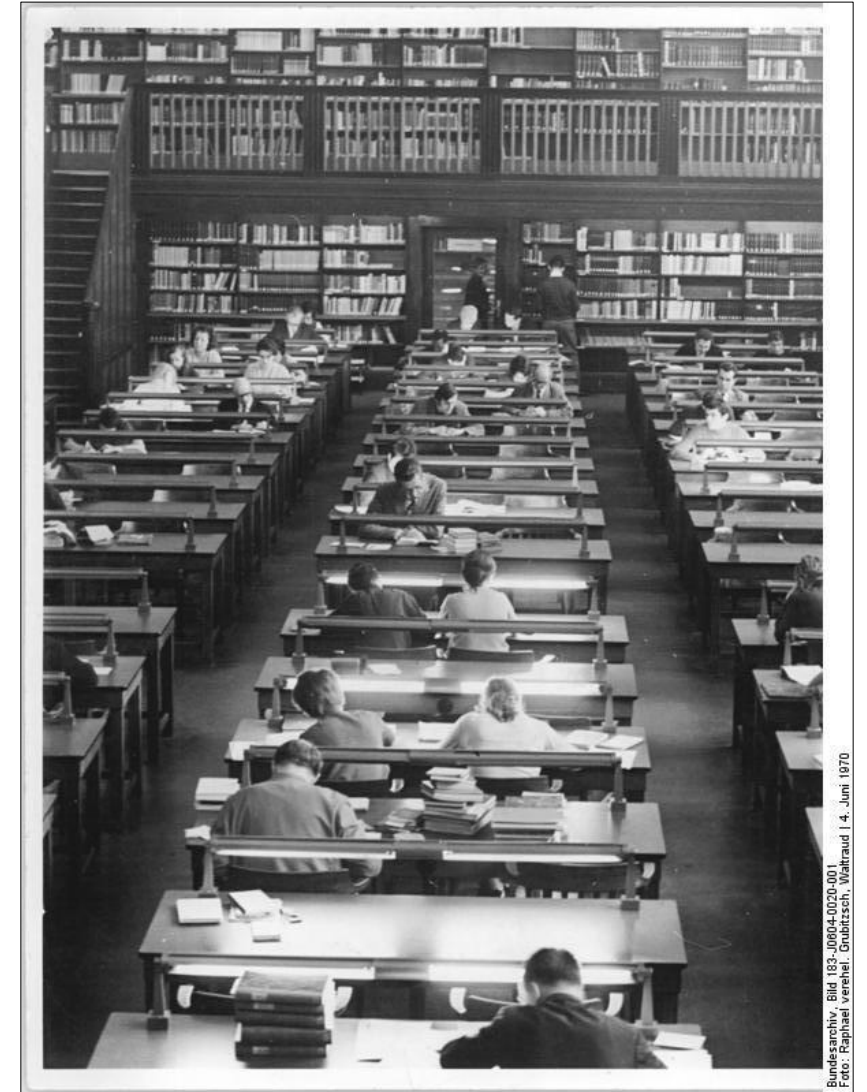
Auswertung:

Gemeinsame Diskussion der Ergebnisse

Datenmanagementpläne

Wieso einen DMP schreiben?

- zunehmend von Förderorganisationen gefordert (z.B. Horizon Europe)
- Kosten für das Datenmanagement (Software, Hardware, technisches Fachwissen) sind häufig förderfähig
- Eigenes Projektmanagement
- Teil eines Beratungsgesprächs
- ...



Bundesarchiv, Bild 183-J0604-0020-001 / Raphael (verehel. Grubitzsch), Waltraud / CC BY SA 3.0.

Wie schreibt man einen DMP?

1. Brainstorming
2. Vorlage nutzen
3. Digitales Werkzeug nutzen

Vorlagen für DMPs

- [Horizon Europe DMP Template by the European Commission, 2021](#)
- [Horizon 2020 DMP Template by the European Commission, 2018](#)
- [ERC Grants DMP Template by the European Research Council, 2021](#)
- [DMP Template by the Swiss National Science Foundation, 2017](#)

How to DMP

- Argos (<https://argos.openaire.eu>)
 - Data Stewardship Wizard (<https://ds-wizard.org>)
 - DMPTool (<https://dmptool.org>)
 - DMPOnline (<https://dmponline.dcc.ac.uk>)
 - RDMO (<https://rdmo.aip.de>)
-
- Helbig, K., Paul-Stüve, T., & Rex, J. (2021): DMP-Toolguide, <http://doi.org/10.5281/zenodo.4632308>.

Beispiele für DMPs

- Digital Curation Centre Example DMPs and Guidance: <http://www.dcc.ac.uk/resources/data-management-plans/guidance-examples>
- LIBER Europe DMP Catalogue: <https://libereurope.eu/dmpcatalogue/>
- Examples for Horizon 2020 DMPs by the University of Vienna: <https://phaidra.univie.ac.at/search#?page=1&pagesize=10&collection=:1140797>

Aspekte eines DMPs

Die DMP umfassen in der Regel die folgenden Aspekte:

1. Beschreibung und Erhebung von Daten
2. Dokumentation und Metadaten
3. Speicherung und Sicherung
4. Rechtliche und ethische Anforderungen
5. Gemeinsame Nutzung und langfristige Aufbewahrung von Daten
6. Zuständigkeiten und Ressourcen

Mögliche Fragen eines DMPs

Die folgenden Fragen können in einem DMP beantwortet werden:

- Welche Daten werden erhoben/genutzt?
- Wie werden die Daten verarbeitet?
- Welche Standards werden verwendet?
- Wie werden die Daten dokumentiert?
- Wo werden die Daten (offen) gespeichert und wann?
- Was, wenn überhaupt, spricht gegen eine Veröffentlichung?
- Welche Kosten entstehen durch das Forschungsdatenmanagement?

Checkliste von der ETH Zürich (https://www.dlcm.ch/download_file/force/66/371)

Kosten für Datenmanagement

- OpenAIRE: <https://www.openaire.eu/how-to-comply-to-h2020-mandates-rdm-costs>
- Data Management Costing Tool by the UK Data Archives: <https://www.ukdataservice.ac.uk/manage-data/plan/costing>

Was fällt Ihnen zu diesem Bild ein?

T-1
Mc Crue Shoals Lighted Whistle Station Buoy 2

4 of 5

L.	W.	H.	T	L	W	L _n	L/W	L/H	L/W
15.0	5.6	8.5	29.1	51.5	19.2	29.2	2.68	1.76	1.52
13.0	4.0	7.8	24.8	52.5	16.1	31.4	2.25	1.67	1.95
14.5	4.1	7.3	25.9	56.0	15.8	28.2	3.54	1.99	1.78
12.3	4.4	7.6	24.3	50.6	18.1	31.3	2.80	1.62	1.73
13.6	4.5	7.3	25.4	53.5	17.7	28.8	3.02	1.86	1.62
11.5	4.7	6.2	22.4	51.3	21.0	27.6	2.45	1.85	1.32
11.6	3.9	6.0	21.5	54.0	18.1	27.9	2.98	1.93	1.54
13.2	5.5	8.4	27.1	48.7	20.3	31.0	2.40	1.57	1.53
10.4	3.6	5.6	19.6	53.1	18.3	28.6	2.89	1.86	1.56
11.1	4.5	6.1	21.7	51.2	20.7	28.1	2.47	1.82	1.35
10.0	3.0	5.0	18.0	55.6	16.6	27.8	3.33	2.00	1.66
9.6	3.4	5.6	18.6	51.5	18.3	30.1	2.42	1.71	1.65
8.5	3.4	4.1	16.0	53.0	21.2	25.7	2.50	2.07	1.21
8.5	3.5	5.5	17.5	48.5	20.0	21.4	2.42	1.54	1.57
7.2	2.2	4.5	13.9	51.7	15.8	32.4	3.28	1.60	2.04
8.4	2.6	4.6	15.0	53.8	16.6	29.5	2.13	1.83	1.77
8.2	3.5	4.4	16.1	50.9	21.7	27.2	2.34	1.86	1.25
7.5	2.2	4.5	14.2	52.8	15.5	31.7	3.40	1.66	2.04
6.6	2.1	3.3	12.0	55.0	17.5	27.5	3.14	2.00	1.57
6.1	2.1	3.2	11.4	53.5	18.4	28.1	2.90	1.91	1.52

Louis W. Hutchins, Buoy fouling survey, research data, Smithsonian Institution Archives, [https://commons.wikimedia.org/wiki/File:Buoy_fouling_survey_research_data_\(15\)_1943-1947_\(20323535739\).jpg](https://commons.wikimedia.org/wiki/File:Buoy_fouling_survey_research_data_(15)_1943-1947_(20323535739).jpg). no known copyright restrictions.

FAIR-Prinzipien

FAIR-Prinzipien

Findable
Accessible
Interoperable
Reusable

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016).
<https://doi.org/10.1038/sdata.2016.18>.

FAIR-Prinzipien

Findable

- F1. (Meta)data are assigned a globally unique and persistent identifier
- F2. Data are described with rich metadata (defined by R1 below)
- F3. Metadata clearly and explicitly include the identifier of the data they describe
- F4. (Meta)data are registered or indexed in a searchable resource

Interoperable

- I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (Meta)data use vocabularies that follow FAIR principles
- I3. (Meta)data include qualified references to other (meta)data

Accessible

- A1. (Meta)data are retrievable by their identifier using a standardised communications protocol
 - A1.1 The protocol is open, free, and universally implementable
 - A1.2 The protocol allows for an authentication and authorisation procedure, where necessary
- A2. Metadata are accessible, even when the data are no longer available

Reusable

- R1. (Meta)data are richly described with a plurality of accurate and relevant attributes
 - R1.1. (Meta)data are released with a clear and accessible data usage license
 - R1.2. (Meta)data are associated with detailed provenance
 - R1.3. (Meta)data meet domain-relevant community standards

Test: F-Uji



<https://www.f-uji.net>

i.e. <https://doi.org/10.5281/zenodo.5566761>

Identifiser aus Recherche-Repository in F-Uji

Metadaten und Lizenzen

Metadaten

Author ?	Giorgetta, Marco  Max Planck Institute for Meteorology 
Study Type ?	simulation/modelling
Description ?	This dataset provides additional information related to the publication "The ICON-A model for direct QBO simulations on GPUs (version icon-cscs:baf28a514)" submitted to Geoscientific Model Development (GMD). Details for the content are given in the README.pdf file.
Subject ?	Geosciences
Keyword ?	Global atmospheric modelling storm resolving GPU port ICON atmosphere model weak and strong scaling quasi-biennial oscillation
Depositor ?	Giorgetta, Marco
Deposit Date ?	2022-03-28
Related Publication ?	Giorgetta, M. A., Sawyer, W., Lapillonne, X., Adamidis, P., Alexeev, D., Clément, V., Dietlicher, R., Engels, J. F., Esch, M., Franke, H., Frauen, C., Hannah, W. M., Hillman, B. R., Kornblueh, L., Marti, P., Norman, M. R., Pincus, R., Rast, S., Reinert, D., Schnur, R., Schulzweida, U., and Stevens, B.: The ICON-A model for direct QBO simulations on GPUs (version icon-cscs:baf28a514), submitted to Geosci. Model Dev., 2022.

Metadaten von <https://doi.org/10.17617/3.5CYUFN>

Metadaten

- Bibliographische Metadaten
 - [DC](#) – Dublin Core Metadata Initiative
 - [SKOS](#) – Simple Knowledge Organization System
- Fachliche Metadaten
 - [RDF](#) – Resource Description Framework
 - [Metadaten-Standards](#)
 - ([Metadaten-Registries](#))
- Technische Metadaten, Lizenzmetadaten, Provenienzmetadaten

Persistente Identifikation

- URL/URI: <http://zuse.zib.de/collection/wl3XoEDHO8v0ImCa/item/VqNgKUacPrlhqPKu>
- URN: urn:nbn:de:bsz:25-opus-14124 (Resolver z.B. <https://nbn-resolving.org>)
- Handle: <hdl:11858/00-001M-0000-0019-D20F-6>
- DOI: <doi:10.1088/0004-637X/715/2/1453>
- ORCID: <0000-0002-2880-8947>
- ROR: <0061msm67>

Digital Object Identifier

- für digitale Objekte
- <https://www.doi.org>
- DOIs für Forschungsdaten durch DataCite (im Gegensatz zu Crossref)



Public domain

Open Researcher and Contributor ID

- für Personen
- <https://orcid.org>



Public domain, CC0

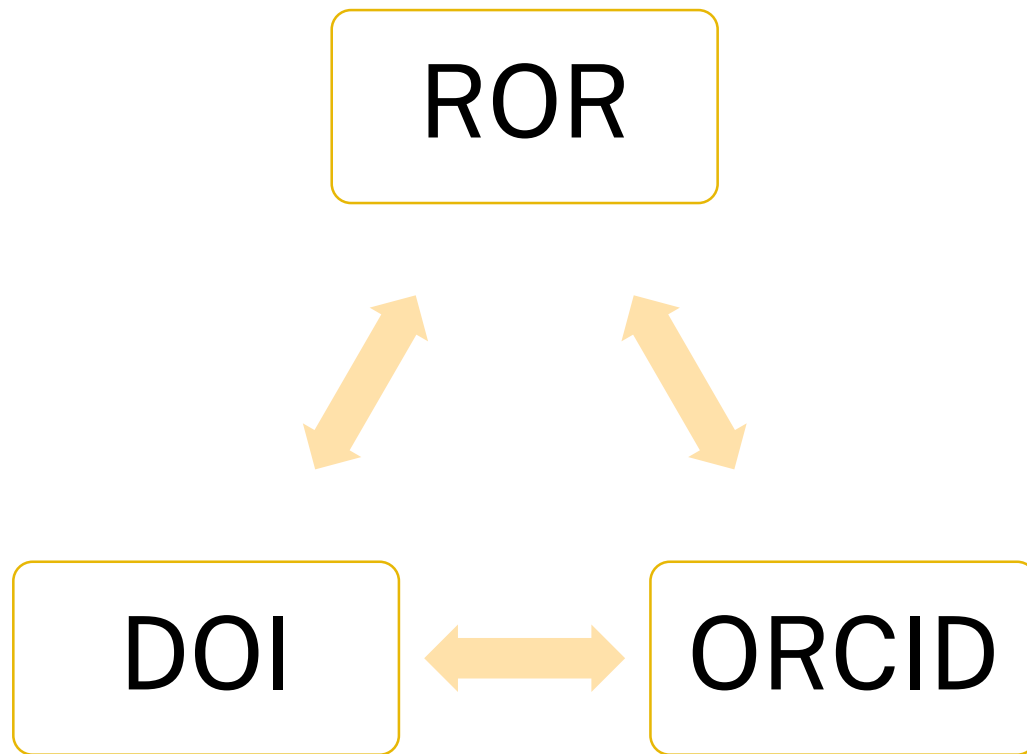
Research Organization Registry

- für Organisationen
- <https://ror.org>



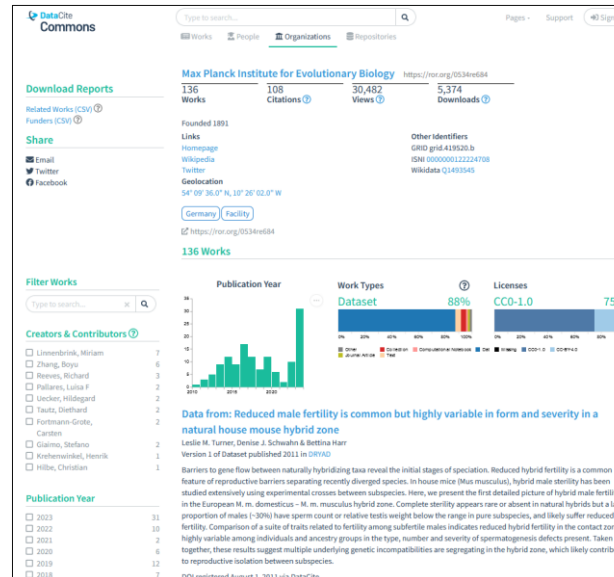
Research Organization Registry, CC BY 4.0,
<https://doi.org/10.5281/zenodo.4701802>

Identifier Network



Max Planck Institute for Evolutionary Biology:

- <https://ror.org/0534re684>
- <https://commons.datacite.org/ror.org/0534re684>



Live Demonstration

Beispiel Ludwig-Maximilians-Universität München:

- <https://ror.org/05591te55>
- <https://commons.datacite.org/ror.org/05591te55>

Daten-Lizenzen

- Creative Commons (<https://creativecommons.org>)
 - CC 0
 - CC BY
- Open Data Commons (<https://opendatacommons.org>)
 - ODC-By
 - PDDL
- Datenlizenz Deutschland

Research Data ScaryTales

Frage:

Was zeigt das Beispiel?

Methode:

Quiz mit Geschichten von FDM Thüringen

Dauer:

9min (3x3min)

Lang et al. (2023): Research Data ScaryTales, v4.1, Zenodo,
<https://doi.org/10.5281/zenodo.10281701>, CC0.

Aus allen Wolken gefallen

Nachdem seine Kollegen zusammen in die Wolken gestarrt hatten, blicken sie ihn finster an.

Eine Gruppe von Studierenden arbeitete am Ende des Semesters an einem Abschlussbericht für eine Lehrveranstaltung. Das Dokument wurde gemeinsam in einem iCloud Drive bearbeitet und stand kurz vor der Fertigstellung. Einer der Studierenden löschte den Bericht in dem Glauben, dass es sich dabei um einen alten Entwurf handelte. Zu dieser Zeit gab es keine Funktion zur Wiederherstellung von Daten und die Gruppe musste den Bericht von Grund auf neu schreiben. Es ist nicht überraschend, dass die restlichen Mitglieder der Gruppe in dieser Zeit nicht unbedingt gut auf ihren Kommilitonen zu sprechen waren.

Das Beispiel zeigt sehr gut, dass die Speicherung an nur einem Ort nicht ausreicht, um eine sichere Aufbewahrung der Daten zu gewährleisten. Generell sollten Daten nach der 3-2-1-Regel gesichert werden. Dementsprechend sollte es mindestens 3 Kopien der Daten auf 2 verschiedenen Speichermedien geben, wobei eine davon an einem externen Ort verwahrt werden sollte. Außerdem sollte man bei der gemeinsamen Arbeit in einem zentralen Speicherort besonders umsichtig beim Entfernen von Dateien sein.

Lang et al. (2023): Research Data ScaryTales, v4.1, Zenodo,
<https://doi.org/10.5281/zenodo.10281701>, CC0, scarytales_de.pdf, S. 5.

