

## Woche der Forschungskompetenzen 2025

# *Forschungsdaten nutzen und verwalten*

07.10.2025

*Johanna Vompras*

Kompetenzzentrum Forschungsdaten  
[data@uni-bielefeld.de](mailto:data@uni-bielefeld.de)



Bildquelle: Jørgen Stamp,  
<https://digitalbevaring.dk/>

# Programm

- Verständnis von Forschungsdaten
- Empfehlungen und Anforderungen
- Daten finden
- Daten nachnutzen
- Daten organisieren, dokumentieren und sichern
- Abschluss

# Programm

- Verständnis von Forschungsdaten
- Empfehlungen und Anforderungen
- Daten finden
- Daten nachnutzen
- Daten organisieren, dokumentieren und sichern
- Abschluss

# Forschungsdaten

## DFG (Deutsche Forschungsgemeinschaft):

„Zu Forschungsdaten zählen u.a. **Messdaten, Laborwerte, audiovisuelle Informationen, Texte, Surveydaten, Objekte aus Sammlungen** oder **Proben**, die in der wissenschaftlichen Arbeit entstehen, entwickelt oder ausgewertet werden.

Methodische Testverfahren, wie Fragebögen, Software und Simulationen **können ebenfalls** zentrale Ergebnisse wissenschaftlicher Forschung darstellen und sollten daher ebenfalls unter den Begriff Forschungsdaten gefasst werden.“ [1]

## Weitere allgemeine Definition (University of Leeds Library):

“Research data is any information that has been collected, observed, generated or created to validate original research findings.” [2]



[1] [www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien\\_forschungsdaten.pdf](http://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten.pdf)

[2] <https://library.leeds.ac.uk/info/14062/research-data-management/61/research-data-management-explained>

# Quellen von Forschungsdaten

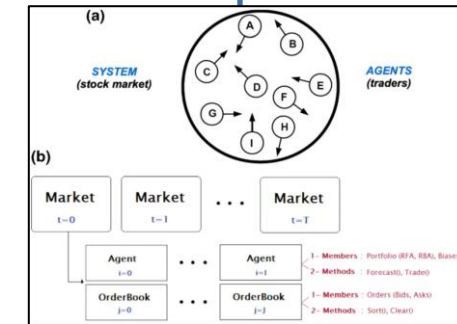
Beobachtungsdaten,  
z. B. Umfragen oder  
Verhalten



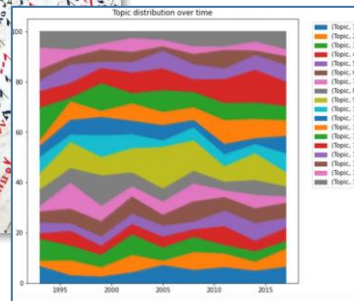
Experimentelle Daten,  
z. B. Labor



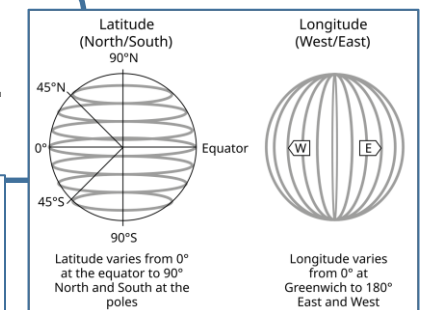
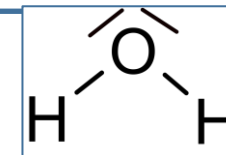
Simulationsdaten,  
z. B. ökonomische  
Modelle




Abgeleitete oder  
zusammengestellte  
Daten, z. B. Quellen oder  
Textmining





Referenz- oder  
kanonische Daten, z. B.  
chemische Strukturen





# Arten von Forschungsdaten


- 
- Dokumente, Tabellenkalkulationen
  - Labornotizbücher, Feldnotizbücher, Tagebücher
  - Fragebögen, Protokolle, Codebücher

- 
- Audio-/Videodateien/-bänder
  - Fotografien, Filme

- 
- Antworten auf Tests
  - Dias, Artefakte, Proben, Muster
  - Sammlungen von digitalen Ergebnissen

- 
- Daten-Files
  - Inhalt einer Datenbank

- 
- Modelle, Algorithmen, Skripte
  - Inhalt einer Anwendung

- 
- Methodologien, Arbeitsabläufe
  - Standardarbeitsanweisungen und -protokolle

- 
- ...
  - ...