Bindende Verträge

Mit der abschließenden Aufzählung wurde einfach zu viel versprochen.

Eine mit Rückgriff auf Mustererklärungen erstellte informierte Einwilligungserklärung für eine Studie nutze eine Formulierung, in der die zugelassenen Verwendungsmöglichkeiten aufgezählt und eine Beschränkung auf die genannten Möglichkeiten versprochen wurden. Konkret enthielt der Text die Aussage "die Daten werden ausschließlich in der wie vorher beschriebenen Form verarbeitet". Da die (anonymisierte) Veröffentlichung in der Liste fehlte, konnten die Daten nicht publiziert werden. Die sehr strenge Formulierung war für beide Seiten bindend.

Es ist sinnvoll alle rechtlich bindenden Texte vor ihrer Nutzung von Dritten mit entsprechenden Erfahrungen gegenlesen zu lassen, um zu prüfen, dass die rechtlichen Konsequenzen der Formulierung tatsächlich den gewünschten Wirkungen entsprechen. Die Nachnutzungsbedingungen der erhobenen Forschungsdaten sollten so formuliert sein, dass sie der Veröffentlichung oder Weiternutzung der Daten in anderen Projekten nicht im Wege stehen.

<https://doi.org/10.18450/dataman/98>

Lang et al. (2023): Research Data ScaryTales, v4.1, Zenodo,
<https://doi.org/10.5281/zenodo.10281701>, CC0, scarytales_de.pdf, S. 57.

Seltsam – Preise haben Geburtstag

Sie konnte ihre Kostenanalyse einfach nicht beenden, da ihr immer wieder Geburtstage dazwischenkamen.

Beim Versuch eine Kostenanalyse zu erstellen, war die zuständige Mitarbeiterin überrascht als in ihrer Tabelle immer wieder abwegige Ergebnisse am Ende ihrer Berechnungen standen. Nach kurzer Recherche war klar, dass beim Import der CSV-Datei für die Nutzungsstatistik elektronischer Medien in Excel aus den Preisen automatisch Datumsangaben gemacht wurden. Nutzungsstatistiken für elektronische Medien werden typischerweise als CSV-Dateien ausgeliefert. Werden diese in den Grundeinstellungen in eine Excel-Datei importiert, werden Preise teilweise in Datumsangaben umgewandelt, obwohl dies nicht gewünscht ist. Das verhindert dann in der Folge eine sinnvolle Interpretation der Nutzungsdaten insgesamt. So lassen sich z.B. keine durchschnittlichen Nutzungskosten mehr berechnen, wenn ein Teil der Preise nicht ausgelesen werden kann.

Das Beispiel zeigt, dass beim Import von Daten in Tabellenkalkulationsprogramme sorgfältig auf die korrekte Formatierung der Zellen geachtet werden muss, da es sonst zu automatischen Formatierungsänderungen kommen kann.

Lang et al. (2023): Research Data ScaryTales, v4.1, Zenodo,
<https://doi.org/10.5281/zenodo.10281701>, CC0, scarytales_de.pdf, S. 2.

Literaturempfehlungen

- allgemein: Fotis Jannidis, Hubertus Kohle, und Malte Rehbein (Hrsg.): Digital Humanities: eine Einführung, Stuttgart 2017.
- zu Repositorien: Friedrich Summann und Andreas Czerniak: „Die deutsche und europäische Repositorien-Landschaft im Überblick“, gehalten auf den Open Access Tagen 2020 (OAT2020), Bielefeld, 17. September 2020 <https://doi.org/10.5281/zenodo.4034075>.
- zu Metadaten: Jana Baur: „FDM erklärt – Metadaten & Metadatenstandards“, 13. August 2021. <https://blog.rwth-aachen.de/forschungsdaten/2021/08/13/fdm-erklaert-metadaten-metadatenstandards/>.
- zu Lizenzen: Brettschneider et al. (2021): Offene Lizenzen für Forschungsdaten: Rechtliche Bewertung und Praxistauglichkeit verbreiteter Lizenzmodelle. O-Bib. Das Offene Bibliotheksjournal 8(3), S. 1–22, <https://doi.org/10.5282/o-bib/5749>.

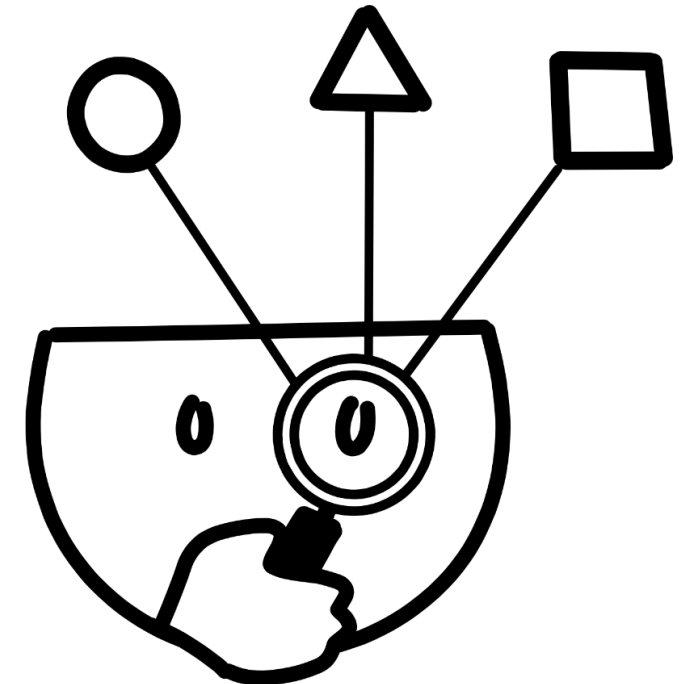
Sortieren und Ordnen

Dateibenennung – Gute Praxis

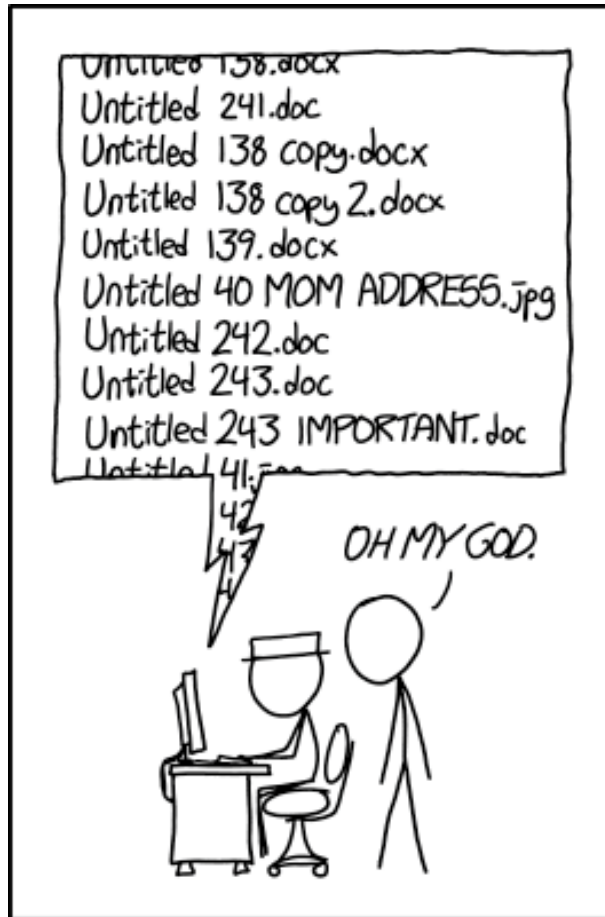
- Fügen Sie relevante **Informationen** in die Dateinamen ein. Dazu können Projekt, Forscher (Initialen), Experiment, Ort, Datum, Parameter, Status der Daten (Rohdaten, verarbeitete Daten usw.), Dateiversion oder andere Informationen gehören.
- Verwenden Sie JJJJ-MM-TT für **Datumsangaben** (z. B. 2016-08-02) wie in ISO 8601 und setzen Sie das Datum an den Anfang oder das Ende von Dateinamen, um die chronologische Sortierung zu erleichtern.
- Verwenden Sie bei Bedarf führende **Nullen** für andere Zahlen (001, 002, 011, 114 statt 1, 2, 11, 114), je nach der zu erwartenden Anzahl von Dateien, und auch, um eine korrekte Sortierung der Dateien zu ermöglichen.
- Verwenden Sie **Bindestriche** (Dateiname.xxx) oder **Unterstriche** (Dateiname.xxx), um Teile eines Dateinamens zu trennen.
- Erstellen Sie eine README-Datei, in der die Dateinomenklatur (einschließlich der Bedeutung von Akronymen oder Abkürzungen), die Dateiorganisation und die Versionsverwaltung erklärt werden. Speichern Sie diese Datei oben in einer Ordnerstruktur, um den Zugriff zu erleichtern.
- **Gutes Beispiel:** 20140618_exp08_co2data_raw_v01.csv

Dateibenennung – Vermeiden

- Die Verwendung von **Sonderzeichen** wie ~ ! / \ @ # \$ % ^ & * () ` ; < > ? , [] { } ' " |, da sie in bestimmten Betriebsumgebungen besondere Bedeutungen haben können. Vermeiden Sie auch ä ö ü ß å ø ñ oder ähnliches.
- Verwendung anderer **Punkte** als vor der Dateierweiterung.
- Die Verwendung von **Leerzeichen** in Dateinamen kann dazu führen, dass sie bei der Dateiübertragung abgeschnitten werden (wobei auch die Dateierweiterung verloren geht und die Dateien möglicherweise unlesbar werden).
- Übermäßig **lange Dateinamen** (empfohlen werden maximal 32 Zeichen). Dies kann zu Kompromissen gegenüber der obigen Empfehlung führen.



Dateibenennung – Beispiele



PROTIP: NEVER LOOK IN SOMEONE ELSE'S DOCUMENTS FOLDER.

xkcd. "Dokumente.",
<https://xkcd.com/1459/>,
CC-BY-NC 2.5.

Zu bevorzugen	Zur Vermeidung von
website-texts-2020-05-v15.docx	Website-Texte-Mai-finalFinal2_neu.docx
Digitized_XY-ZZ_E-2_HT-493887.tiff	Bild0001.tiff
Digitized_XY-ZZ_M-5_LS-345-c.tiff	Bild0002.tiff
Digitized_XY-ZZ_E-1_M-296778.tiff	Bild0003.tiff
Umfrage_OpenAccess_bereinigt_v04.csv	OA-Ergebnisse-sauber-4.csv

Beispiele aus <https://www.fu-berlin.de/en/sites/forschungsdatenmanagement/in-der-praxis/durchfuehrung/benennung.html>.

Struktur der Dateiorganisation

- Bevor Sie mit dem Sammeln von Dateien oder Daten beginnen, sollten Sie auf jeden Fall eine **Konvention** festlegen, um einen Rückstau an unorganisierten Inhalten zu vermeiden.
- Im Allgemeinen gelten für die **Benennung** von Ordnern die gleichen Regeln wie für Dateien.
- Je nach der Struktur Ihrer Daten und Unterlagen müssen Sie entscheiden, wie Sie Ihre Dateien und Ordner anordnen.
- Dies kann auch Abwägungen und Kompromisse erfordern: eine sehr tiefe Ordnerstruktur (Unterordner-Unterordner) kann ebenso unpraktisch sein wie ein Ordner mit 100 oder mehr einzelnen Dateien.
- Falls verschiedene Projektmitglieder unterschiedliche Zugriffsbeschränkungen auf Dateien haben sollten, kann dies durch Ihre Ordnerstruktur abgebildet werden.

Beispiele für Verzeichnisstrukturen

Projekt- Ordner

1. Projektleitung

1. Vorschläge

2. Finanzen

3. Berichte

2. Ethik und Governance

1. Ethische Genehmigungen

2. Zustimmungformulare

3. Experiment 01

1. Eingaben

2. Daten

3. Datenanalyse

4. Ausgänge

4. Verbreitung

1. Präsentationen

2. Veröffentlichungen

3. Öffentlichkeitsarbeit

Nach Suse Prejawa (2021). Ich glaube, ich habe wirklich gute Daten ... aber ich kann sie nirgendwo finden: Die strukturierte Organisation von Forschungsdaten. Vortrag präsentiert auf Human Research Data in Practice, Virtual Workshop. 2021-04-20, <https://hdl.handle.net/21.11116/0000-0008-662A-7>.

Beispiele für Verzeichnisstrukturen

Organisiert nach Dateityp

```
DatensatzA.tar.gz
|- Daten/
|   |- Verarbeitet/
|   |- Roh/
|- Ergebnisse/
|   |- Abbildung1.tif
|   |- Abbildung2.tif
|   |- Modelle/
```

Organisiert nach Analyse

```
DatensatzB.tar.gz
|- Abbildung1/
|   |- Daten/
|   |- Ergebnisse
|       |- Abbildung1.tif
|- Abbildung2/
|   |- Daten/
|   |- Ergebnisse/
|       |- Abbildung2.tif
```

Beispiel README-Vorlage

```
# Titel des Datensatzes

[Zugriff auf diesen Datensatz auf XYZ](DOI-Link des Datensatzes)

Geben Sie eine kurze Zusammenfassung des Inhalts des Datensatzes, kontextualisiert in experimentellen Verfahren und Ergebnissen.

## Beschreibung der Daten und der Dateistruktur

In diesem Abschnitt können Sie frei beschreiben, wie die Daten strukturiert sind und wie ein potenzieller Verbraucher sie nutzen könnte. Seien Sie so anschaulich wie nötig. Denken Sie daran, dass die Nutzer Ihrer Daten möglicherweise neu auf dem Gebiet sind und mit der gängigen Terminologie, den Metriken usw. nicht vertraut sind.

Beschreiben Sie die Beziehungen zwischen den Dateien, fehlende Datencodes und andere verwendete Abkürzungen. Seien Sie so anschaulich wie möglich.

## Freigabe/Zugriffsinformationen

Dies ist ein Abschnitt für Links zu anderen Möglichkeiten des Datenzugriffs und für Links zu Quellen, aus denen die Daten stammen, falls vorhanden.

Links zu anderen öffentlich zugänglichen Stellen, an denen die Daten zu finden sind:

- [http://...] (http://...)

Die Daten stammen aus den folgenden Quellen:

- [] ()

## Code/Software

Dies ist ein optionaler, freier Abschnitt zur Beschreibung des Codes in Ihrem Beitrag und der Software, mit der er ausgeführt wird.

Beschreiben Sie alle Skripte, Codes oder Notebooks (z. B. R, Python, Mathematica, MatLab) sowie die Softwareversionen (einschließlich geladener Pakete), die Sie zur Ausführung dieser Dateien verwendet haben. Wenn Ihr Repository mehr als eine Datei enthält, deren Beziehung zu anderen Skripten nicht offensichtlich ist, geben Sie Informationen über den Arbeitsablauf an, den Sie zur Ausführung dieser Skripte und Notebooks verwendet haben.
```

Kurze eigene Arbeit

Frage: Schauen Sie sich die Dateien und Ordner in Ihrem aktuellen Projekt an. Welche Probleme sehen Sie?

Methode: Einzelarbeit

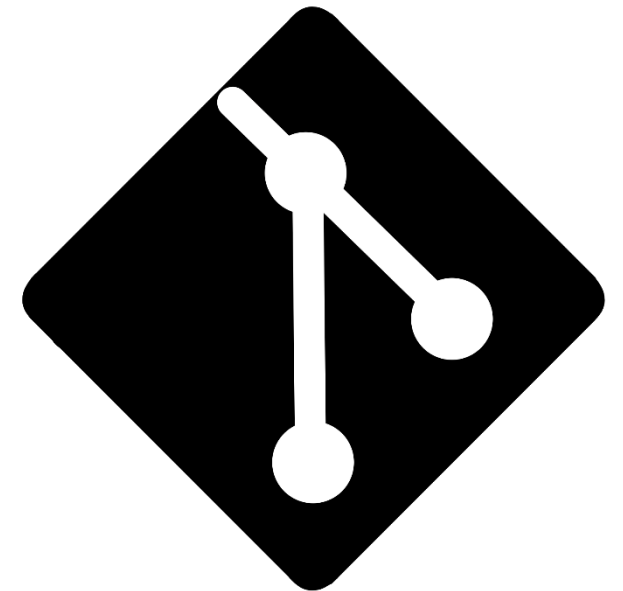
Zeit: 5 Minuten

Wissen sichern: Anschließend Austausch innerhalb der Gruppe

Versionskontrolle

Dokumente können sich im Laufe der Zeit weiterentwickeln, und mehrere Personen können an aufeinanderfolgenden Änderungen beteiligt sein. Sowohl für Daten (jeglicher Art, z. B. Zahlen oder Bilder) als auch für Textdokumente dient die Dateiversionierung zwei Zwecken:

- Sie können bei Bedarf zu früheren Versionen **zurückkehren**.
- Sie können Änderungen **nachverfolgen**, einschließlich der Dokumentation über die zugrunde liegende Logik und die beteiligten Personen.
- Die Versionskontrolle kann entweder **manuell** mit Hilfe von Namenskonventionen oder mit einem **Versionskontrollsystems** erfolgen.



<https://doi.org/10.5281/zenodo.3674561>

Manuelle Versionskontrolle

- Bei der Arbeit mit einem **manuellen Kontrollsystem** sollten die Versionen fortlaufend nummeriert werden; so können größere Änderungen (v1, v2, v3, ...) von kleineren (v1-1, v1-2, v1-3 oder 1a, 1b, 1c) unterschieden werden. Verwenden Sie führende Nullen, wenn Sie mehr als neun Versionen erwarten.
- Verwenden Sie eine solche Nummerierung jedoch nicht, wenn Sie eine Versionskontrollsoftware verwenden, da sie die automatische Versionierung beeinträchtigt.
- Bezeichnungen wie "roh" oder "verarbeitet" für Daten oder "Entwurf" oder "intern" für Dokumente sind nützlich.
- Aber Begriffe wie "final2", "final-revised", "final-changed_again", "final_ready" können verwirrend sein.
Mit anderen Worten: **Vermeiden** Sie die Benennung "final".



<https://doi.org/10.5281/zenodo.3674561>

Versionskontrollsysteme

- Ein bekanntes Beispiel für ein **Versionskontrollsystem** ist Git, ein verteiltes Versionskontrollsystem zur Verfolgung von Änderungen. Es ist im Bereich der Softwareentwicklung weit verbreitet. Auch eine Versionskontrolle für Forschungsdaten ist mit Git möglich.
- Vielleicht gibt es sogar eine **lokale** Git-Instanz an Ihrer eigenen Institution? Fragen Sie danach!
- Eine **Anleitung** zur Verwendung von Git gibt es bei [Data Carpentry](#) oder bei [GitHub Education](#).
- Einen hilfreichen **Überblick** über Werkzeuge zur Versionskontrolle von Daten bietet [The Turing Way](#).



Dateiformate

Proprietäre Datenformate vs. offene Datenformate

- **Proprietäre** Datenformate sind oft ohne die entsprechende (kommerzielle) Software nicht lesbar und können in Zukunft veraltet sein.
- Ein offenes Datenformat ist ein Format mit "einer frei zugänglichen, veröffentlichten Spezifikation, deren Verwendung keinen finanziellen oder sonstigen Beschränkungen unterliegt". (siehe <https://opendefinition.org/ofd/>)
- **Die Konvertierung** von proprietären Formaten in offene Formate ist oft möglich, kann aber zu einem gewissen Datenverlust führen.



Empfohlene Dateiformate I

Art der Daten	Empfohlene Formate
Text	Einfacher Text (.txt) Rich Text Format (.rtf) Offener Dokumententext (.odt; ISO 26300) PDF (PDF/A-1)
Tabellarische Daten (minimale Metadaten)	Komma-getrennte Werte (.csv) Offene Dokumentenkalkulationstabelle (.ods; ISO 26300) Tabulator-getrennt (.tab)
Tabellarische Daten (umfangreiche Metadaten)	SPSS portables Format (.por) NetCDF HDF5

Diese Liste ist ein Leitfaden und nicht bindend.

Empfohlene Dateiformate II

Art der Daten	Empfohlene Formate
Bilder	TIFF6.0 unkomprimiert (.tif) JPEG2000 (.jp2) JPEG (.jpg) GIF (.gif) Skalierbare Vektorgrafiken (.svg) FITS (.fits) PNG (.png)
Video	MPEG-4 (.mp4) OGG-Video (.ogv, .ogg)
Audio	Freier verlustfreier Audiocodec (.flac) Wellenform-Audio-Dateiformat (.wav)

Diese Liste ist ein Leitfaden und nicht verbindlich.

Teil 3

Forschungsdatenmanagement in Bibliotheken

Inhalt von Teil 3

1. FDM an Bibliotheken
2. Herausforderungen beim FDM-Wissen
3. Institutionelle Selbstevaluierung
4. Science Clouds
 1. Nationale Forschungsdateninfrastruktur
 2. European Open Science Cloud
 3. GOSC & GAIA-X

FDM an Bibliotheken, warum?

- Neue Technologien
- Neue Anforderungen
- Neue kommerzielle Dienste
- Wahrnehmung
- Krise der Bibliotheken? Bibliotheken im Wandel?

Kommission Virtuelle Bibliothek (KVB) im BVB: Die Rolle der wissenschaftlichen Bibliotheken Bayerns beim Forschungsdatenmanagement – Herausforderungen, Aufgabenfelder, Handlungsempfehlungen, 2016, https://www.bib-bvb.de/documents/11301/1476009/PositionspapierKVB_August_2016.pdf/369b9b2c-0513-45e9-b3eb-72819613de9d.

LIBER

- Ligue des bibliothèques européennes de recherche (<http://libereurope.eu>)
- 450 Forschungsbibliotheken
- 62 deutsche Bibliotheken
- München: UB TU, BSB, UB UniBwM, MPDL
- <https://libereurope.eu/liber-participants/>



Final report of the LIBER working group on E-Science / Research Data Management, 2012, <https://libereurope.eu/wp-content/uploads/2020/11/The-research-data-group-2012-v7-final.pdf>.

Liber Recommendations for Libraries

1. Offer research data management support, including data management plans for grant applications, intellectual property rights advice and information materials. Assist faculty with data management plans and the integration of data management into the curriculum.

2. Engage in the development of metadata and data standards and provide metadata services for research data.

3. Create Data Librarian posts and develop professional staff skills for data librarianship.

4. Actively participate in institutional research data policy development, including resource plans. Encourage and adopt open data policies where appropriate in the research data life cycle.

5. Liaise and partner with researchers, research groups, data archives and data centers to foster an interoperable infrastructure for data access, discovery and data sharing.

6. Support the lifecycle for research data by providing services for storage, discovery and permanent access.

7. Promote research data citation by applying persistent identifiers to research data.

8. Provide an institutional Data Catalogue or Data Repository, depending on available infrastructure.

9. Get involved in subject specific data management practice.

10. Offer or mediate secure storage for dynamic and static research data in co-operation with institutional IT units and/or seek exploitation of appropriate cloud services:

Forschungsnahе Dienste durch Bibliotheken

Positionspapier der Kommission für forschungsnahе Dienste des VDB:

- Forschungsdatenmanagement (FDM)
- Publikationsdienste: Open Access, Repositorien, Academic Networking
- Publikationsdatenmanagement: Affiliation und Autor*innen-Identität
- Bibliometrien, Evaluierung und Leistungsmessung
- Systematic Reviews: Recherche bei komplexen Forschungsprojekten
- Digitale Integration von (Forschungs-) Infrastrukturen

Stille et al. (2021): Forschungsunterstützung an Bibliotheken: Positionspapier der Kommission für forschungsnahе Dienste des VDB, In: o-bib. Das offene Bibliotheksjournal 8(2), <https://doi.org/10.5282/o-bib/5718>, S. 3-8.

Empfehlungen

- „Support is user experience“
- „Problem first, solution second“
- Zuhören, verstehen, Fragen stellen
- Ergebnisse oder Aufgaben notieren
- Schnelle Lösungen sind nicht immer die Besten
- ggf. auch andere Kolleg_innen und/oder Expert_innen hinzuziehen
- ...

Herausforderungen beim FDM-Wissen

Sieben Herausforderungen beim FDM-Wissen



Herausforderung 1: FDM-Wissen entwickeln

Als Institution

- Eigenen Wissenskanon etablieren
- Schulungsangebote entwickeln
 - Universitäre Curricula
- Infrastruktur und Ressourcen bereitstellen
- Anreize schaffen

Als Person

- Weiterbildungen
- Autodidaktische Aneignung

Slowig et al., *Der Zertifikatskurs Forschungsdatenmanagement in NRW*

Der Zertifikatskurs Forschungsdatenmanagement in NRW
Eine modular aufgebaute Weiterqualifikation für das professionelle Datenmanagement

1. Ausgangslage

Im Zuge der Digitalisierung der Wissenschaft verändern sich die Methoden zur Erkenntnisgewinnung und Informationserschließung. Während ursprünglich ausschließlich analoge bzw. mechanische Verfahren genutzt wurden, nimmt der Anteil von digitalen Forschungsdaten und damit verbundenen Prozessen rasant zu.¹ In diesem Kontext gewinnt das Handlungsfeld des Forschungsdatenmanagements (FDM) an zentraler Bedeutung, um wertvolle Forschungsdaten nachhaltig zu sichern und sie für eine weitere Verwendung im Forschungszyklus zu halten. Hierdurch werden aber auch neue Herausforderungen für wissenschaftliche und infrastrukturelle Einrichtungen geschaffen: Um digitale Forschungsdaten reproduzierbar zu verarbeiten und für eine Nachnutzung aufzubereiten, bedarf es nicht nur entsprechender technischer Informations- und Analyseinfrastrukturen, sondern vor allem Personen, die über fundierte Kompetenzen verfügen, verschiedene Zielgruppen zu einem nachhaltigen Umgang mit Forschungsdaten und zur Nutzung einschlägiger Infrastrukturen zu beraten und zu schulen. Ein systematischer Auf- und Ausbau solcher Kompetenzen wurde bereits 2016 vom Rat für Informationsstrukturen (RifI) mit der Empfehlung formuliert, neue Berufsbilder, Studiengänge und Qualifizierungswege zu schaffen.² Die stetig wachsende Nachfrage nach Personal mit einschlägigen Expertisen zeigt, dass dieses Desiderat nach wie vor besteht.

Dieser Bedarf ergibt sich aus der Bildung standortbezogener Infrastrukturen und überregionaler Strukturen, Netzwerke und Organisationen, die von einzelnen Hochschulen oder Forschungszentren über FDM-Landesinitiativen und der Einrichtung der Nationalen Forschungsdateninfrastruktur (NFDI) bis zur Bildung internationaler Zusammenschlüsse wie der European Open Science Cloud (EOSC) reichen. Die NFDI hat den Auftrag, für die gesamte deutsche Forschungsgemeinschaft Lösungen für die effiziente Speicherung, Analyse und Nachnutzung von Forschungsdaten zu finden. Der Start der NFDI mit inzwischen 19 der 30 fachlich oder methodisch ausgerichteten Konsortien³ befeuert die Nachfrage nach FDM-versiertem Personal maßgeblich und adressiert das Thema einer entsprechenden Qualifizierung im Rahmen der Querschnittsthemen der NFDI (u. a. die Sektion „Training & Education“).⁴ Auf europäischer Ebene findet die NFDI ihr Pendant in der European Open Science Cloud (EOSC), deren Ziel es ist, Forschenden in Europa eine verteilte, offene und multidisziplinäre

1 Kündling, Maxi; Schimbacher, Peter: „Die digitale Forschungswelt“ als Gegenstand der Forschung, in: *Information - Wissenschaft & Praxis* 64 (2-3), 2013, S. 127-136. Online: <<https://doi.org/10.1515/iwp-2013-0017>>.

2 Rat für Informationsinfrastrukturen: *Leistung aus Vielfalt. Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland*, Göttingen 2016, S. 5-160. Online: <www.rif-i.org/10.1515/iwp-2013-0017>.

3 NFDI, *Nationale Forschungsdateninfrastruktur. Konsortien*, <<https://www.nfdi.de/konsortien/>>, Stand 13.06.2022.

4 Herres-Pawlis, Sonja; Pelz, Peter; Kockmann, Norbert u. a.: *Sektionskonzept zur Einrichtung einer Sektion im Verein Nationale Forschungsdateninfrastruktur (NFDI) e.V. (2.0)*, Zenodo, 2022, S. 5-13. Online: <<https://doi.org/10.5281/zenodo.5999769>>.

o|bib 2022/3 <https://doi.org/10.5282/o-bib/5833> 1

Slowig et al.: „Zertifikatskurs Forschungsdatenmanagement in NRW“, o-bib. Das offene Bibliotheksjournal 9,3 (2022), CC BY 4.0.

Herausforderung 2: FDM-Wissen vermitteln

Als Institution

- Onboarding-Prozesse etablieren
- Schulungen anbieten
- Informationsangebote aufbauen

Als Person

- Train-the-Trainer-Fortbildung
- Schulungen anbieten
- Peer-to-Peer-Austausch



Biernacka et al (2021): Train-the-Trainer-Konzept zum Thema Forschungsdatenmanagement, CC BY 4.0, <https://doi.org/10.5281/zenodo.5773203>.

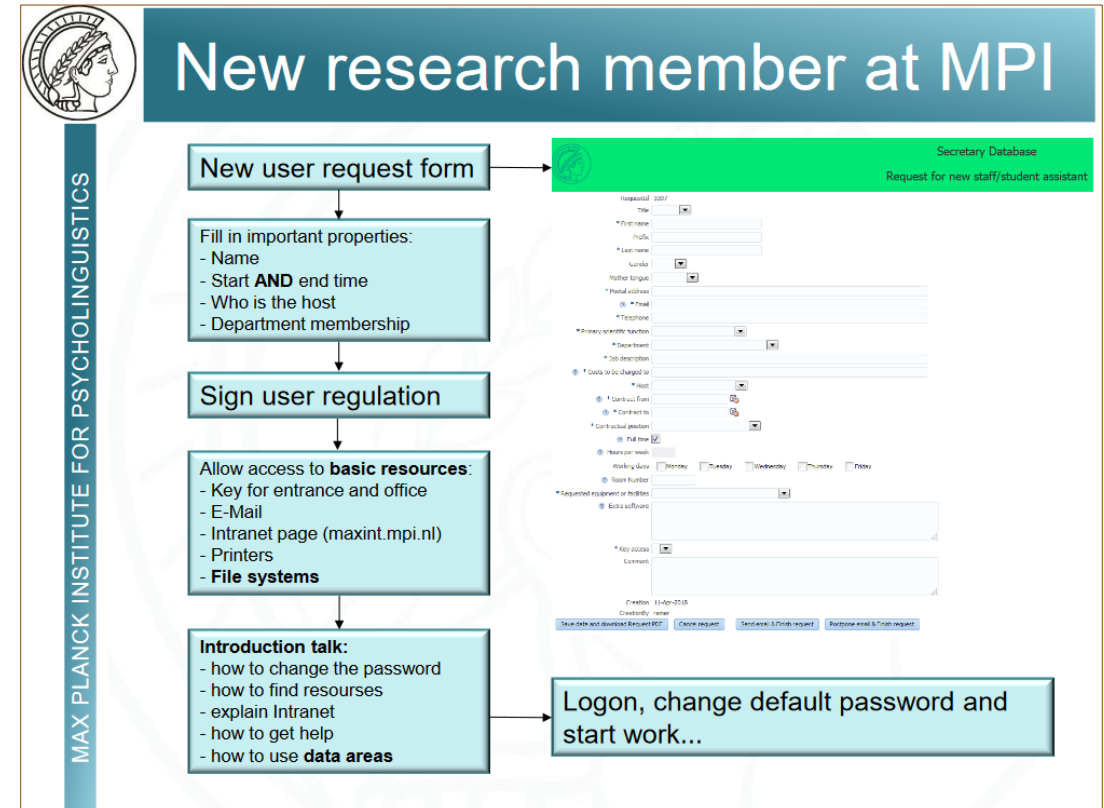
Herausforderung 3: FDM-Wissen etablieren

Als Institution

- *Übergang implizitem zu explizitem Wissen*
- FDM-Strukturen wie Abläufe, Infrastruktur, Ansprechpartner etc. (verbindlich) etablieren

Als Person

- *Übergang implizitem zu explizitem Wissen*
- Schulungen besuchen
- Wissen rekapitulieren
- Austausch mit Kolleg_innen



Reiner Dirksmeyer und Karin Kastens, IT und Bibliothek – Infrastruktur am MPI für Psycholinguistik, FDM-Workshop München 2018, S. 4, https://fdm-workshop.mpg.de/wp-content/uploads/2018/05/Kastens_FDM_MPI_RD_KK.pdf.

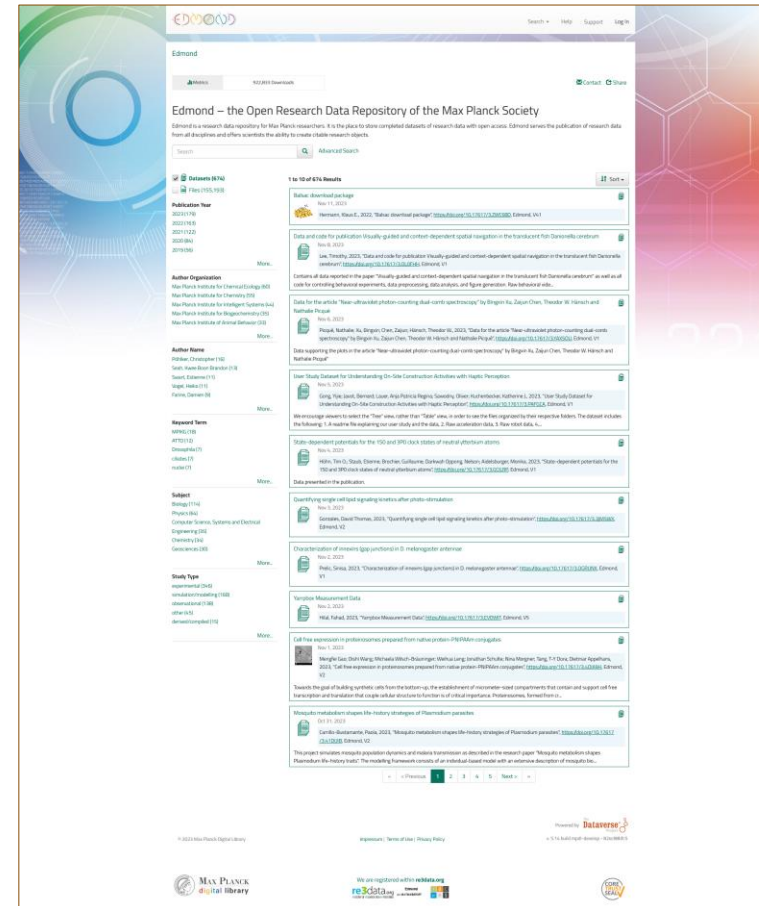
Herausforderung 4: FDM-Wissen anwenden

Als Institution

- FDM-Strukturen anwenden
- ...

Als Person

- Wissen in der Praxis anwenden
- Support-Kanäle nutzen
- ...



<https://edmond.mpg.de>

Herausforderung 5: FDM-Wissen erhalten

Als Institution

- Kontinuierliche Verbesserungsmaßnahmen, wie Angebote unterbreiten
 - Eine einmalige Schulung ist nicht ausreichend
 - (Erneutes) Onboarding
- Evaluation der eigenen Institution
 - Evaluation als Iteration zu einer Policy
- Instanz am Institut mit Beständigkeit (z.B. Bibliothek, Forschungscoordination, IT)

Als Person

- Kontinuierliche Rekapitulation der eigenen Wissensstände?
- (Erneute) Schulung zum Erhalt von Wissen



The image shows the cover of a report. At the top right is a small logo. The main title is "Strategieentwicklung für institutionelles Forschungsdatenmanagement - Erfahrungsbericht zum Einsatz von RISE-DE". Below the title, it says "5. FDM-Workshop 2022 - FDM in der MPG" and "13. - 14.09.2022". The author is "Dr. Steffen Henicke (MPIWG)". At the bottom right is the MPIWG logo and the text "MPIWG MAX PLANCK INSTITUTE FOR THE HISTORY OF SCIENCE".

Henicke, Steffen: „Strategieentwicklung für institutionelles Forschungsdatenmanagement – Erfahrungsbericht zum Einsatz von RISE-DE“, 5. FDM-Workshop 2022: Forschungsdatenmanagement in der Max-Planck-Gesellschaft, <https://hdl.handle.net/21.11116/0000-000B-2B50-B>.