# Datenformate (Beispiel: Soziologie)

Verschiedene Quellen/Typen von Daten		Datenformate
Quantitative Daten	Daten aus Erhebungen, Volkszählungen, ... ( <b>numerische</b> Daten)	Tabellenkalkulationsformate, z. B. CSV, Excel (.xlsx)
	Erhebungsinstrumente und Fragebögen ( <b>textuelle</b> Daten)	Textformate, z. B. PDF, Word (.docx), offene Textformate (.odt) oder einfache Textdateien (.txt)
	Instrumente zur Datenanalyse ( <b>statistische</b> Daten)	RData, SPSS, Stata, ... oder allgemein gebräuchliche Formate: CSV, Excel (.xlsx)
Qualitative Daten	Daten, die durch Interviews, Fokusgruppen, ethnografische Studien gesammelt wurden, z. B. Transkripte, Feldnotizen ( <b>textuelle</b> Daten)	Textformate, z. B. PDF, Word (.docx), offene Textformate (.odt) oder einfache Textdateien (.txt)
	Aufzeichnungen der oben genannten Daten ( <b>Audio-/Videodaten</b> )	MP3, MP4, WAV, AVI, ...
	Fotografien oder Bilder ( <b>visuelle</b> Daten)	JPEG, PNG, TIFF
	Instrumente zur Datenanalyse ( <b>qualitative</b> Daten)	MAXQDA (.mx18, .mx19), Atlas.ti (*.atlproj) oder allgemein übliche Formate: CSV, TXT, ...

# Datenformate (Beispiel: Chemodiversität von Pflanzen)

Verschiedene Quellen/Typen von Daten	Datenformate
Beschreibende Informationen/ Dokumentation/ Beobachtungen ( <b>textuelle</b> Daten)	XML, CSV, TXT, Markdown/txt (Wiki), programmspezifisch (Word-Dateien, ...)
Strukturierte Daten in Tabellenkalkulationen oder Datenbanken ( <b>tabellarische</b> Daten)	Datenmatrizen (in XLSX, CSV, SQL-Datenbanken)
Fotografische oder mikroskopische Bilder ( <b>Bilddaten</b> )	JPEG, PNG, TIFF
Sequenzen und Annotationen ( <b>genomische</b> Daten)	FASTA, FASTQ (rohe Sequenzlesungen) und GFF, BED (genomische Annotationen)
Rohe/verarbeitete Daten z. B. aus der Massenspektrometrie ( <b>Spektral</b> daten)	mzML, mzXML, oder proprietäre Formate
Daten mit geografischen Komponenten ( <b>Geodaten</b> )	Shapefiles (SHP), GeoJSON, KML
Instrumente zur Datenanalyse ( <b>statistische</b> Daten)	RData, SPSS, MATLAB, Stata, ... oder allgemein gebräuchliche Formate: CSV, Excel



# Komplexität von Forschungsdaten

- **die empirische Basis** für wissenschaftliche Analysen weitet sich kontinuierlich aus
- stetige Zuwachs an verfügbaren Daten, durch **methodische** und **technische Neuerungen** (z.B. Social Media, stetig wachsende Datenquellen)
- **Komplexität** der Daten deutlich **erhöht**, durch
  - Verknüpfung von Daten anderer Disziplinen
  - unendliche Vielzahl an Weiternutzungsmöglichkeiten

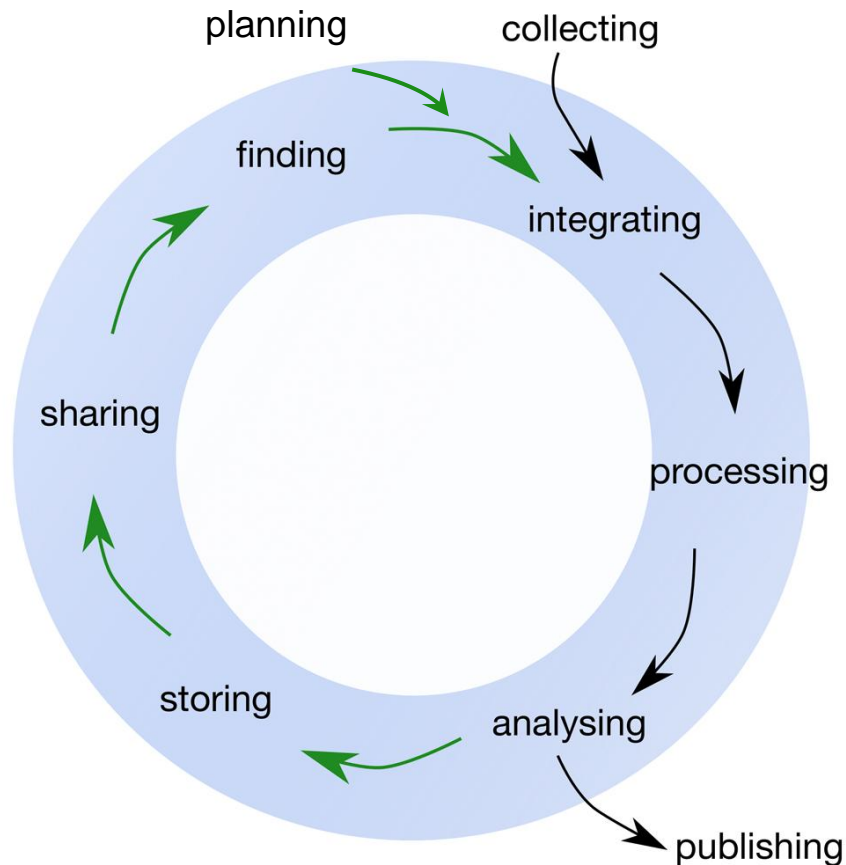


Bildquelle: KI-generiert: Wissenschaftler/in untersucht externe Datenquellen

# Programm

- Verständnis von Forschungsdaten
- **Empfehlungen und Anforderungen**
- Daten finden
- Daten nachnutzen
- Daten organisieren, dokumentieren und sichern
- Abschluss

# Aspekte rund um den Data Life Cycle



## Data Life Cycle

Vereinfachte **Repräsentation** eines jeden Forschungsprojektes

Forschungsdatenmanagement (FDM): die **gesamte Organisation** und **Verwaltung der Prozesse** zusammen, die in Zusammenhang mit der Entstehung, Verarbeitung, Dokumentation, Weitergabe und der nachhaltigen Archivierung von Forschungsdaten stehen.