Herausforderung 6: FDM-Wissen weiterentwickeln

Als Institution

- Strukturen, Prozesse, Wissensstände etc. weiterentwickeln
- Aktualität erhalten
- Innerhalb der eigenen Institution über FDM verständigen
- Weiteres Bewusstsein aufbauen
- Anreize schaffen
- Menschen mitnehmen

Als Person

- Anreize nutzen
- Fortbildungen besuchen

Laura Rothfritz, Vivien Petras, Maxi Kindling und Heike Neuroth

3.4 Aus- und Weiterbildung für das Forschungsdatenmanagement in Deutschland

Abstract: Der vorliegende Beitrag gibt einen Überblick über Aus- und Weiterbildungsmöglichkeiten im Bereich des wissenschaftlichen Forschungsdatenmanagements (FDM) in Deutschland. Sowohl internationale als auch nationale Koordinierungsmaßnahmen für das FDM wie die European Open Science Cloud und die Nationale Forschungsdateninfrastruktur erfordern gut qualifiziertes Personal. Jedoch befinden sich sowohl Rollen- als auch Berufsbilder noch in der Entwicklung und es besteht die Herausforderung, Kompetenzbereiche entsprechend zu definieren.

Grundlage für ein gutes FDM ist ein hohes Maß an Datenkompetenz (Data Literacy), die zu einem kritischen und lösungsorientierten Umgang befähigt. Hierzu werden beispielsweise vom Stifterverband Projekte für Qualifizierungsmaßnahmen gefördert. Kompetenzbereiche und entsprechende Profile werden auf internationaler Ebene ausgearbeitet – in Deutschland gibt es bislang keine klaren Bezeichnungen und Kompetenzanforderungen.

Bestehende Qualifizierungsmaßnahmen lassen sich in grundständige und weiterbildende Studiengänge sowie in Fort- und Weiterbildungen wie Workshops oder Lehrgänge unterteilen. Innerhalb von Studiengängen werden bislang vorwiegend im Bereich der Informationswissenschaft Fähigkeiten vermittelt. Insbesondere für fachspezifische Qualifizierungsmaßnahmen bieten sich Fort- und Weiterbildungen an. An dieser Stelle ist die exakte Definition der Zielgruppe von Qualifizierungsansätzen sowie die daraus abgeleiteten benötigten Kenntnisse weiterhin eine große Herausforderung.

Eine gemeinsame Qualifizierungsstrategie fehlt bislang in Deutschland. Zukünftig könnten neben der Ausarbeitung von Rollen- und Kompetenzbereichen sowie Berufsbezeichnungen und Karrierewegen auch gemeinsam abgestimmte Qualifizierungsangebote, beispielsweise durch Qualifizierungsallianzen dazu beitragen, das Forschungsdatenmanagement systematisch und nachhaltig im deutschen Wissenschaftsbetrieb zu verankern.

Einleitung

Durch die Initiative European Open Science Cloud (EOSC) wird eine nachhaltige gemeinsame europäische Infrastruktur mit vernetzten, standardisierten und so offen wie möglich zugänglichen Forschungsdaten in vertrauenswürdigen Repositorien

Open Access. © 2021 Laura Rothfritz, Vivien Petras, Maxi Kindling und Heike Neuroth, publiziert von De Gruyter. Dieses Werk ist lizenziert unter der Creative Commons Attribution 4.0 Lizenz. <https://doi.org/10.1515/9783110657807-015>

Laura Rothfritz, Maxi Kindling und Heike Neuroth, Vivien Petras: „Aus- und Weiterbildung für das Forschungsdatenmanagement in Deutschland“, in: Putnigs, Markus, Heike Neuroth und Janna Neumann (Hrsg.): Praxishandbuch Forschungsdatenmanagement, 2021, S. 255-276, <https://doi.org/10.1515/9783110657807-015>, CC BY 4.0.

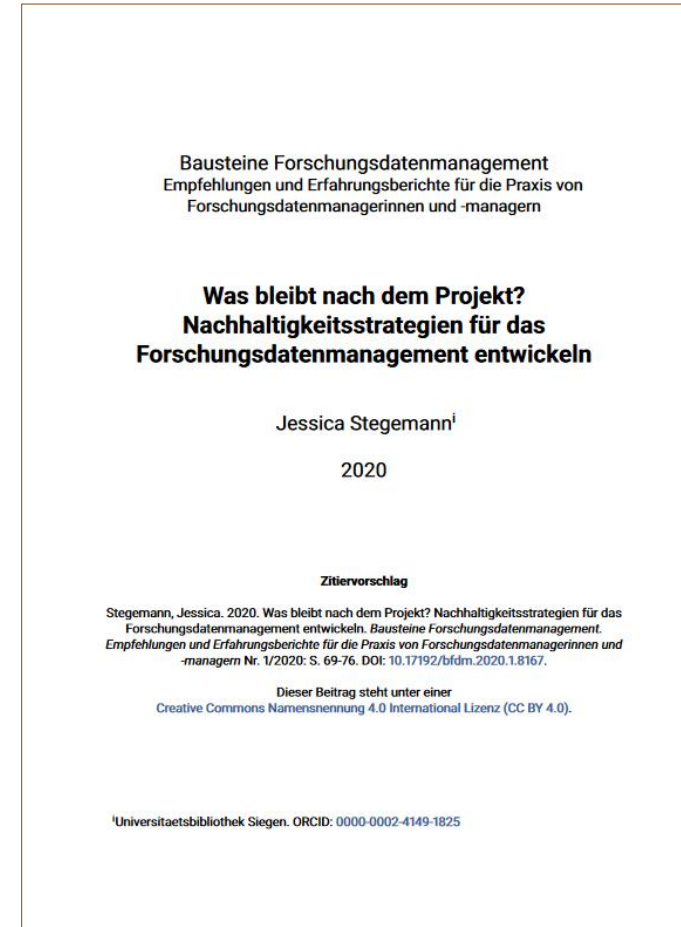
Herausforderung 7: FDM-Wissen weitergeben

Als Institution

- Offboarding nach Ende des Dienstverhältnisses
- Gewährleistung der 10-jährigen Erhaltungsphase nach Projektende
- Nachnutzung von Forschungsdaten

Als Person

- Projekt-/Aufgaben-Übergabe
- Nachnutzung von Forschungsdaten



Stegemann (2020): „Was bleibt nach dem Projekt? Nachhaltigkeitsstrategien für das Forschungsdatenmanagement entwickeln“, CC BY 4.0, <https://bausteine-fdm.de/article/view/8167>.

FDM-Austausch

Frage:

Wie Sehen Sie FDM in Ihrer Bibliothek? Welche Dienste werden angeboten? Welche Chancen und Gefahren sehen Sie?

Methode:

Speed Talking, mit der/dem Nachbar_in

Dauer:

10min (2x5min)

Institutionelle Selbstevaluierung

RISE

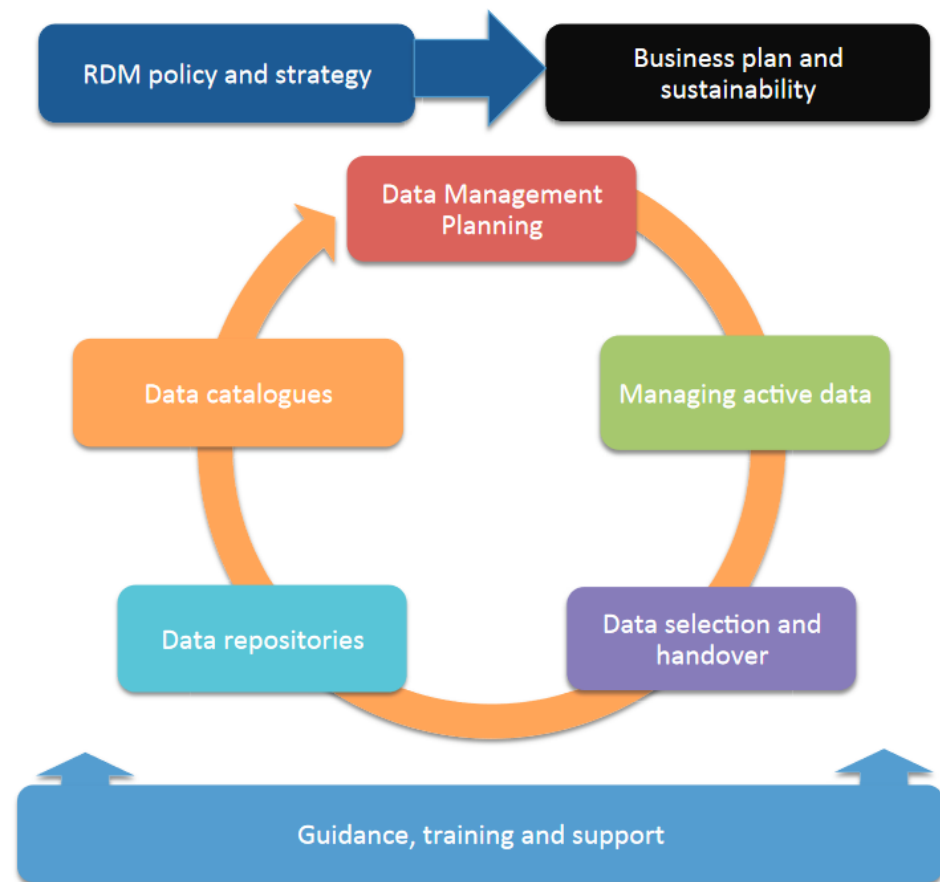
- Research Infrastructure Self-Evaluation Framework (RISE v1.1) des UK Digital Curation Centre (DCC) von 2017
- <https://www.dcc.ac.uk/guidance/how-guides/RISE>



Rans, J and Whyte, A. (2017). 'Using RISE, the Research Infrastructure Self-Evaluation Framework' v.1.1 Edinburgh: Digital Curation Centre. Available online: www.dcc.ac.uk/guidance/how-guides, CC BY 4.0.

RISE

- Komponenten der Unterstützungsdienste für das Forschungsdatenmanagement



Jones, S., Pryor, G. & Whyte, A. (2013). 'How to Develop Research Data Management Services – A guide for HEIs'. DCC How-to Guides. Edinburgh: Digital Curation Centre. Available online: [/guidance/how-guides](#), CC BY 4.0, p. 5.

RISE-DE

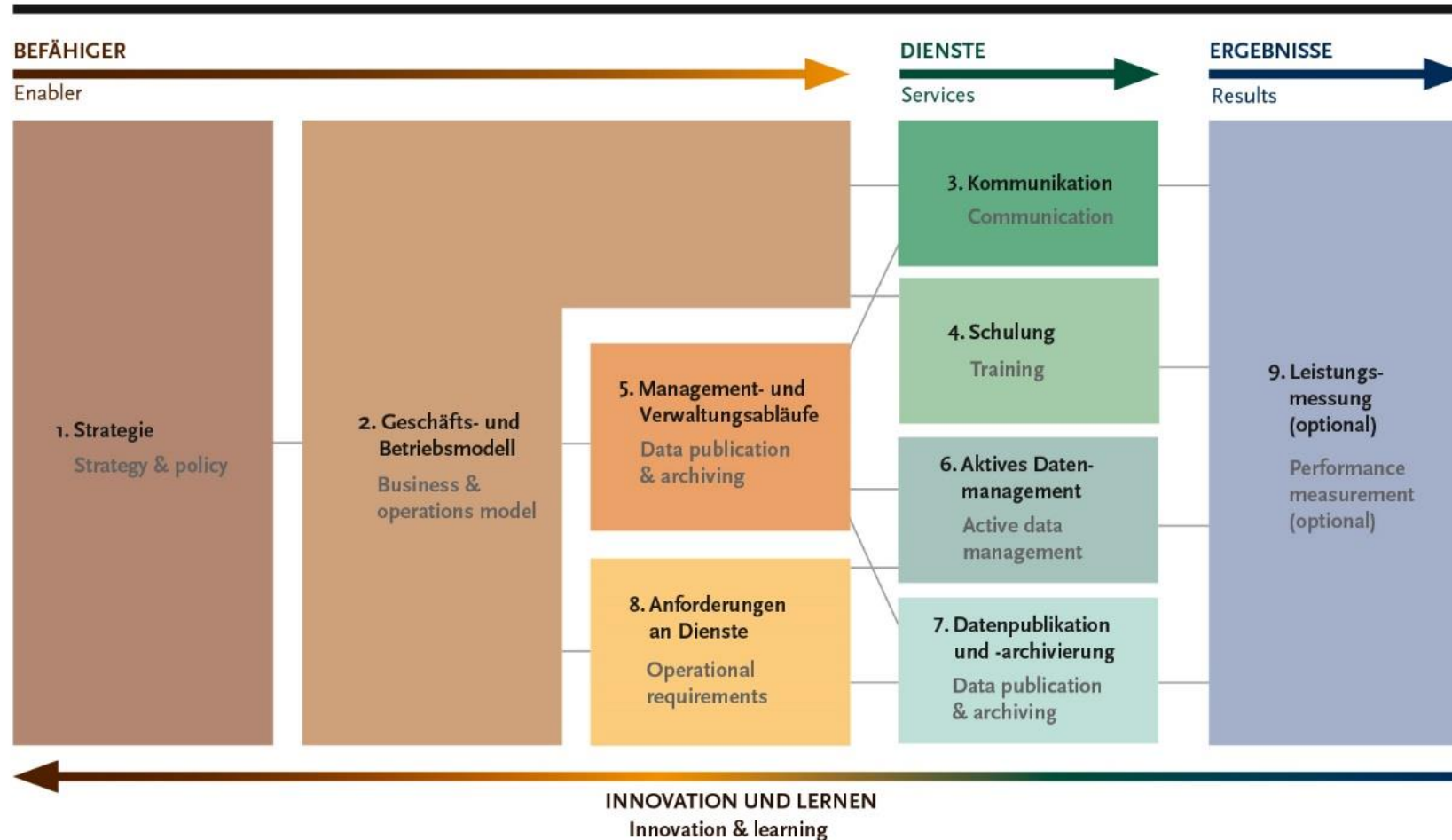
- <https://zenodo.org/record/3585556>
- RISE-DE orientiert sich an dem Research Infrastructure Self-Evaluation Framework (RISE v1.1)
- Anpassung an deutsche Spezifika in der Bildungslandschaft



Hartmann, Jacob und Weiß (2019). RISE-DE – Referenzmodell für Strategieprozesse im institutionellen Forschungsdatenmanagement (1.0), CC BY 4.0, <https://doi.org/10.5281/zenodo.3585556>.

RISE-DE

RISE-DE Referenzmodell



Hartmann, Niklas K., Jacob, Boris, & Weiß, Nadin. (2019). RISE-DE – Referenzmodell für Strategieprozesse im institutionellen Forschungsdatenmanagement (1.0). Zenodo. <https://doi.org/10.5281/zenodo.3585556>, CC BY 4.0, p. 9.

Themenfelder von RISE-DE

1 Strategy

- 1.1 Strategy development
- 1.2 Strategic goals
- 1.3 Commitment and incentives
- 1.4 RD Policy

2 Business and Operating Model

- 2.1 Governance and organisational development
- 2.2 Human resource investments
- 2.3 Technology investments
- 2.4 Cost modelling

3 Communication

- 3.1 Information
- 3.2 Advice
- 3.3 Public Relations

4 Training

- 4.1 Training materials
- 4.2 Training events

Themenfelder von RISE-DE

5 Management processes and administrative procedures

5.1 Integration into project planning

5.2 Integration with research information

5.3 Integration with legal and ethical processes

6 Active research data management

6.1 Basic IT services

6.2 Services for Cooperative Work

7 Data publication and archiving

7.1 Publication services

7.2 Archiving services

7.3 Risk management

8 Requirements for services

8.1 Fail-safety

8.2 Data security

8.3 Metadata and indexing

8.4 Standard conformity of publication and LTA service

Stufen der RISE-DE-Bewertungsskala

Level	Definition	Typical Requirement
0	No (significant) activity	None
1	Minimal activity	External Requirements, Preservation of Third party funding
2	Committed activity	Needs of the researchers of the institution
3	Outstanding activity	National or international industry leader

Hartmann, Niklas K., Jacob, Boris, & Weiß, Nadin. (2019). RISE-DE – Referenzmodell für Strategieprozesse im institutionellen Forschungsdatenmanagement (1.0). Zenodo. <https://doi.org/10.5281/zenodo.3585556>, CC BY 4.0, p. 10.

RISE-DE Evaluation Tool

- <https://zenodo.org/record/3585556/files/RISE-DE%20V1.0%20Evaluations-Tool.xlsx?download=1>
- ein xlsx-Dokument

1.1 Strategieentwicklung	
Das Thema Strategieentwicklung beschreibt die Qualität und den Umfang des Prozesses zur Selbstevaluation und Strategieentwicklung.	
Stufe 0	Aktivitäten ohne strategische Planung Die Aktivitäten der Zentralen Einrichtungen in Bezug auf Forschungsdaten, insbesondere Auf- und Ausbau von Diensten, erfolgen ohne strategische Planung.
Stufe 1	Einfache Strategieentwicklung Die Strategieentwicklung erfolgt durch von der Leitung beauftragte Personen. Als Basis der Selbstevaluation dienen Informationen über die Forschungsdatenpraxis an der eigenen Einrichtung, die aus Verwaltungsdaten und Befragungen gewonnen werden. Die Ziele werden von der Leitung festgelegt und ggf. durch Gremien legitimiert.
Stufe 2	Strukturierte Strategieentwicklung unter Einbindung einzelner Akteure Die Strategieentwicklung erfolgt durch von der Leitung beauftragte Personen, die gezielt Kontakt zu relevanten Akteuren aufnehmen, um diese einzubinden und weitere Informationen zu gewinnen. Die Strategieentwicklung wird durch Verwendung strukturierter Instrumente zur Selbstbewertung und Zielbestimmung unterstützt. Die Selbstbewertung wird von den relevanten Gremien zur Kenntnis genommen und die Zielbestimmung durch diese legitimiert.
Stufe 3	Partizipative strukturierte Strategieentwicklung Zur Strategieentwicklung werden Selbstevaluation und Zielbestimmung im Rahmen eines strukturierten, Teilorganisations- und Hierarchieebenen-übergreifenden Stakeholder-orientierten, durch Leitung und Gremien legitimierten Prozesses durchgeführt.

Hartmann, Jacob und Weiß (2019). RISE-DE – Referenzmodell für Strategieprozesse im institutionellen Forschungsdatenmanagement (1.0), CC BY 4.0, <https://doi.org/10.5281/zenodo.3585556>.

Empfehlungen für FDM- und Strategieprozess-Einsteiger

Bei der Bewertung des institutionellen FDM einer Einrichtung und der Entwicklung einer Strategie zur Einrichtung oder Verbesserung von Diensten ist es sinnvoll, in mehreren Schritten vorzugehen:

1. Grundsatzentscheidungen, Vorüberlegungen und Initiierung
2. Evaluierung mit Beteiligung der Betroffenen
3. Entwicklung und Vereinbarung eines Textentwurfs
4. Verabschiedung und Veröffentlichung

Weitere Evaluierungsmodell: DIAMANT

Lemaire, Marina / Gerhards, Lea / Kellendonk, Stefan / Blask, Katarina / Förster, André:
Das DIAMANT-Modell 2.0. Modellierung des FDM-Referenzprozesses und Empfehlungen für die Implementierung einer institutionellen FDM-Servicelandschaft.



(Universität Trier eSciences Working Papers, Nr. 05), Trier 2020,

<https://doi.org/10.25353/ubtr-xxxx-f5d2-fffb>,
CC BY 4.0.



Weitere Evaluierungsmodell: UpdateFDM

- Reifegradmodell zur Evaluation des institutionellen Forschungsdatenmanagements
- Verknüpfung von:
 1. FDM als Bestandteil der datenbasierten Forschungslandschaft
 2. und Nutzenden- und Serviceorientierung des IT-Betriebs.

Icon	Fragengruppe	Definition der Fragengruppe	Fragen	Ausfällhilfe und Beispiele	
	Service Design	Das Service Design betrifft die ausführliche Beschreibung der Funktionsweise des Services. Neben dieser allgemeinen Definition des Services werden auch Rollen und Verantwortlichkeiten erfasst.	Governance	[1] Ist die Funktionsweise des Services definiert?	Die schriftliche Beschreibung der Funktionsweise umfasst den Zweck sowie den Funktionsumfang eines Services.
				[2] Gibt es einen oder mehrere technische und strategische betreuete werden kann bzw. soll?	Verantwortliche können sowohl mit der technischen Betreuung als auch mit der Administration eines Services beauftragt sein.
			Dokumentation	[3] Ist definiert, wie der Service intern dokumentiert?	Verantwortliche können einsehen, inwiefern die Servicequalität für Nutzende sichergestellt werden. Bsp.: Service Level Agreement.
				[4] Erfolgt die Nutzung des Services im Rahmen gesetzlicher und institutioneller Verpflichtungen?	Es wird sichergestellt, dass Maßnahmen zur Einhaltung geltender Gesetze und institutioneller Richtlinien bei der Nutzung des Services umgesetzt werden. Bsp.: DSGVO, Urheberrecht, Sicherheitskonzepte, FAIR-Prinzipien, etc.
	Service Request Management	Das Service Request Management betrifft alle Anfragen zur Nutzung des Services. Changes (Veränderungen) und Problemmeldungen sind ausgenommen. Auch die Definition von Rollen und Verantwortlichkeiten gehören zu dieser Fragengruppe.	Governance	[1] Ist definiert, wer Anfragen an den Service stellen kann bzw. darf?	Die Definition umfasst ggf. ausgeschlossene Personengruppen.
				[2] Ist definiert, in welcher Form Anfragen an den Service gestellt werden können?	Es gibt eine oder mehrere Kontaktmöglichkeiten, die definiert sind.
			Dokumentation	[3] Ist definiert, welche Informationen in den Anfragen zu einem Service mindestens	Es ist definiert, welche Informationen mindestens in Anfragen enthalten sein müssen, sodass diese strukturiert bearbeitet werden können.
				[4] Erfolgt das Stellen von Anfragen im Rahmen gesetzlicher und institutioneller Verpflichtungen?	Es wird sichergestellt, dass Maßnahmen zur Einhaltung geltender Gesetze und institutioneller Richtlinien beim Stellen von Anfragen umgesetzt werden. Bsp.: DSGVO, Urheberrecht, Sicherheitskonzepte, FAIR-Prinzipien, etc.
			Dokumentation	[1] Ist intern dokumentiert, wer Anfragen stellen kann, wie die Einrichtung von Anfragen erfolgt und welche Mindestangaben	Mitarbeitende der Einrichtung haben Zugriff auf das Service Request Management. Sind nur Teilbereiche dokumentiert, bedarf es einer Einzelfallentscheidung.
				[2] Ist öffentlich dokumentiert, wer Anfragen stellen kann, wie die Einrichtung von Anfragen erfolgt und welche Mindestangaben	Alle Interessierten, insbesondere die Nutzenden haben Zugriff auf das Service Request Management. Sind nur Teilbereiche dokumentiert, bedarf es einer Einzelfallentscheidung.
				[3] Werden die Anfragen für ein internes Wissensmanagement	Anfragen und deren Bearbeitung werden strukturiert für Mitarbeitende zur Verfügung gestellt.
				[4] Wird die Dokumentation regelmäßig überarbeitet und	Die Informationen zum Service Request Management werden in regelmäßigen Abständen evaluiert und bei Bedarf überarbeitet.

Lehmann, A., Dreyer, M., Odebrecht, C., & Helbig, K. (2023). UpdateFDM (Version 1). Zenodo. <https://doi.org/10.5281/zenodo.8210530>, CC BY 4.0.

Literaturempfehlung

- Petersen et al. (2022): Lernzielmatrix zum Themenbereich Forschungsdatenmanagement (FDM) für die Zielgruppen Studierende, PhDs und Data Stewards, <https://doi.org/10.5281/zenodo.7034478>.
- DINI AG Forschungsinformationssysteme (2022): Management von Forschungsinformationen in Hochschulen und Forschungseinrichtungen, <https://doi.org/10.18452/25440>.
- Helbig, Kerstin, Katarzyna Biernacka, Petra Buchholz, Dominika Dolzycka, Niklas Hartmann, Thomas Hartmann, Beate Maria Hiemenz, u. a. (2019): Lösungen und Leitfäden für das institutionelle Forschungsdatenmanagement, in: o-bib. Das offene Bibliotheksjournal 6 (3): S. 21–39, <https://doi.org/10.5282/o-bib/2019H3S21-39>.

FDM-Tabu

Teilnehmende:

2 Mannschaften

Spielregeln:

Eine Person bekommt einen FDM-Begriff und muss ihn erklären, ohne den Begriff oder Bestandteile zu verwenden. Die beiden Mannschaften müssen den Begriff erraten.

Ziel:

Die Mannschaft mit den meisten erratenen Begriffen gewinnt.

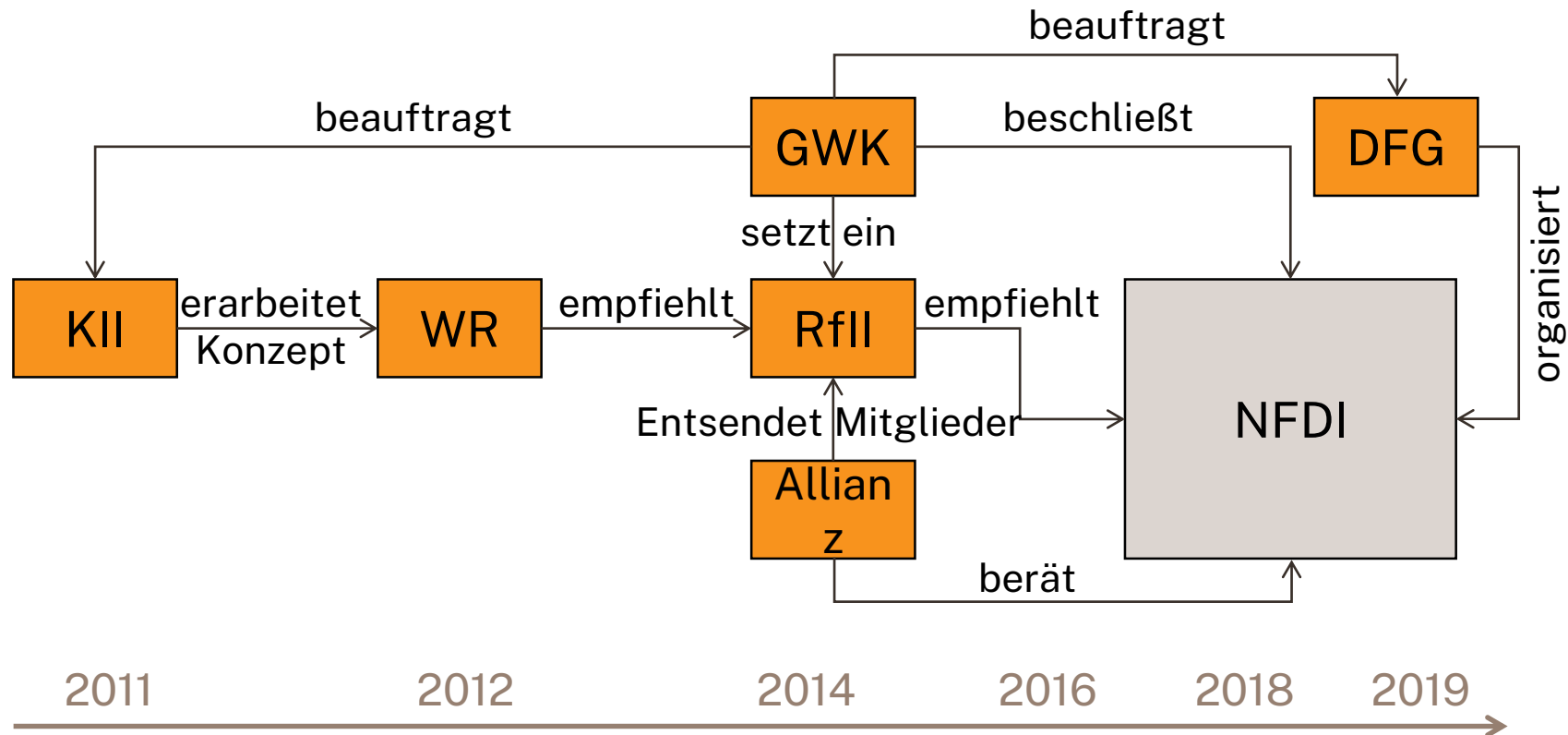
Nationale Forschungsdateninfrastruktur

Nationale Forschungsdateninfrastruktur

„Die nationale Forschungsdateninfrastruktur (NFDI) soll die Datenbestände von Wissenschaft und Forschung systematisch erschließen, nachhaltig sichern und zugänglich machen sowie (inter-)national vernetzen. Sie wird in einem aus der Wissenschaft getriebenen Prozess als vernetzte Struktur eigeninitiativ agierender Konsortien aufgebaut werden.“ (DFG-Website)

- NFDI: <https://www.nfdi.de>
- DFG: <https://www.dfg.de/foerderung/programme/nfdi/index.html>

Nationale Forschungsdateninfrastruktur



DFG: Deutsche Forschungsgemeinschaft

GWK: Gemeinsame Wissenschaftskonferenz

KII: Kommission Zukunft der Informationsinfrastruktur

RfII: Rat für Informationsinfrastruktur

WR: Wissenschaftsrat

siehe auch <https://www.nfdi.de/verein/#historie>

Werbevideo zur NFDI



M-Create, Düsseldorf für DFG, CC BY 4.0, <https://youtu.be/x3Cvn1vNQ98>

1. Runde der NFDI-Konsortien 2019/20

Akronym	Name	weiterführende Informationen
DataPlant	Data in PLANT research	http://nfdi4plants.de/
GHGA	German Human Genome-Phenome Archive	https://ghga.dkfz.de/
KonsortSWD	Konsortium für die Sozial-, Bildungs-, Verhaltens-und Wirtschaftswissenschaften	https://www.konsortswd.de/
NFDI4BioDiversity	NFDI4BioDiversity: Biodiversity, Ecology & Environmental Data	https://www.nfdi4biodiversity.org/
NFDI4Cat	NFDI for Catalysis-Related Sciences	http://gecats.org/NFDI4Cat.html
NFDI4Chem	Fachkonsortium Chemie für die Nationale Forschungsdateninfrastruktur	https://www.nfdi4chem.de/
NFDI4Culture	Consortium for research data on material and immaterial cultural heritage	https://nfdi4culture.de/
NFDI4Health	National Research Data Infrastructure for Personal Health Data	https://www.nfdi4health.de/
NFDI4Ing	Nationale Forschungsdateninfrastruktur für die Ingenieurwissenschaften	https://nfdi4ing.de/

2. Runde der NFDI-Konsortien 2020/21

Akronym	Name	weiterführende Informationen
BERD@NFDI	Business, Economic and Related Data @ NFDI	https://www.berd-nfdi.de/
DAPHNE4NFDI	Data from Photon and Neutron Instruments for NFDI	https://www.sni-portal.de/de/daphne-nfdi
FAIRmat	FAIR Data Infrastructure for Condensed-Matter Physics and the Chemical Physics of Solids	https://www.fair-di.eu/fairmat/fairmat_/consortium
MaRDI	Mathematical Research Data Initiative	https://www.mardi4nfdi.de/
NFDI4DataScience	National Research Data Infrastructure for Data Science	https://www.nfdi4datascience.de
NFDI4Earth	NFDI Consortium Earth System Science	http://www.nfdi4earth.de
NFDI4Microbiota	National Research Data Infrastructure for Microbiota Research	http://nfdi4microbiota.de
NFDI-MatWerk	National Research Data Infrastructure for Materials Science & Engineering	https://nfdi-matwerk.de/
PUNCH4NFDI	Particles, Universe, NuClei and Hadrons for the NFDI	https://www.punch4nfdi.de/
Text+	Text+: Language- and Text-Based Research Data Infrastructure	https://www.text-plus.org

3. Runde der NFDI-Konsortien 2021/22

Akronym	Name	weiterführende Informationen
BASE4NFDI	Basisdienste für die NFDI	https://base4nfdi.de
FAIRAgro	FAIR für Agrosystemforschung	https://fairagro.net
NFDI4BioImage	National Research Data Infrastructure for Bio-Imaging	https://nfdi4bioimage.de
NFDI4Energy	National Research Data Infrastructure for the Interdisciplinary Energy System Research	https://nfdi4energy.uol.de
NFDI4Immuno	National Research Data Infrastructure for Immunology	https://www.nfdi4immuno.de
NFDI4Memory	NFDI4Memory – The Consortium for the Historically Oriented Humanities	https://4memory.de
NFDI4Objects	Research Data Infrastructure for the Material Remains of Human History	https://www.nfdi4objects.net
NFDIxCS	National Research Data Infrastructure for and with Computer Science	https://nfdixcs.org

NFDI-Statistiken der DFG

DFG: Statistische Übersicht zu den Förderentscheidungen in beiden Ausschreibungsrunden, June 2021,

https://www.dfg.de/download/pdf/foerderung/programme/nfdi/nfdi_auswertung_2021.pdf.

DFG: NFDI – Statistische Übersichten zum Antragseingang, November 2019,

https://www.dfg.de/download/pdf/foerderung/programme/nfdi/nfdi_auswertung_2019.pdf.

DFG: Statistische Übersichten zum Antragseingang – Zweite Ausschreibungsrunde, September 2020,

https://www.dfg.de/download/pdf/foerderung/programme/nfdi/nfdi_auswertung_2020.pdf.

Beispiele für Dienste in NFDI-Konsortien

KonsortSWD: Data access
via data centres
(<https://www.konsortswd.de/en/konsortswd/tasks/data-access/>)

NFDI4Ing: Vocabularies
and Ontology Service
(<https://terminology.nfdi4ing.de/ts4ing/ontologies>)

NFDI4Culture: Software
Consulting and Tools
Software Consulting
Agency
(<https://nfdi4culture.de/what-we-do/task-areas/task-area-3.html>)

Kurzrecherche

Frage:

Welches NFDI-Konsortium könnte für Ihren Bereich von Relevanz sein?
Wieso?

Methode:

Individuelle Online-Recherche über

https://www.dfg.de/foerderung/programme/nfdi/gefoerderte_konsortien/index.html

Dauer:

5 Minuten

European Open Science Cloud

European Open Science Cloud

*For the European Commission the EOSC “is a **trusted digital platform for the scientific community**, providing seamless access to data and interoperable services that address the whole research data cycle, from discovery and mining to storage, management, analysis and re-use across borders and scientific disciplines”.*

European Commission, 2019, European open science cloud: A new paradigm for innovation and technology. Publications Office, S. 2, <https://data.europa.eu/doi/10.2759/016783>.

European Open Science Cloud

*The central goal of the EOSC is nothing less than **global EU leadership in research data management**. This goes hand in hand to ensure that European researchers have access to all the benefits of data-driven research.*

Achim Streit und Jos van Wezel (2021): Deutschland in der European Open Science Cloud, in: Putnings et al., Praxishandbuch Forschungsdatenmanagement, S. 32, <https://doi.org/10.1515/9783110657807-003>.

*EOSC is not a cloud “made in Brussels”, it should help Europe to build a “**Schengen Area for digital Data**”.*

Budroni, P., Claude-Burgelman, J., & Schouppe, M. (2019): Architectures of Knowledge: The European Open Science Cloud, in: ABI Technik 39(2), S. 140, <https://doi.org/10.1515/abitech-2019-2006>.

European Open Science Cloud

*From a more technical perspective EOSC is an **integration layer**. It aims at establishing interoperability among the existing and forthcoming European Research Infrastructures. So, by federating existing research data infrastructures, the EOSC leverages national investments and adds value in terms of scale, interdisciplinary and faster innovation.*

European Commission, <https://digital-strategy.ec.europa.eu/en/policies/open-science-cloud>.

European Open Science Cloud

Sie ist:

- nicht rein europäisch, da auch außereuropäische Partner teilnehmen
- nicht nur offen, da es auch Dienste mit Zugangsbeschränkungen gibt
- nicht nur wissenschaftlich sondern beinhaltet auch Datenservice für z.B. die Industrie
- keine Cloud, sondern eher ein föderiertes Infrastruktursystem

EOSC Catalogue and Marketplace

- <https://marketplace.eosc-portal.eu>
- Integrierte Plattform zur Übersicht
- einfacher Zugang zu zahlreichen Diensten für verschiedene Forschungsbereiche sowie zu Daten
- Suche nach wissenschaftlichem Bereich, aktueller Forschungstätigkeit oder Anbieter

GOSC & GAIA-X

Global Open Science Cloud

“The Global Open Science Cloud (GOSC) initiative will encourage cooperation, alignment, and ultimately interoperability, between existing and emerging Open Science Clouds”

CODATA, S. 3, <https://codata.org/wp-content/uploads/2021/06/GOSC-Introduction-Event-Objectives-and-Agenda.pdf>.

GAIA-X

- [Deutsch-Französische Initiative 2020 für eine Data Cloud, mit besonderem Schwerpunkt auf die Wirtschaft](#)
- Erklärtes Ziel von GAIA-X ist es, die digitale Souveränität von Wirtschaft, Wissenschaft, Regierung und Gesellschaft zu stärken, indem die Entwicklung von Innovationsökosystemen gefördert wird.
- Außerdem geht es bei GAIA-X auch um große Politik und Marketing: Es geht um europäische Werte, um Datenschutz und vor allem um Unabhängigkeit.
- Standards, weniger Infrastruktur

And More

- [China Science and Technology Cloud \(CSTCloud\)](#)
- [African Open Science Platform](#)
- [LA Referencia](#) for Latin America
- ...