# Was erwartet/bietet die

- Checkliste „Umgang mit Forschungsdaten“  
<https://www.dfg.de/resource/blob/174732/for>
- Fachspezifische Empfehlungen: <https://www.dfg.de/resource/blob/174732/for>
- In vielen Förderprogrammen sind Angaben zu den Daten (Aspekte: z. B. Umfang der Daten, Dauer der Erhebung, spätere Wiederverwendung und langfristige Verfügbarkeit) erforderlich
  - **Opt Out** bezüglich der Datenweitergabe ist nicht möglich  
→ **Erklärung**
- Aspekte, die möglicherweise bereits in der Projektbeschreibung zu beachten!

## Fachspezifische Empfehlungen

Im Folgenden finden Sie Fachkollegien der DFG, die sich mit dem Umgang mit Forschungsdaten befassen.

### Geistes- und Sozialwissenschaften

- Fachkollegium "Philosophie" (2022)
- Fachkollegium "Rechtswissenschaften" (2020)
- Fachkollegium "Historische Linguistik und Sprachwissenschaft"
- Fachkollegium "Sprachwissenschaft und Kommunikationswissenschaft"
- ...

### Lebenswissenschaften

- Fachkollegien "Phyto- und Biodiversitätsforschung"

### Naturwissenschaften

- Fachkollegium "Mathematik"
- Fachkollegien "Physik" zum Umgang mit Forschungsdaten im Fachbereich Physik (2023)

## Checkliste zum Umgang mit Forschungsdaten

- Datenbeschreibung**  
 Auf welche Weise entstehen in Ihrem Projekt neue Daten? Werden existierende Daten wiederverwendet? Welche Datentypen, im Sinne von Datenformaten (z. B. Bilddaten, Textdaten oder Messdaten) entstehen in Ihrem Projekt und auf welche Weise werden sie weiterverarbeitet? In welchem Umfang fallen diese an bzw. welches Datenvolumen ist zu erwarten?
- Dokumentation und Datenqualität**  
 Welche Ansätze werden verfolgt, um die Daten nachvollziehbar zu beschreiben (z. B. Nutzung vorhandener Metadaten- bzw. Dokumentationsstandards oder Ontologien)? Welche Maßnahmen werden getroffen, um eine hohe Qualität der Daten zu gewährleisten? Sind Qualitätskontrollen vorgesehen und wenn ja, auf welche Weise? Welche digitalen Methoden und Werkzeuge (z. B. Software) sind zur Nutzung der Daten erforderlich?
- Speicherung und technische Sicherung während des Projektverlaufs**  
 Auf welche Weise werden die Daten während der Projektlaufzeit gespeichert und gesichert? Wie wird die Sicherheit sensibler Daten während der Projektlaufzeit gewährleistet (Zugriffs- und Nutzungsverwaltung)?
- Rechtliche Verpflichtungen und Rahmenbedingungen**  
 Welche rechtlichen Besonderheiten bestehen im Zusammenhang mit dem Umgang mit Forschungsdaten in Ihrem Projekt? Sind Auswirkungen oder Einschränkungen in Bezug auf die spätere Veröffentlichung bzw. Zugänglichkeit zu erwarten? Auf welche Weise werden nutzungs- und urheberrechtliche Aspekte sowie Eigentumsfragen berücksichtigt? Existieren wichtige wissenschaftliche Kodizes bzw. fachliche Normen, die Berücksichtigung finden sollten?
- Datenaustausch und dauerhafte Zugänglichkeit der Daten**  
 Welche Daten bieten sich für die Nachnutzung in anderen Kontexten besonders an? Nach welchen Kriterien werden Forschungsdaten ausgewählt, um diese für die Nachnutzung durch andere zur Verfügung zu stellen? Planen Sie die Archivierung Ihrer Daten in einer geeigneten Infrastruktur? Falls ja, wie und wo? Gibt es Sperrfristen? Wann sind die Forschungsdaten für Dritte nutzbar?
- Verantwortlichkeiten und Ressourcen**  
 Wer ist verantwortlich für den adäquaten Umgang mit den Forschungsdaten (Beschreibung der Rollen und Verantwortlichkeiten innerhalb des Projekts)? Welche Ressourcen (Kosten; Zeit oder anderes) sind erforderlich, um einen adäquaten Umgang mit Forschungsdaten im Projekt umzusetzen? Wer ist nach Ende der Laufzeit des Projekts für das Kuratieren der Daten verantwortlich?

Quelle: DFG Checkliste "Umgang mit Forschungsdaten"

# Forschungsdaten Policy der Uni Bielefeld

<https://www.uni-bielefeld.de/ub/digital/forschungsdaten/policy/>

2011: Grundsätze zum Umgang mit Forschungsdaten

Forschende .. [ ] .. sollen  
Forschungsdaten als  
**wertvolles  
wissenschaftliches Werk**  
behandeln ...

2013: Resolution Forschungsdaten

Die Universität Bielefeld  
unterstützt [Forschende], die  
Grundsätze .. [ ] .. im Sinne  
**eines verbindlichen und  
qualitätsbewussten**  
Forschungsdatenmana-  
gements **umzusetzen.**

2021: Sicherung guter wiss. Praxis

**§12:** Herstellung von  
öffentlichem **Zugang** zu  
Forschungsergebnissen  
**§13:** Autorenschaft: .. [ ] ..  
wissenschaftlichen Text-,  
**Daten-** oder  
**Softwarepublikation ...“**

**§16:** **Archivierung** von  
Forschungsdaten .. [ ] ..

Universität Bielefeld



# FAIR Data

## Subject-specific Recommendations on the Handling of Research Data

Here you will find current subject-specific statements developed in cooperation with DFG committees.

### Humanities and social sciences

- "Psychology" Review Board on the handling of research data (2020) →
- "Educational Research" Review Board on the handling of research data (2020) (in German only) →
- "Social Sciences" Review Board on the handling of research data in sociology, political science and communication science (2020) (in German only) →
- "Ancient Cultures" Review Board on the handling of research data (2020) (in German only) →
- "Social and Cultural Anthropology, Non-European Cultures, Jewish Studies and Religious Studies" Review Board on the handling of research data (2019) (in German only) →
- "Economics" Review Board on the handling of research data (2025) →
- "Linguistics" Review Board on data standards and tools in the collection of language corpora (2019) (in German only) →
- "Literary Studies" Review Board on funding criteria for scholarly editions in literary studies (2015) (in German only) →

### Life sciences

- "Plant Sciences" and "Zoology" Review Boards on the handling of research data in biodiversity research (2015) (in German only) →

#### Further information

- Handling of Research Data →
- Research funding →
- Infrastructure funding →
- Resources available →
- FAQ →
- DFG areas of activity →
- Subject-specific recommendations →

## Stellungnahme des Fachkollegiums 111 „Sozialwissenschaften“ zum Forschungsdatenmanagement in der Soziologie, der Politikwissenschaft und der Kommunikationswissenschaft

(1) Die Archivierung von Forschungsdaten ist, sofern sie nicht InformantInnen oder Forschungsbeziehungen beschädigt, sinnvoll und notwendig. Archivierung oder Nicht-Archivierung sollten in Forschungsanträgen begründet werden. Im Falle der Archivierung können auch Finanzmittel in angemessenem Umfang beantragt und bereitgestellt werden.

(2) Wissenschaftlich nachnutzbare Daten sind nützlich für Offenheit und Transparenz vieler Studien. Insbesondere sind sie eine große Unterstützung, um empirische Ergebnisse intersubjektiv nachvollziehbar zu machen (etwa durch Reproduktion der Analysen).

(3) Wenn Forschungsdaten archiviert werden, sind Qualitätsstandards einzuhalten, wie sie in den FAIR-Prinzipien festgehalten sind (findable, accessible, interoperable, reusable).

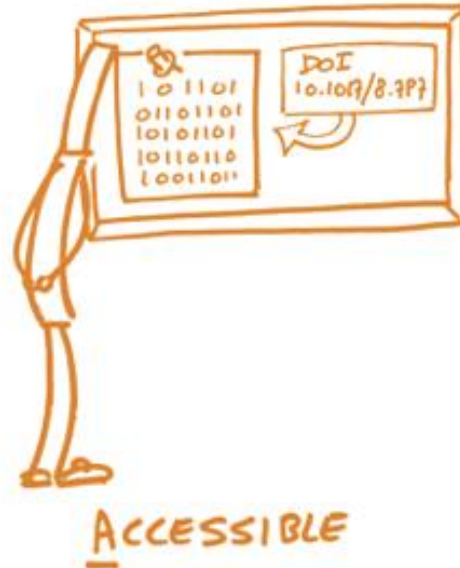
(4) Nicht immer werden archivierte Daten an andere Forschende weitergegeben werden können. Dies hat keinen Einfluss auf die Frage der Förderwürdigkeit eines Forschungsvorhabens.

„FAIR“ als **Leitprinzip** für das Forschungsdatenmanagement

# FAIR Data Principles

Wie sollte man optimalerweise Daten behandeln?  
Wie könnte man FAIR Data umsetzen?