Literaturempfehlung

- Zu NFDI: Kammerer, Dietmar, Schrade Torsten, und Stellmacher Martha (2021): NFDI4Culture: Konsortium für Forschungsdaten zu materiellen und immateriellen Kulturgütern, in: Bausteine Forschungsdatenmanagement, Nr. 2 (Juli), S. 23-33, <https://bausteine-fdm.de/article/view/8341>.
- Zu EOSC: Achim Streit, und Jos van Wezel (2021) Deutschland in der European Open Science Cloud, in: Praxishandbuch Forschungsdatenmanagement, herausgegeben von Markus Putnings, Heike Neuroth, und Janna Neumann, S. 31-52, <https://doi.org/10.1515/9783110657807-003>.
- Zu GAIA-X: Mahn, Jan (2020): Gaia-X: Wie Europa in der Cloud unabhängig werden soll, c't Magazin, 25. Juni 2020. <https://www.heise.de/ct/artikel/Gaia-X-Wie-Europa-in-der-Cloud-unabhaengig-werden-soll-4783596.html>.

Research Data ScaryTales

Frage:

Was zeigt das Beispiel?

Methode:

Quiz mit Geschichten von FDM Thüringen

Dauer:

9min (3x3min)

Lang et al. (2023): Research Data ScaryTales, v4.1, Zenodo,
<https://doi.org/10.5281/zenodo.10281701>, CC0.

Nur schauen, nicht anfassen

Hätte sie den Spruch "Nur schauen, nicht anfassen!" ernster genommen, wäre ihr die zusätzliche Arbeit erspart geblieben.

Eine Forscherin hatte die Daten eines Experimentes auf ihrem Computer gespeichert. Eines Tages öffnete sie die originalen Rohdaten in Microsoft Excel, wobei in einigen Spalten die Formatierung automatisch geändert wurde. Dadurch wurden die ursprünglichen Werte unwiederbringlich umformatiert und waren nicht länger auswertbar. Da keine Kopie der Rohdaten existierte, musste das Experiment wiederholt werden.

Das Beispiel zeigt, dass bei der langfristigen Sicherung von Rohdaten besondere Aufmerksamkeit erforderlich ist. Diese sollten mit entsprechenden Backups gesichert werden, wobei der Zugriff generell auf eine Leseberechtigung beschränkt werden sollte. Dadurch erfolgen Auswertungen in separaten Dateien und die ursprünglichen Rohdaten können nicht verändert werden.

<https://www.nature.com/articles/d41586-019-01040-w>

Lang et al. (2023): Research Data ScaryTales, v4.1, Zenodo,
<https://doi.org/10.5281/zenodo.10281701>, CC0, scarytales_de.pdf, S. 16.

Unbeschriebenes Blatt

Die Daten existierten, konnten aber trotz größter Bemühungen nicht nachgenutzt werden.

Zu Beginn seiner Promotion wurde einem jungen Wissenschaftler mitgeteilt, er solle an unveröffentlichten Daten arbeiten, die drei Jahre zuvor erhoben wurden. Er erhielt mehrere Ordner voller Daten. Darin enthalten waren Dateien mit identischem Namen, aber unterschiedlichem Inhalt, Skripte von denen niemand mehr wusste, was sie tun oder warum sie existieren und Tabellen mit unklaren Spaltenbezeichnungen. Noch dazu war teilweise unbekannt, welche Geräte und/oder Einstellungen genau für die Datenerhebung verwendet wurden. Da die Daten mehrere Jahre alt waren, konnten weder intensive Gespräche mit den Herstellern der identifizierten Geräte noch mit den damaligen Forschern die Nachnutzbarkeit der Daten ermöglichen. Am Ende konnten die Daten einfach nicht mehr verwendet werden.

Dies zeigt, wie essenziell das Beschreiben und Dokumentieren von Datensammlungen und Analyseprozessen ist. Auch wenn Datendokumentation Zeit braucht, ist es noch zeitintensiver schlecht dokumentierte, jahrealte Daten aufzubereiten. Obwohl viele Forschende denken, dass sie ihre Daten kennen, ist es sehr wahrscheinlich, dass die meisten von ihnen einen Großteil der Details innerhalb weniger Jahre vergessen. Daher sollte die Datendokumentation immer so umfangreich, detailliert, präzise und für Dritte leicht verständlich sein wie möglich.

<https://www.lcrdm.nl/horror-lack-of-documentation>

Lang et al. (2023): Research Data ScaryTales, v4.1, Zenodo,
<https://doi.org/10.5281/zenodo.10281701>, CC0, scarytales_de.pdf, S. 31.

Unscharfer Mondmann

Durch diesen Vorfall entstand viel Freiraum für Interpretationen.

2006 sollte das Goddard Center's Data Evaluation Laboratory geschlossen werden, welches das letzte Institut war, das mit seiner Ausrüstung die Original-Mondlandungsaufnahmen im SSTV-Format noch lesen konnte. Die Schließung veranlasste die Suche nach den Originalbändern in den Archiven der NASA. Wie sich jedoch 2009 herausstellte, waren diese schon Anfang der 80er Jahre entweder verschwunden oder gar überspielt worden, wie es damals gängige Praxis war. Nur wenige andere Fotos oder Videos konnten noch ausfindig gemacht werden. Die heutzutage bekannten Aufnahmen von der Mondlandung stammen meist von der Umwandlung in das eigentlich bessere NTSC-Format für Fernsehgeräte. In der Umwandlung von SSTV hatte man in den meisten Fällen aber einfach nur einen Bildschirm abgefilmt, wodurch es starke Verluste in Kontrast, Helligkeit und Auflösung gab. Die bis heute nicht aufgefundenen Originalbänder sorgten für eine Welle von Verschwörungstheorien.

Diese Geschichte zeigt, wie wichtig es ist bedeutungsvolle Forschungsdaten richtig zu archivieren. Merkmale einer Archivierung sind die vollständige Übertragung, die Unveränderbarkeit, die Zugriffsbeschränkung und die Dokumentation des Zugangs und Umzugs der Daten. Gegebenenfalls müssen die Daten auf neue Medien transferiert werden, falls die alten obsolet geworden sind.

<https://www.zeit.de/online/2009/30/mondlandung-aufnahmen-verschwunden>

Lang et al. (2023): Research Data ScaryTales, v4.1, Zenodo, <https://doi.org/10.5281/zenodo.10281701>, CC0, scarytales_de.pdf, S. 56.

Teil 4

Weiterführende Überlegungen

Inhalt von Teil 4

1. Open (Research) Data
2. Open Science
3. Wissenschaftliche Software
4. Fortbildungsmöglichkeiten

Open Research Data

Open Research Data

Vier Schritte zu offenen Daten von der Europäischen Kommission:

1. Bereiten Sie Ihre Daten für die gemeinsame Nutzung vor
2. Wählen Sie ein Repository
3. Fügen Sie eine Datenverfügbarkeitserklärung zu Ihrem Artikel hinzu
4. Verknüpfen Sie Ihre Datensätze mit Ihrem Artikel

https://think.f1000research.com/open-research-europe-submission/opendataguide/?utm_source=CPB&utm_medium=cms&utm_campaign=JQC19229

5 ★ Model for Open Data by Tim Berners-Lee

★ Daten online

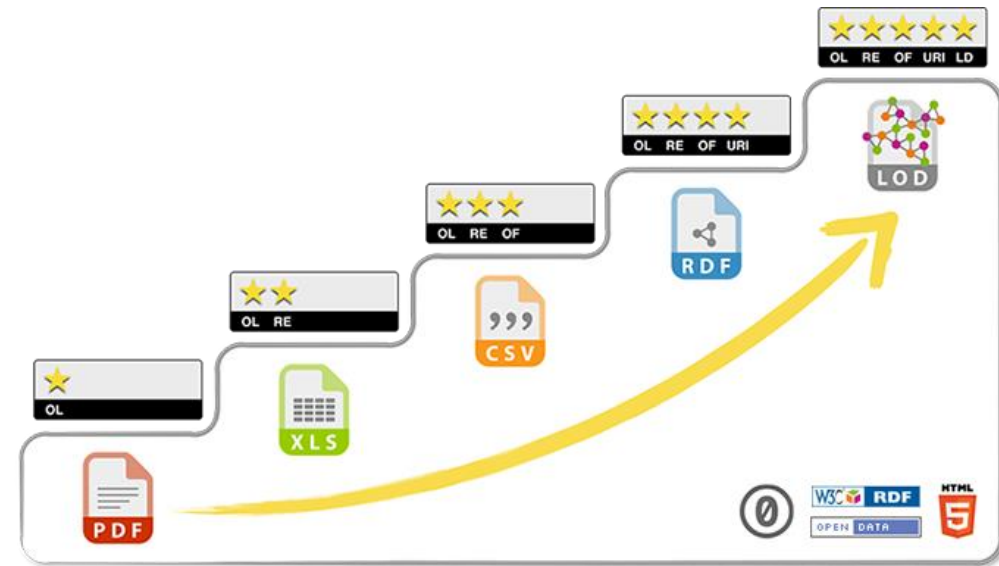
★★ Online verfügbare Daten als strukturierte Daten

★★★ Online verfügbare Daten in einem nicht-proprietären offenen Format

★★★★ Verwenden Sie Identifikatoren, um Dinge zu kennzeichnen, damit andere auf Ihre Daten verweisen können

★★★★★ Verknüpfen Sie Ihre Daten mit anderen Daten, um einen Kontext zu schaffen

<https://5stardata.info>

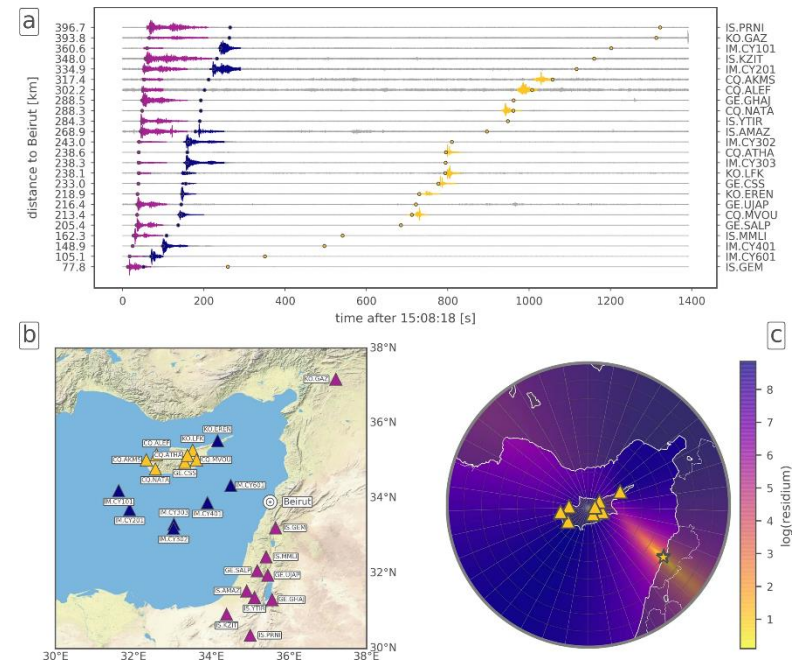


<https://5stardata.info/en/>, CC0.

Beispiel für Open Data

Explosion im Hafen von Beirut

- Kombination von verschiedenen offenen Datenquellen zur Rekonstruktion der Explosionen
- Pilger, C., Gaebler, P., Hupe, P. et al. (2021): Yield estimation of the 2020 Beirut explosion using open access waveform and remote sensing data, Scientific Reports 11, 14144, <https://doi.org/10.1038/s41598-021-93690-y>.



Pilger, C., Gaebler, P., Hupe, P. et al., 2021, CC BY 4.0, <https://www.nature.com/articles/s41598-021-93690-y/figures/1>

Spezifikationen für Open Data

- Open Data ↔ Open Research Data
- Open Data ↔ FAIR Data

- Open Data Portal München
- [Kompetenzzentrum für Open Data des Bundes](#)
- [Open Data Portal der EU](#)
- ...

Open Science

Aspekte von Open Science

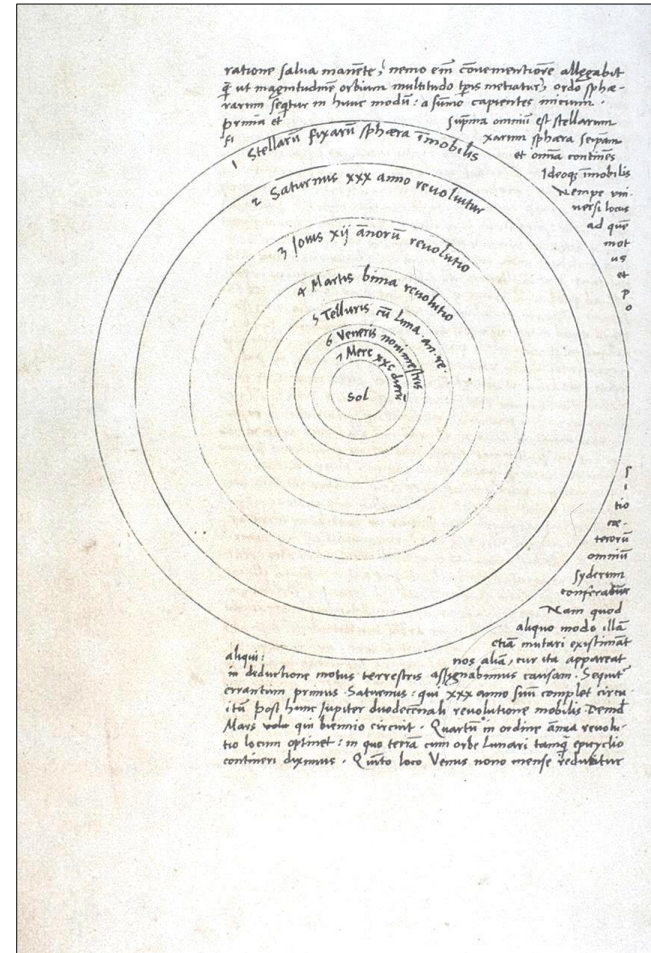
- **Open Access:** Öffentlicher Zugang zu Forschungsergebnissen in Form von Veröffentlichungen
- **Open Research Data:** Freie Verfügbarkeit von Forschungsdaten/Rohdaten
- **Linked Open Data:** Zugänglichkeit und Vernetzung von öffentlichen Repositorien
- **Open Peer Review:** Suche nach Alternativen zu den traditionellen Begutachtungsverfahren der Zeitschriftenverlage, um eine größere Transparenz in diesem Bereich zu gewährleisten

Aspekte von Open Science

- **Open Source:** Entwicklung und Nutzung von Open-Source-Software für die Wissenschaft
- **Citizen Science:** Einbindung von Nicht-Wissenschaftlern in den wissenschaftlichen Prozess
- **Open Educational Resources:** Kostenloser Zugang zu digitalem Lehrmaterial und (Aufzeichnungen von) Kursen
- **Pre-registration:** Hinterlegung und Begutachtung eines Projekt- und Methodenplans mit Hypothesen, die auf einer geeigneten Plattform getestet werden sollen

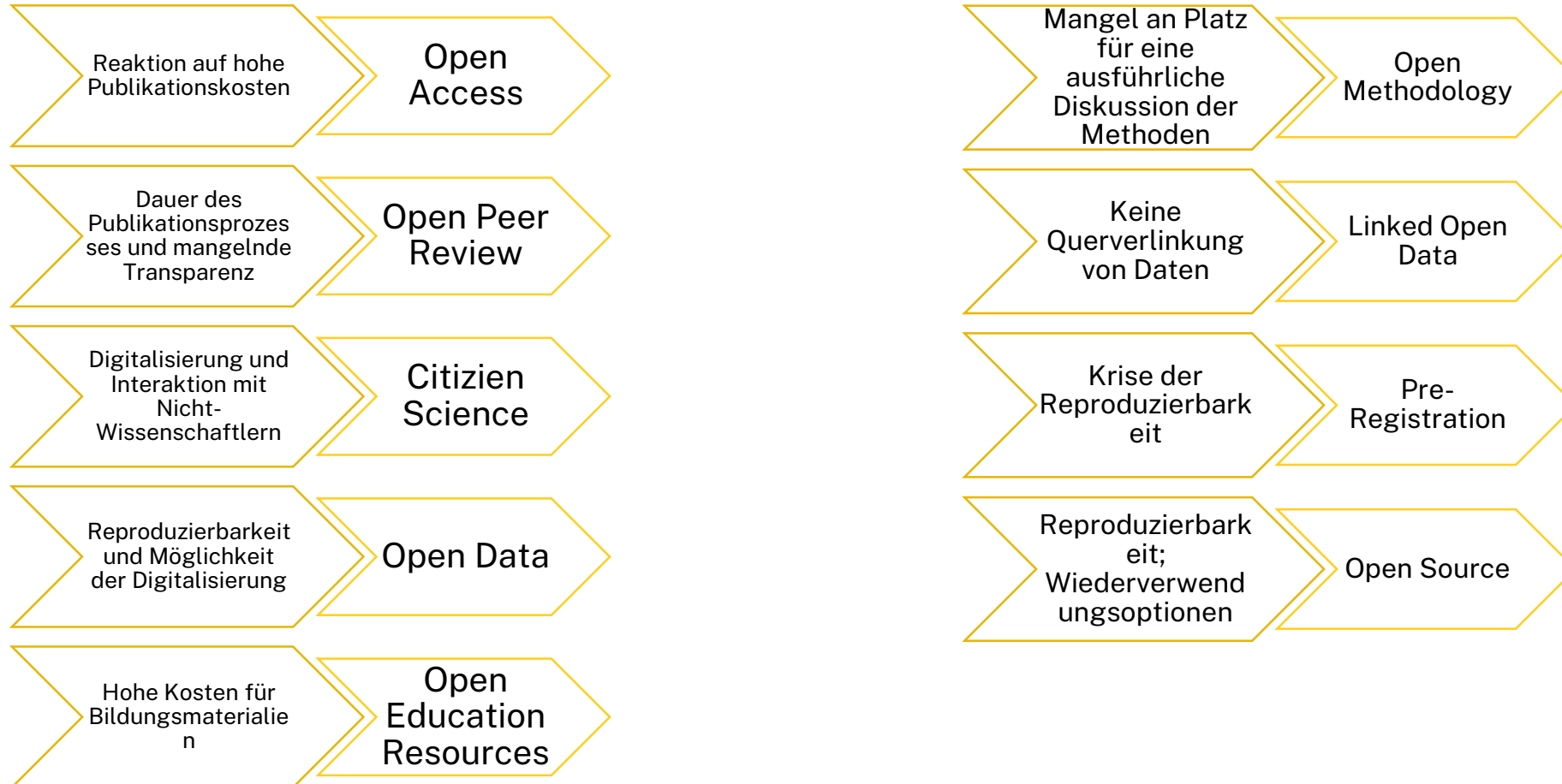
Wechselndes „Weltbild“

- Ähnlich wie in der kopernikanischen Wende (Abkehr vom geozentrischen hin zum heliozentrischen Weltbild)
- Übertragen auf Open Science:
 - Wechsel vom **artikelzentrierten** Blick (Publikation am Ende der Forschung)
 - hin zum **prozessorientierten** Blick (Transparenz und Reproduzierbarkeit aller Schritte, von Anfang mit Präregistrierung bis hin zur Abschlusspublikation)



Nicolaus Copernicus (1473–1543), “De revolutionibus”, fol. 9 verso, Jagiellonische Bibliothek, Krakau, gemeinfrei.

Hypothesis: OS als Reaktion auf ...



Open Science Communities

Example The Netherlands

- <https://www.osc-nl.com>
- <https://www.openscience.nl>



<https://www.osc-nl.com>, CC BY 4.0

Openness in Libraries

“The Commission for Research-Related Services sees the promotion of openness as a central task of libraries. In this sense, it is strongly committed to the establishment and further development of innovative library services, and committed to the principles of openness.”

Openness in Libraries
Position Paper of the Commission for Research-Related Services of the VDB

Alexander Berg-Weiß
University Library of the Ludwig-Maximilians-Universität München
<https://orcid.org/0000-0002-7435-8676>


Sibylle Hermann
Information and Communication Center of the University of Stuttgart
<https://orcid.org/0000-0001-9239-8789>

Miriam Kötter
University of Duisburg-Essen, University Library
<https://orcid.org/0000-0002-7253-2933>

Caroline Leiß
Universitätsbibliothek der TUM
<https://orcid.org/0000-0002-2792-2625>

Christoph Müller
Library of the Ibero-Amerikanisches Institut Stiftung Preußischer Kulturbesitz
<https://orcid.org/0000-0001-9399-6676>

Annette Strauch-Davey
University Library of Hildesheim
<https://orcid.org/0000-0003-2355-325X>



PDF

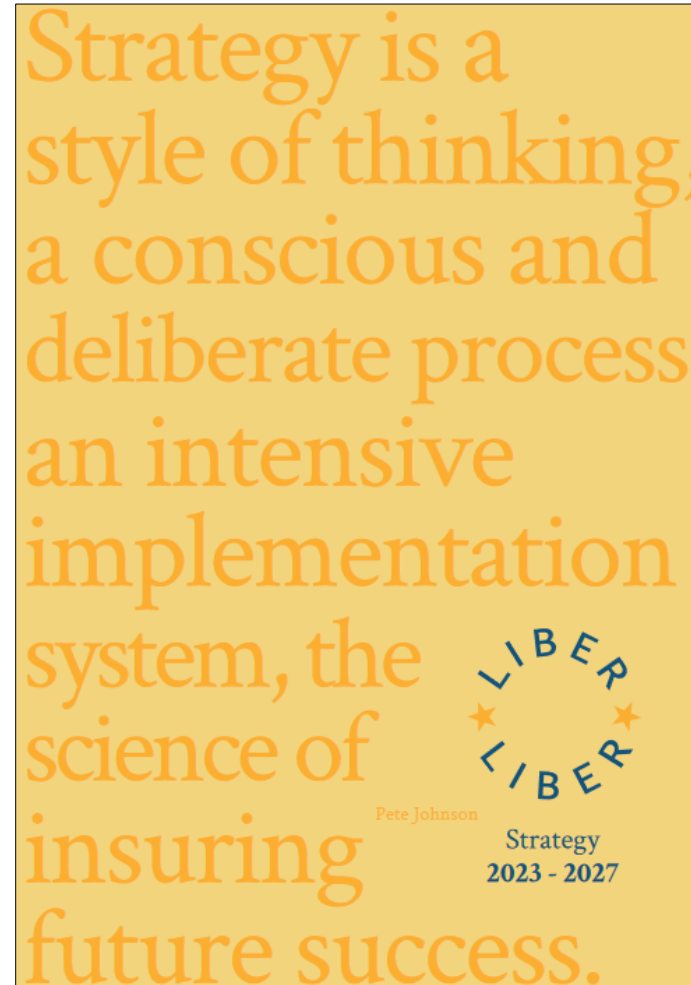
HTML

Berg-Weiß, A., Hermann, S., Kötter, M., Leiß, C., Müller, C., & Strauch-Davey, A. (2022): Openness in Libraries: Position Paper of the Commission for Research-Related Services of the VDB, In: o-bib. Das offene Bibliotheksjournal 9(2), S. 1–4. <https://doi.org/10.5282/o-bib/5834>, CC BY 4.0

LIBER Strategy 2023-2027

“D. ADVANCING OPEN SCIENCE:
By 2027, in collaboration with
researchers, research libraries stimulate,
facilitate, co-develop, and manage
infrastructures and practices designed to
take Open Science to the next level.”

LIBER Strategy 2023-2027,
https://libereurope.eu/wp-content/uploads/2022/01/LIBER_STRAT_A5_digital-final-1.pdf, S. 3



LIBER 2023-2027 Strategy,
CC BY-NC-ND,
https://libereurope.eu/wp-content/uploads/2022/01/LIBER_STRAT_A5_digital-final-1.pdf.

French Open Science Monitor

FRENCH MINISTRY OF HIGHER EDUCATION AND RESEARCH
French Open Science Monitor

en

The general monitor | The health monitor | The local monitors | About

Welcome to
the French Open Science Monitor

Measure the evolution of open science in France using reliable, open and controlled data.

See the last communication →

Glossary

The key figures
Data updated on Dec 1, 2022 with publications released between 2013 and 2021

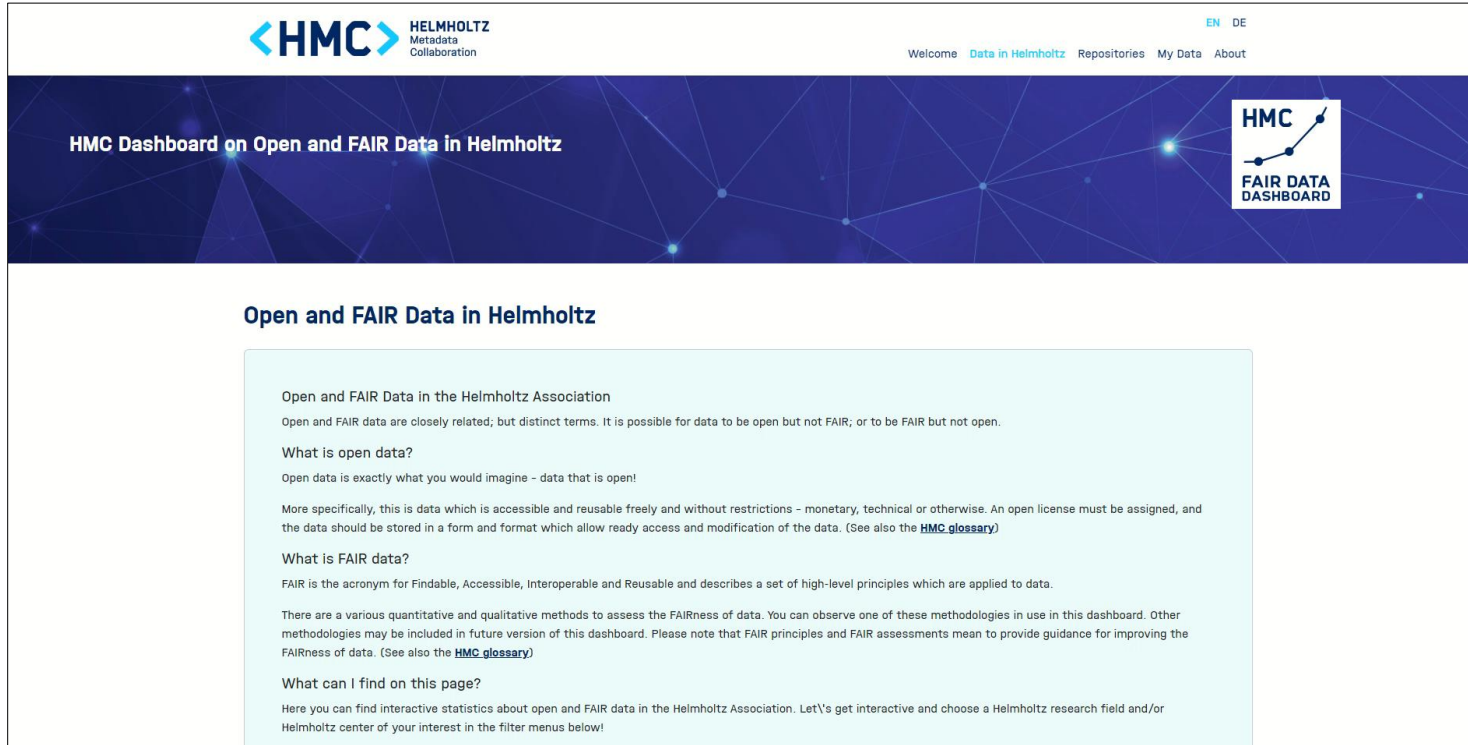
The publications

Publications in **open access** refer to publications from scientific research that are made available online in open access for all, without technical or financial barriers. The French Open Science Monitor focuses on **french publications**, i.e. publications where at least one of the authors is affiliated in France. It is therefore the activity of French research that is taken into account, and not that of French scientific publishers. The open access rate represents the ratio of the number of open access publications to the total number of publications on the same perimeter (e.g. by year, discipline or publisher).
The generalisation of open access to scientific publications is one of the axes of the French national open science strategy, with the objective of a 100% open access rate.

- Datenbeschaffung mit KI (z.B. Softcite)
- Inkludiert Open Data, Open Access und Open Code
- Differenziert zwischen Allgemeinem, PhD und Gesundheitsforschung
- Filtermöglichkeiten wie Verleger
- Webseitenaufbau mit weniger Graphen pro Seite stattdessen auf mehreren Unterseiten verteilt

<https://frenchopensciencemonitor.esr.gouv.fr>

Helmholtz Dashboard on Open and FAIR Data



The screenshot shows the homepage of the Helmholtz Dashboard on Open and FAIR Data. The header features the HMC logo (Helmholtz Metadata Collaboration) and navigation links for Welcome, Data in Helmholtz, Repositories, My Data, and About. The main content area is titled "Open and FAIR Data in Helmholtz" and contains several sections of text explaining the concepts of open and FAIR data, their relationship, and how to find interactive statistics on the site.

Open and FAIR Data in Helmholtz

Open and FAIR Data in the Helmholtz Association

Open and FAIR data are closely related; but distinct terms. It is possible for data to be open but not FAIR; or to be FAIR but not open.

What is open data?

Open data is exactly what you would imagine – data that is open!

More specifically, this is data which is accessible and reusable freely and without restrictions – monetary, technical or otherwise. An open license must be assigned, and the data should be stored in a form and format which allow ready access and modification of the data. (See also the [HMC glossary](#))

What is FAIR data?

FAIR is the acronym for Findable, Accessible, Interoperable and Reusable and describes a set of high-level principles which are applied to data.

There are a various quantitative and qualitative methods to assess the FAIRness of data. You can observe one of these methodologies in use in this dashboard. Other methodologies may be included in future version of this dashboard. Please note that FAIR principles and FAIR assessments mean to provide guidance for improving the FAIRness of data. (See also the [HMC glossary](#))

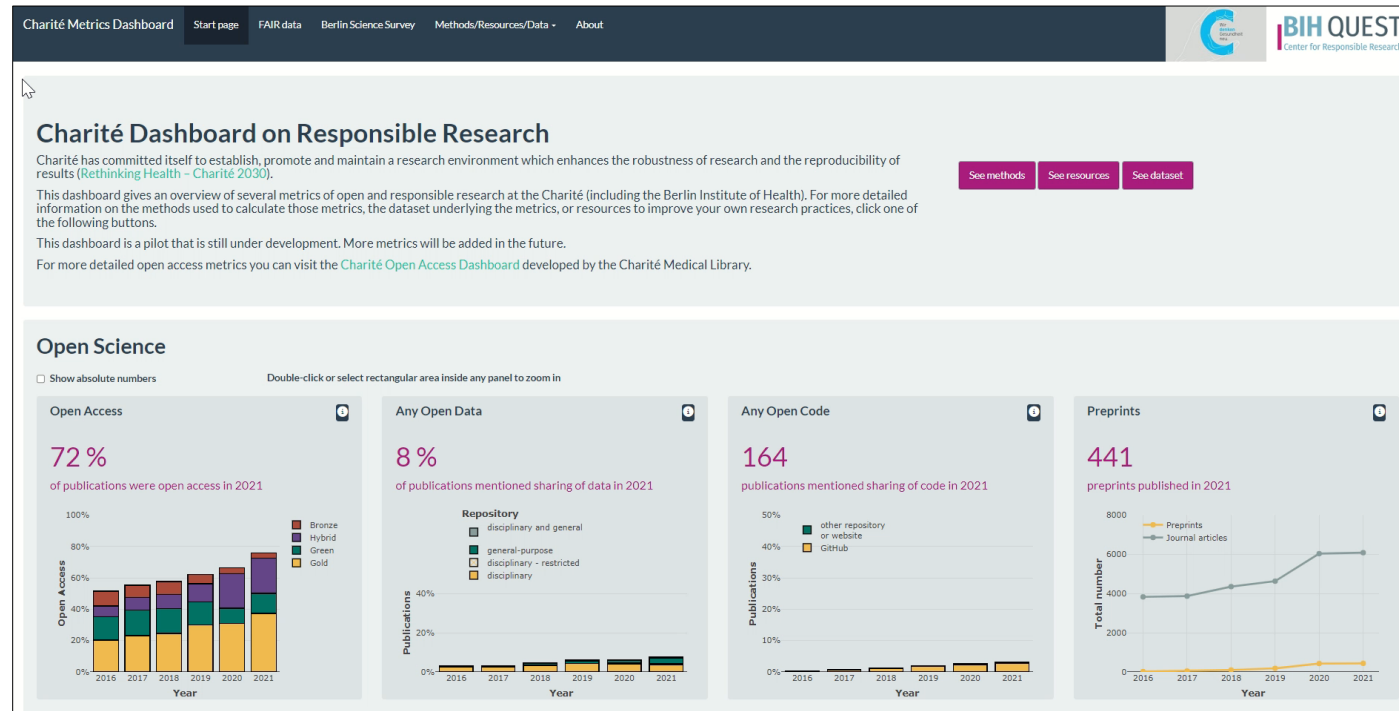
What can I find on this page?

Here you can find interactive statistics about open and FAIR data in the Helmholtz Association. Let's get interactive and choose a Helmholtz research field and/or Helmholtz center of your interest in the filter menus below!

- Nutzt Third-Party APIs (z. B. F-UJI API) für die Datenerhebung
- Enthält Informationen zu Veröffentlichungen und FAIR-Bewertungen
- Webseitenaufbau mit weniger Graphen pro Seite stattdessen auf mehreren Unterseiten verteilt

<https://fairdashboard.helmholtz-metadaten.de>

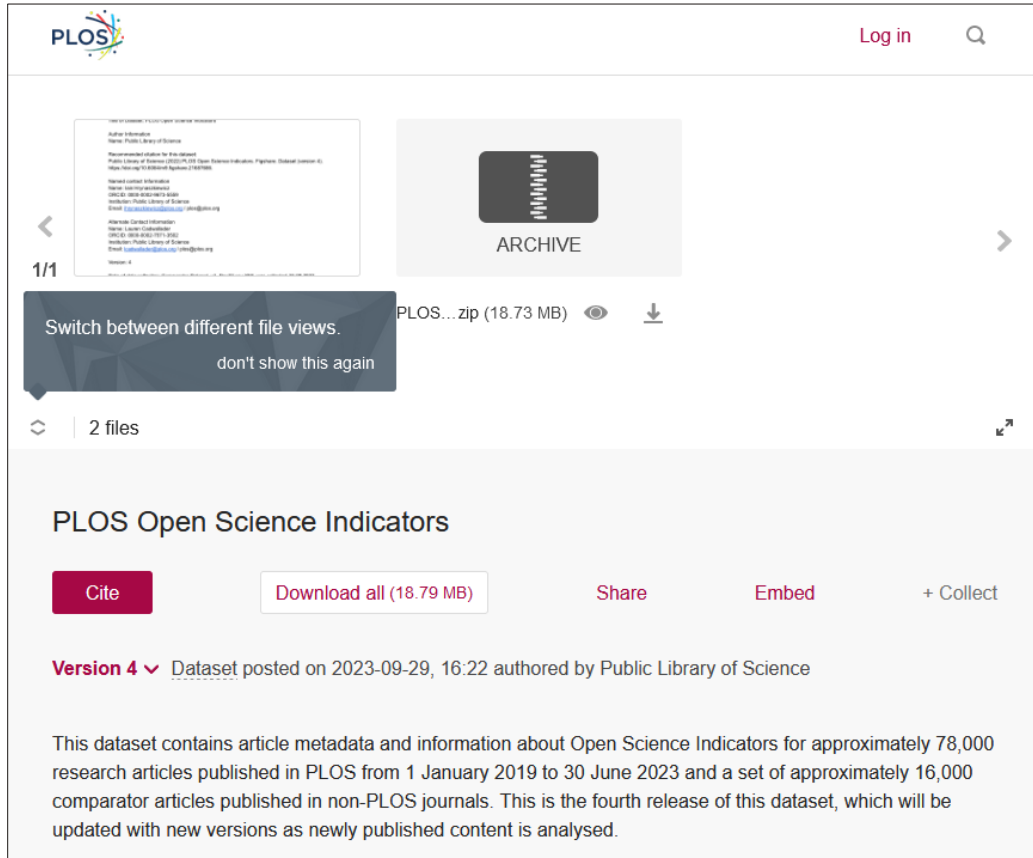
Charité Open Science Monitor



- Datenerhebung durch Nutzung bestehender Lösungen (z.B. ODDPub)
- Beinhaltet Daten zu Open Data, Open Code und den FAIR-Prinzipien
- Verwendung einer „Cockpit“ Seite, die eine Vielzahl von Graphen auf einer einzigen Seite zusammenführt

<https://quest-dashboard.charite.de>

PLOS Open Science Indicators



1/1

Switch between different file views. don't show this again

PLOS...zip (18.73 MB)

2 files

PLOS Open Science Indicators

Cite Download all (18.79 MB) Share Embed + Collect

Version 4 Dataset posted on 2023-09-29, 16:22 authored by Public Library of Science

This dataset contains article metadata and information about Open Science Indicators for approximately 78,000 research articles published in PLOS from 1 January 2019 to 30 June 2023 and a set of approximately 16,000 comparator articles published in non-PLOS journals. This is the fourth release of this dataset, which will be updated with new versions as newly published content is analysed.