## FAIR DATA PRINCIPLES



Bildquelle: <https://book.fosteropenscience.eu/>

# FINDABLE



- Nutzung einer **persistenten Kennung** (z.B. DOI) Beispiel: 10.1038/s41586-020-2649-2



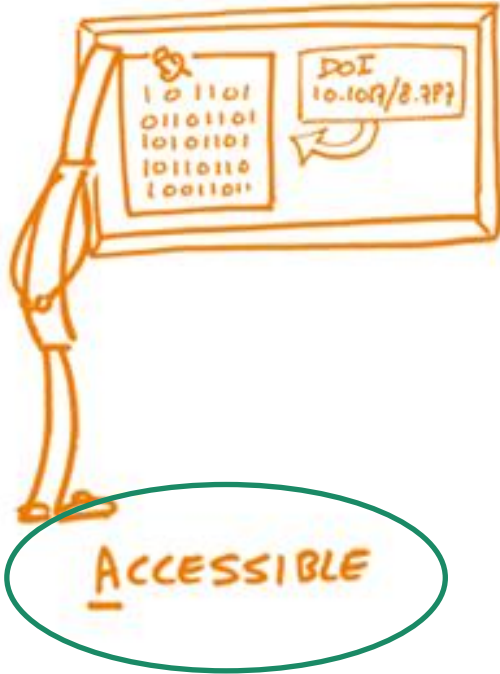
- Prüfung der **Anforderungen zur Auffindbarkeit** der wissenschaftlichen Verlage (z.B. in der „**Data Availability Policy**“)
- Erhöhung der Sichtbarkeit der Daten **durch Verlinkung** (z.B. Daten in eigenen Publikationen oder grauer Literatur, wie „*working papers*“ oder „*technical reports*“ referenzieren)
- **Zitation** der eigenen Daten:

Source:  
<https://www.fosteropenscience.eu/learning/assessing-the-fairness-of-data/>

**Romer, Christina D., and David H. Romer.** 2010. "Replication data for: The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks." *American Economic Association* [publisher], Inter-university Consortium for Political and Social Research [distributor]. <https://doi.org/10.3886/E112357V1>.



# ACCESSIBLE



Source:

<https://www.fosteropenscience.eu/learning/assessing-the-fairness-of-data/>

- Ergänzen Sie Ihre Datenpakete um **formale** und **technische** Metadaten!
- **Offene Protokolle** nutzen: Repositorium mit **HTTPS/FTP** wählen.



- Anbieter mit „Pay Walls“ **vermeiden** (zwecks Kostenfreiheit und Barrierefreiheit)
- Zur Einschränkung der **Zielgruppe** oder **Nutzungszwecke**: Repositorien mit Authentifizierung und Autorisierung nutzen.

Beispiele:

**Repositorien:** EBI NCBI, GenBank, RePEc, GESIS, usw.

**Versionierung/Entwicklung:** GitLab, GitHub

# INTEROPERABLE

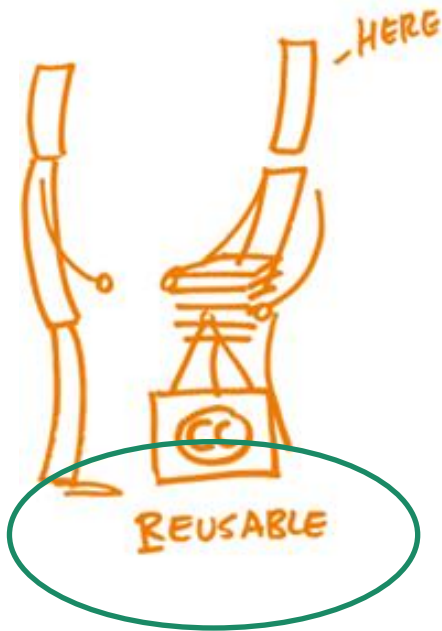


Source:  
<https://www.fosteropenscience.eu/learning/asses-sing-the-fairness-of-data/>

- ... Bedingung für **Datenausch** und **Nutzung durch Tools / Dritte**
- ... möglichst **offene Formate** und Datenformate der **Community** (wenn ja, welche?), wenn **neu definierte Formate**: Definition der Dokumentation beifügen ...
- Dokumentation aller Datenpakete (z.B. Variablen, Ressourcen, Parameter für Berechnungen, etc.) in einer standardisierten Art und Weise:

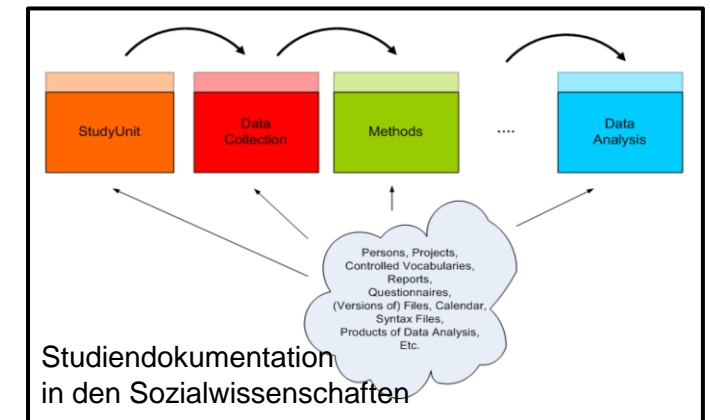
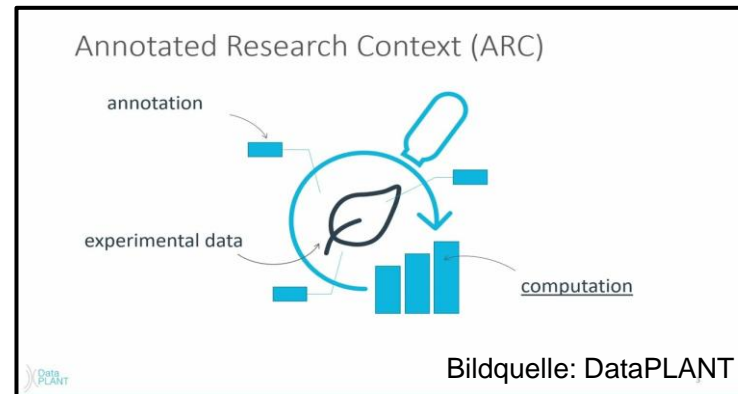
```
<subjects>
  <subject schemeURI="http://www.nlm.nih.gov/mesh/meshhome.html"
    subjectScheme="Mesh">Neoplasms</subject>
</subjects>
```
- ... auch hier Angabe von **Metadaten** (für die Interoperabilität), z.B. Softwareversion zum Öffnen, Definition zu Schnittstellen (APIs), usw.
- ... Daten in **verschiedenen Formaten anbieten** (um Interoperabilität zu erhöhen)