- Identifizieren und quantifizieren bestimmte Open-Science-Praktiken in der veröffentlichten PLOS-Literatur
- Verwendung von KI (z.B. [DataSeer](#))
- ca. 78 000 ausgewertete PLOS Journal Article

<https://doi.org/10.6084/m9.figshare.21687686.v4>

Brainwriting Pool

Frage:

Wie wird Open Science die deutschen Bibliotheken zukünftig prägen?

Methode:

Gemeinsames Schreiben an drei Textabschnitte

Einteilung:

Reihe 1 übernimmt Teil A, Reihe 2 B und Reihe 3 + 4 C

Werkzeug:

https://pad.gwdg.de/s/5kKpH_EPz# in Markdown

Dauer:

15 Minute (alle 5 Minuten Wechsel A → B → C)

Ziel:

Kreative Annäherung an das Thema, gemeinsamer Text

Und was schreibt ChatGPT dazu?

von <https://chat.openai.com>

Frage: “Wie wird Open Science die deutschen Bibliotheken zukünftig prägen?”

Antwort:

Open Science hat das Potenzial, die deutschen Bibliotheken zukünftig in verschiedener Hinsicht zu prägen. Einerseits kann es dazu beitragen, die Verfügbarkeit wissenschaftlicher Erkenntnisse und Ressourcen zu erhöhen und die Zugänglichkeit von Wissen für alle zu verbessern.

Andererseits kann es auch dazu beitragen, die Rolle der Bibliotheken als Orte der Wissensvermittlung und des Austauschs von Wissen zu stärken. Es kann auch die Möglichkeiten der Bibliotheken erweitern, Wissenschaftler bei der Verbreitung und Veröffentlichung ihrer Arbeiten zu unterstützen. Insgesamt kann Open Science dazu beitragen, die Bibliotheken in ihrer Rolle als zentrale Anlaufstellen für wissenschaftliches Wissen zu stärken und zukunftsfähig zu machen.

Blablameter-Index: 0.46

Wissenschaftliche Software

Drei Säulen an wissenschaftlichen Ergebnissen?

Text

- Journal Article
- Book
- Poster Presentation
- ...

Data

- Research Data
- Laboratory Data
- ...

Code

- Software
- Software Documentation
- ...

Second French Plan for Open Science

Path Three “Opening up and promoting source code produced by research”

French Ministry of Higher Education, Second French Plan for Open Science. Generalizing open science in France 2021-2024, July 2021, https://www.ouvrirlascience.fr/wp-content/uploads/2021/10/Second_French_Plan_for-Open-Science_web.pdf pp. 1 and 16-19, CC BY 4.0.



Wissenschaftliche Software

Begrifflichkeit:

- Eigenentwickelte Forschungssoftware
- Softwareanwendungen für die Forschung
- Infrastruktursoftware/-dienste

Wissenschaftliche Software

Gemeinsamkeiten von Software und Forschungsdaten im **Forschungsprozess:**

- beides spielt eine immer bedeutendere Rolle in der Forschung
- beides ist notwendig, um Forschungsergebnisse reproduzierbar zu machen
- beides stellt potentiell einen Wert für Nachnutzer dar
- beides wird bisher nicht (oder nur in geringem Maße) als eigenständige Forschungsleistung anerkannt

Wissenschaftliche Software

Gemeinsamkeiten von Software und Forschungsdaten im **Management**:

- beides kann mit Metadaten versehen und archiviert werden
- beides muss kuratiert werden, um die Nutzbarkeit über längere Zeiträume zu gewährleisten
- bei beidem fehlt oft das notwendige Wissen für ein adäquates Management

Wissenschaftliche Software

Unterschiede von Software zu Forschungsdaten:

- die Kuratierung von Software ist wesentlich aufwändiger
- [Metadaten für Software](#) sind wesentlich homogener als die für Forschungsdaten
- bei Software gibt es viel Vorerfahrung mit offenen/kommerziellen [Lizenzen](#)
- eine genaue Versionierung spielt bei Software eine wesentlich größere Rolle

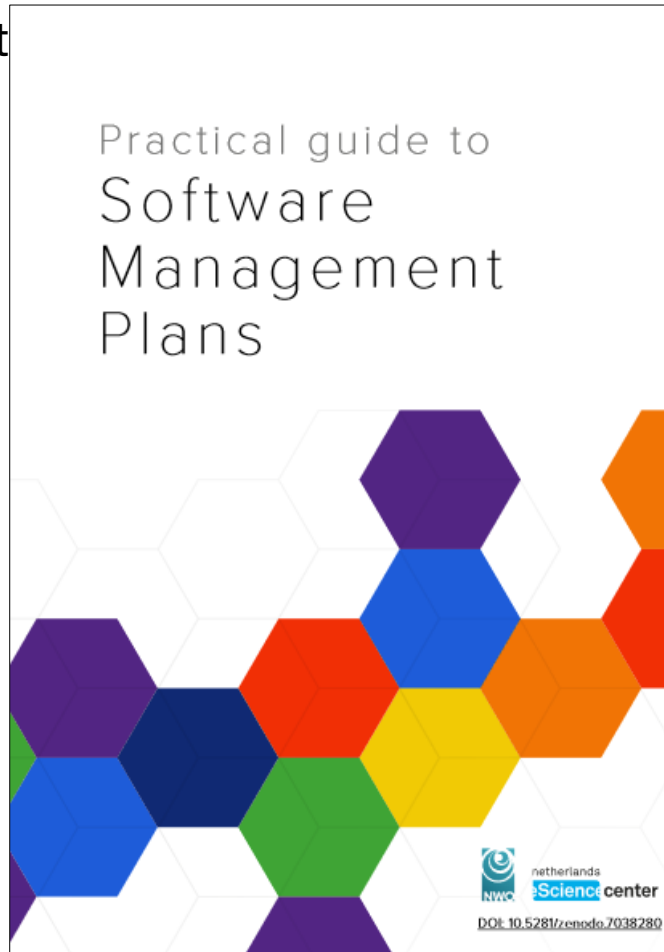
Wissenschaftler_innen schreiben wissenschaftliche Software

Beobachtungen:

- häufig keine ausgebildeten Software-Entwickler_innen, eher Autodidakten
- Funktionalität geht vor Dokumentation geht vor langfristiger Erhaltung
- erst Textpublikation, dann lange nichts, und dann vielleicht Datenpublikation und Software-Publikation
- Software wird von Generation zu Generation „vererbt“
- ...

SMP aktuell in Diskussion

Martinez-Ortiz et al. (27.10.2022): Practical guide to Software Management Plans, v1.0, <https://doi.org/10.5281/zenodo.7248877>, CC BY 4.0.



DFG: **Call for Proposals to Increase the Usability of Existing Research Software** durch „Research Software – Quality Assured and Re-usable“, (Information für die Wissenschaft Nr. 85 | 8. November 2022) http://www.dfg.de/en/research_funding/announcements_proposals/2022/info_wissenschaft_22_85.

FAIR Principles for Research Software

- FAIR4RS Principles
- Adaption der FAIR Data Principles für Forschungssoftware
- von der Community unterstützte Grundsätze

FAIR Principles for Research Software
(FAIR4RS Principles)



RDA Recommendation

DOI: 10.15497/RDA00068

Authors: Neil F. Chue Hong*, Daniel S. Katz*, Michelle Barker*, Anna-Lena Lamprecht, Carlos Martinez, Fotis E. Psomopoulos, Jen Harrow, Leyla Jael Castro, Morane Gruenpeter, Paula Andrea Martinez, Tom Honeyman, Alexander Struck, Allen Lee, Axel Loewe, Ben van Werkhoven, Catherine Jones, Daniel Garijo, Esther Flomp, Françoise Genova, Hugh Shanahan, Joanna Leng, Maggie Hellström, Malin Sandström, Manodeep Sinha, Mateusz Kuzak, Patricia Herterich, Qian Zhang, Sharif Islam, Susanna-Assunta Sansone, Tom Pollard, Udayanto Dwi Atmojo, Alan Williams, Andreas Czerniak, Anna Niehues, Anne Claire Fouilloux, Bala Desinghu, Carole Goble, Céline Richard, Charles Gray, Chris Erdmann, Daniel Nüst, Daniele Tartarini, Elena Rangelova, Hartwig Anzt, Ilian Todorov, James McNally, Javier Moidon, Jessica Burnett, Julián Garrido-Sánchez, Khalid Belhajjame, Laurents Sesink, Lorraine Hwang, Marcos Roberto Tovani-Paloni, Mark D. Wilkinson, Mathieu Servillat, Matthias Liffers, Merc Fox, Nadica Miljković, Nick Lynch, Paula Martinez Lavandhy, Sandra Gesing, Sarah Stevens, Sergio Martinez Cuesta, Silvio Peroni, Stian Soiland-Reyes, Tom Bakker, Tovo Rabemanantsoa, Vanessa Sochat, Yo Yehudi, [FAIR4RS WG](#)

(*) lead authors with equal contributions

Published: 24th May 2022
Version: 1.0

Abstract: To improve the sharing and reuse of research software, the FAIR for Research Software (FAIR4RS) Working Group has applied the FAIR Guiding Principles for scientific data management and stewardship to research software, bringing together existing and new community efforts. Many of the FAIR Guiding Principles can be directly applied to research software by treating software and data as similar digital research objects. However, specific characteristics of software — such as its executability, composite nature, and continuous evolution and versioning — make it necessary to revise and extend the principles.

This document presents the first version of the FAIR Principles for Research Software (FAIR4RS Principles), and includes explanatory text to aid adoption. It is an outcome of the FAIR for Research Software Working Group (FAIR4RS WG) based on community consultations that started in 2019.

The FAIR for Research Software Working Group was jointly convened as a Research Data Alliance (RDA) Working Group, FORCE11 Working Group, and Research Software Alliance (ReSA) Task Force. Going forward, the RDA Software Source Code Interest Group is the maintenance home for the principles. Concerns or queries about the principles can be raised at RDA plenary events organized by the SSC IG, where there may be opportunities for adopters to report back on progress. The full maintenance and retirement plan for the principles can be found on the RDA website.

Chue Hong et al.
(2022): FAIR
Principles for
Research Software
(FAIR4RS
Principles, Zenodo,
CC BY 4.0,
<https://doi.org/10.15497/RDA00068>.

Unterschiede FAIR Data ↔ FAIR4RS

Findable: Software, and its associated metadata, is easy for both humans and machines to find

- (=) F1 Software is assigned a globally unique and persistent identifier
 - (new) F1.1 Components of the software representing levels of granularity are assigned distinct identifiers
 - (new) F1.2 Different versions of the software are assigned distinct identifiers
- (=) F2 Software is described with rich metadata
- (=) F3 Metadata clearly and explicitly include the identifier of the software they describe
- (=) F4 Metadata are FAIR, searchable and indexable

Accessible: Software, and its metadata, is retrievable via standardised protocols

- (=) A1 Software is retrievable by its identifier using a standardised communications protocol
 - (=) A1.1 The protocol is open, free, and universally implementable
 - (=) A1.2 the protocol allows for an authentication and authorization procedure, where necessary
- (=) A2 metadata are accessible, even when the software is no longer available

Interoperable: Software interoperates with other software by exchanging data and/or metadata, and/or through interaction via application programming interfaces (APIs), described through standards

- (≠) I1. Software reads, writes and exchanges data in a way that meets domain-relevant community standards
- ~~I2. (meta)data use vocabularies that follow FAIR principles~~
- (=) I2~~3~~. Software includes qualified references to other objects

Reusable: Software is both usable (can be executed) and reusable (can be understood, modified, built upon, or incorporated into other software)

- (=) R1. Software is described with a plurality of accurate and relevant attributes
 - (=) R1.1. Software is given a clear and accessible license
 - (=) R1.2. Software is associated with detailed provenance
- (new) R2 Software includes qualified references to other software
- (=) R1~~3~~. Software meets domain-relevant community standards

Policies für wissenschaftliche Software

- kann helfen, Standardverfahren zu etablieren
- kann bei der Einrichtung von Veröffentlichungsprozessen helfen
- kann bei der Auswahl von Lizenzen helfen
- kann bei der Klärung der Eigentumsverhältnisse helfen
- kann bei der Kommerzialisierung der Software helfen
- kann bei der Wartung der Software helfen
- kann helfen, die Software in eine Gemeinschaft einzubringen
- kann die Akzeptanz der Veröffentlichung von Software als wissenschaftlicher Wert verbessern

Beispiel für eine Policy

TU Delft

<https://doi.org/10.5281/zenodo.4629635>

- Licensing
- Registration
- Commercialisation

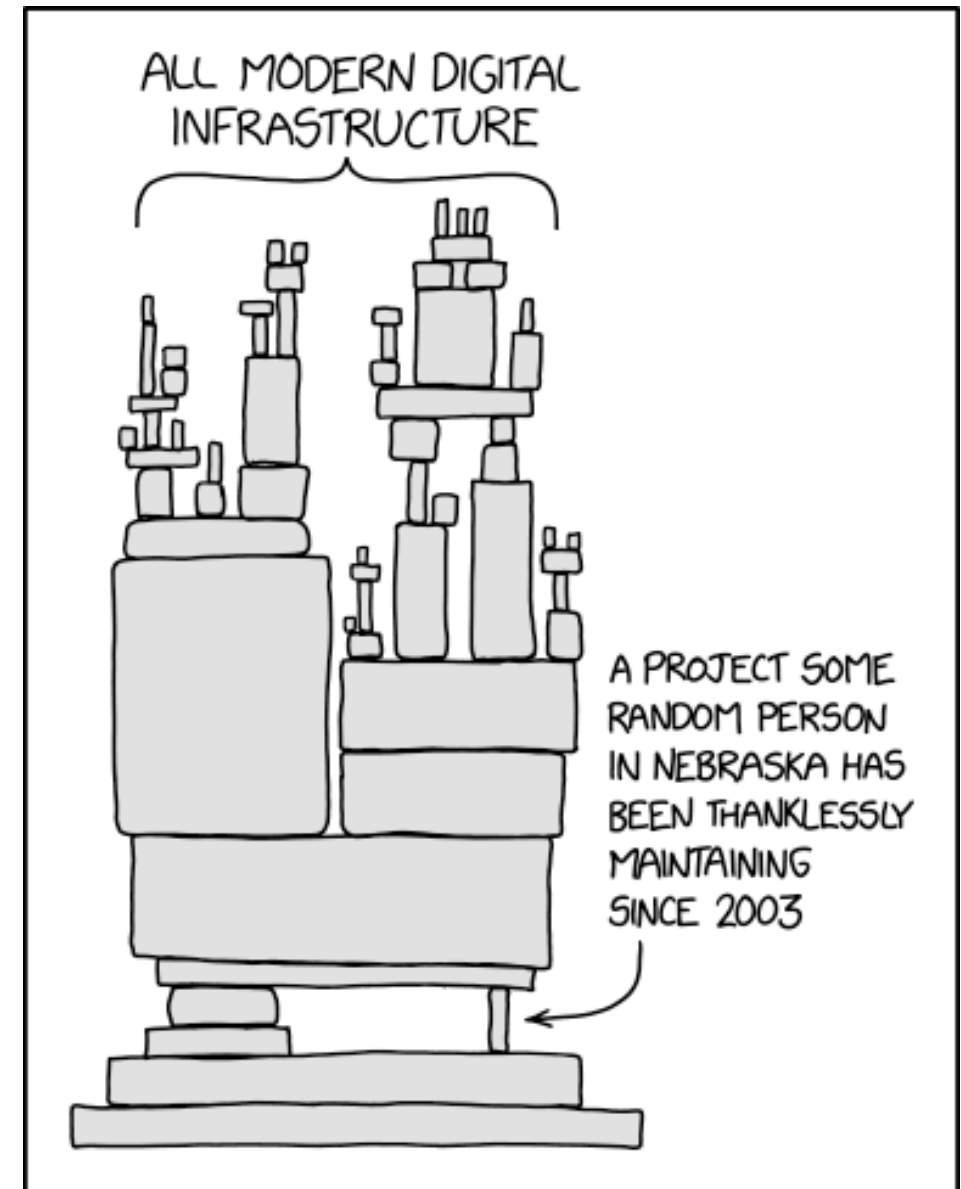


Bazuine et al. (2021): TU Delft Guidelines on Research Software: Licensing, Registration and Commercialisation, CC BY 4.0, <https://doi.org/10.5281/zenodo.4629635>.

Open Source Software

Mögliche Vorteile von Open Source:

1. Software ist überprüfbar
2. Flexibilität bei der Entwicklung (eigenes Hosting, lokal oder als Cloud)
3. reduzierte Kosten
4. offen
5. Veröffentlichung der Software



xkcd, Dependency, <https://xkcd.com/2347/>, CC BY-NC 2.5

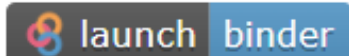
Literaturempfehlung

- Frank Löffler, Christian Busse, und Stephan Janosch: Forschungssoftware in Deutschland, gehalten auf der RDA Deutschland Tagung 26. Februar 2021, <https://doi.org/10.5281/zenodo.4564161>.
- Messerschmidt et al. (2021): Checkliste zur Unterstützung der Helmholtz-Zentren bei der Implementierung von Richtlinien für nachhaltige Forschungssoftware, Potsdam, <https://doi.org/10.48440/os.helmholtz.031>.
- Pampel et al. (2022): Helmholtz Open Science Briefing. Helmholtz Open Science Forum Forschungssoftware. Report, Potsdam, <https://doi.org/10.48440/os.helmholtz.045>.

Live Demonstration Jupyter Notebook

DataCite_Value-Parser

A short Jupyter Notebook to parse DataCite Values via REST API for Statistics

 launch binder

DOI 10.5281/zenodo.7257106

https://mybinder.org/v2/gh/yvgrossmann/DataCite_Value-Parser/HEAD

Aufgabe:

Öffnen Sie das jypnb in Mybinder und lassen es mit einer ROR-ID Ihrer Institution laufen

Abschließendes

- Folien werden zugänglich gemacht
- Evaluation: Frau Schindler meldet sich diesbezüglich bei Ihnen

Ende