# REUSABLE



Source:  
<https://www.fosteropenscience.eu/learning/asses-sing-the-fairness-of-data/>

- ... sind gut **dokumentiert**,
- ... es gibt einen Plan für die **Veröffentlichung von Forschungsdaten?** (Wann?, Embargo? Patente? rechtliche Aspekte?)
- **Nachnutzbarkeit/Nachvollziehbarkeit:** Alle Quellen, Daten, Dokumente sollten verfügbar gemacht werden und die **(Zwischen-)Schritte** der Datengenerierung und Verarbeitung gut dokumentiert werden.

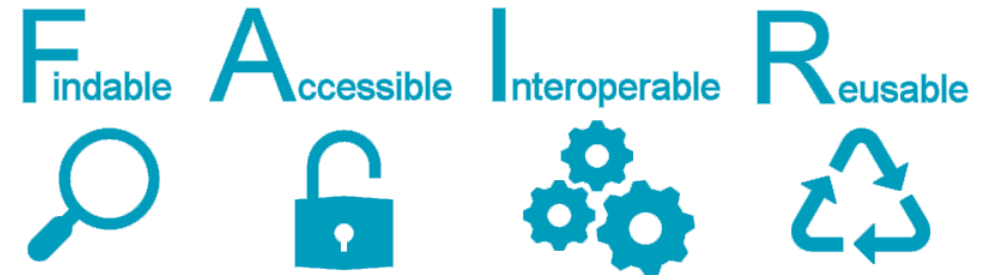


- ... haben klare Lizenz- und Herkunftsangaben (bestenfalls für die **breite Nachnutzung**)
- ... der gesamte „Forschungoutput“ ist gut untereinander verlinkt (z.B. Daten, Code, Working Papers, usw.)

# FAIR – Zusammenfassung

Generelles Ziel: FAIR soll die **effiziente Datennutzung** in der wissenschaftlichen Gemeinschaft fördern, indem es:

- die **Reproduzierbarkeit** stärkt,
- **Forschungserkenntnisse** beschleunigt,
- die **Transparenz** erhöht,
- und **Zusammenarbeit** erleichtert.



Es wird ein **Ökosystem** geschaffen, in dem Daten nutzbarer sind, und somit

- **bessere Forschungsergebnisse,**
- **schnellere wissenschaftliche Fortschritte,**
- **mehr Vertrauen der Öffentlichkeit in Forschung** ermöglicht werden sollen.

# Programm

- Verständnis von Forschungsdaten
- Empfehlungen und Anforderungen
- **Daten finden**
- Daten nachnutzen
- Daten organisieren, dokumentieren und sichern
- Abschluss

# Daten finden

## Datenquellen

Abhängig vom **Forschungsgebiet**  
möglicherweise unterschiedliche An

### Einige Beispiele:

1. Klimadaten
2. Model databases, e.g. C
3. European fact data (e.g. Flash Eurobarometer 315:  
Attitudes of European entrepreneurs towards eco-innovation)

<https://re3data.org>:

Registry of **certified** and **reliable** research data repositories

The collage consists of three overlapping screenshots. The top screenshot is from the 'Flash Eurobarometer 315' page, showing the survey title, publisher (Directorate-General for Communication), and a list of data volumes (A, AA, A', B, B', C). It also includes download links for 'F315\_TableC.zip' and 'fl315\_TableA.zip'. The bottom-left screenshot shows the 're3data.org' website with the 'CORE TRUST SEAL' logo and a search bar. The bottom-right screenshot shows the 're3data.org' website with the 'Registry of Research Data Repositories' logo and a search bar.

# Daten finden

<https://re3data.org>:

Verzeichnis zertifizierter und vertrauenswürdiger Forschungsdaten-Repositoryen

The screenshot displays the re3data.org website interface. At the top, the navigation bar includes 'Search', 'Browse', 'Suggest', 'Resources', and 'Contact'. The main search bar contains the term 'economics'. Below the search bar, a pagination control shows '1' as the active page, with 'Previous' and 'Next' links. A 'Sort by' dropdown menu is also visible. The left sidebar features a 'Filter' section with a 'Reset all' button and a 'Subjects' list. The list includes categories like 'Humanities and Social Sciences (30)', 'History (30)', 'Linguistics (9)', and 'Literary Studies (5)', each with sub-items and counts. The main content area displays 'Found 30 result(s)' and highlights 'The National Archives' as the first result. This entry includes a description of the archive, its subject areas (Humanities and Social Sciences, Humanities, History, etc.), repository type (institutional), provider type (dataProvider), and country (United Kingdom).

re3data.org

Search Browse Suggest Resources Contact

← → ↺ re3data

ceda

Filter

Reset all

Subjects

Humanities and Social Sciences (30)

Humanities (30)

Ancient Cultures (4)

Prehistory and World Archaeology (1)

Greek and Latin Philology (1)

Ancient History (2)

Egyptology and Ancient Near Eastern Studies (1)

History (30)

Medieval History (4)

Early Modern History (3)

Modern and Contemporary History (12)

History of Science (3)

Art History, Music, Theatre and Media Studies (5)

Art History (2)

Musicology (1)

Linguistics (9)

Individual Linguistics, Historical Linguistics (1)

Literary Studies (5)

Social and Cultural Anthropology, Non-European Cultures, Jewish Studies and Religious Studies (10)

Social and Cultural Anthropology and

economics

Search

Toggle short help

Sort by

← Previous 1 2 Next →

Found 30 result(s)

**The National Archives**

The National Archives Discovery

Subject(s)

Humanities and Social Sciences Humanities History Social and Behavioural Sciences Social Sciences

Political Science Economics Economic and Social History Jurisprudence Principles of Law and Jurisprudence

Repository type(s)

institutional

Provider type(s)

dataProvider

Country

United Kingdom

The National Archives is home to millions of historical documents, known as records, which were created and collected by UK central government departments and major courts of law. Data of the former National Digital Archive of Datasets (NDAD) collection, which was active from 1997 to 2010 and preserves and provides online access to archived digital datasets and documents from UK central government departments, is integrated. Access to records held by The National Archives and more than 2,500 other archives.



# Daten finden – Beispiele

## Fachspezifische Portale für offene Forschung

- [KonsortSWD](#) – Suche nach Sozial- und Wirtschafts Forschungsdatenzentren
- [The Economics Network](#) – Quellen für frei verfügbare
- [GenBank](#) – Die NIH-Datenbank für genetische Sequenzierung Sammlung aller öffentlich verfügbaren DNA-Sequenzen
- [GESIS](#): Leibniz-Institut für Sozialwissenschaften
- [IEDB](#): Immune Epitope Database & Tools

## Portale mit “Open Government Data”

- [Australia](#), [Canada](#), [EU](#), [Japan](#), [USA](#), [Netherlands](#)
- Daten des Statistischen Bundesamtes [Destatis](#)
- Daten der Europäischen Kommission [Eurostat](#)
- [Open Government Data Germany – offizielle Staat](#)
- [Open Government Data EU](#)

## Kommerzielle Faktendatenbanken (lizenziert)

- [Refinitiv](#) (Finanz- und Marktdaten)
- [OECD iLibrary](#) (Publikationen und Datensätze der OECD)
- Embase (medizinische Forschungsdatenbank, Elsevier)

The screenshot displays two web interfaces. The top interface is the GESIS Leibniz-Institut für Sozialwissenschaften website, featuring a search bar and navigation links. The bottom interface is the Destatis (Statistisches Bundesamt) website, showing a search result for 'EU-SILC - Persons in primary residence households: Germany, years, household type, poverty or social exclusion'. The table below shows data for 2024.

Year	Persons in primary residence households				
	Total	Persons at risk of poverty or exclusion	Persons at risk of poverty	Severely materially and socially deprived persons	Persons with no access to basic necessities
2024	1000	1000	1000	1000	1000
Households without children	48,692	10,689	8,566		2,812
Persons living alone	17,160	6,145	5,237		1,591
Two adults without children	25,827	3,894	2,938		1,009
Three or more adults without children	5,705	650	391		212
Households with children	35,501	6,912	4,495		2,236
Lone parents	4,205	1,806	1,103		571
Two adults with children	27,204	4,551	3,082		1,446
Three or more adults with children	4,092	555	309		219
Total	84,193	17,601	13,061		5,048

→ Fragen Sie Ihre Fachreferentin/Ihren Fachreferenten an Ihrer Einrichtung nach Verfügbarkeit ...



# Daten finden – Bibliotheksbasierter Zugang

## Database Information Systems (DBIS) – Ein Zugang zu wissenschaftlichen Datenbanken

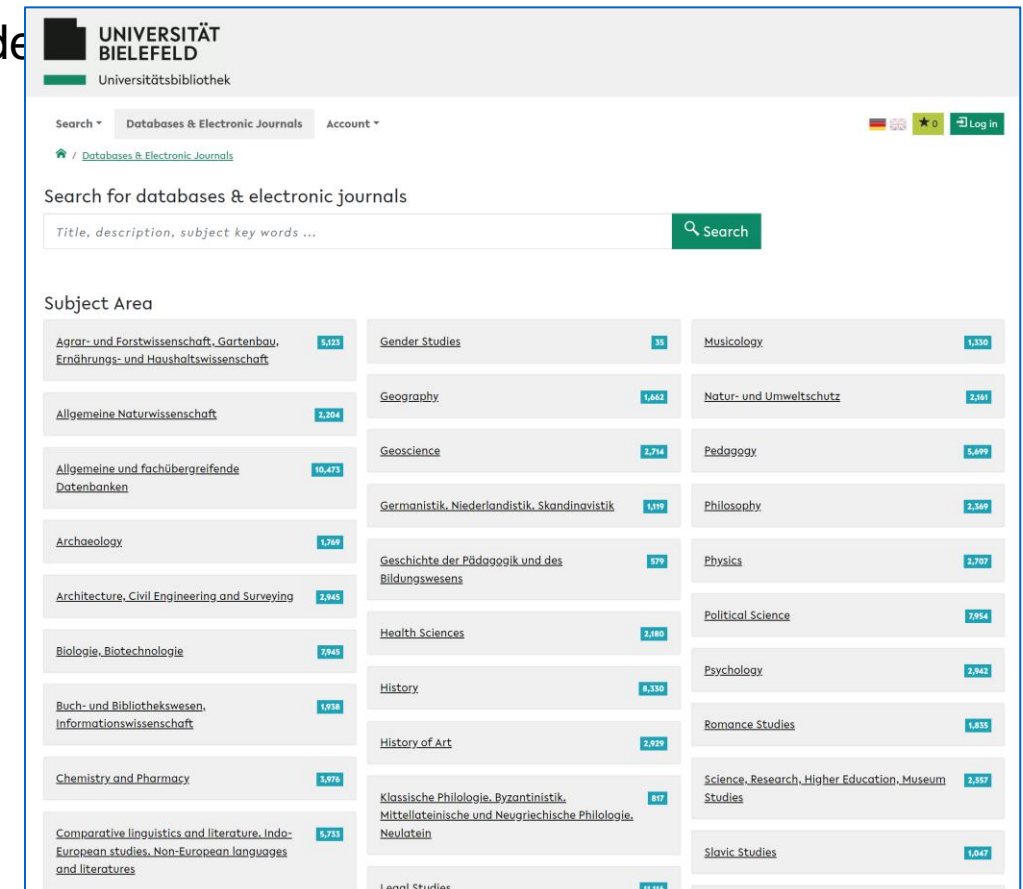
Die **DBIS** (bereitgestellt von Bibliotheken) sind ein hervorragender Zugang zu wissenschaftlichen Datenbanken:

Suche nach Fachgebiet oder gezielt nach Datenbanken:

<https://ub-bielefeld.digibib.net/eres/subject-area/>

### Beispiele der dort gelisteten Informationsressourcen:

- **Statista:** Statistiken zu Wirtschaft, Gesellschaft & Medien.  
<https://www.statista.com/>
- **OECD iLibrary:** Eine der größten Online-Bibliotheken für Daten und Analysen zu Wirtschaft, Gesellschaft und Umwelt.  
<https://www.oecd-ilibrary.org/>
- **ViFa Recht (Virtual Library Law):** Ein Instrument für die rechtliche Online-Recherche und Zugang zu wissenschaftlichen Informationen zu Recht und Rechtsgeschichte.  
<https://vifa-recht.de/>



# Daten finden – Allgemeine Suchmöglichkeiten

## (Disziplin-)übergreifende Portale und Metasuchmaschinen

- [DataCite search portal](#) ist eine internationale Suchmaschine für Forschungsergebnisse, mit der nach registrierten Datensätzen gesucht werden kann.
- [European Union Open Data Portal](#) – Open-Data-Portal der Europäischen Union
- [Research Data Australia](#) – Australiens Open Research Data Portal
- [Google Dataset Search](#) – (Kommerzielle) Suchmöglichkeit für Forschungsdaten
- [gesisDataSearch](#) – Suche nach Daten zu Sozial- und Wirtschaftsforschung in Datenrepositorien und Metadaten-Diensten.
- [OpenAIRE](#) – eine internationale Suchmaschine für Forschungsergebnisse
- [Zanran](#) – Suchmaschine für Daten, mit der man nach Grafiken und Tabellen in Dokumenten suchen kann
- [Mendeley Data](#) – (kommerziell & multidisziplinär, Elsevier): Suche in einer Sammlung von Datenrepositorien
- [Data Portals](#) ist eine umfassende *Liste* von Open-Data-Portalen aus aller Welt.
- [Open Data Impact Map](#) eine öffentliche Datenbank von Organisationen weltweit, die Open Government Data nutzen

# Daten finden – In Zeits

## Beispiel aus der Mikrobiologie

- Verweis über eine lokale ID (Zugangsnummer) eines Repositoriums/Datenbank
- Datenverfügbarkeit im Artikel prüfen, z.B. im Unterkapitel zu „**data availability**“

# Whole-Genome Sequence of *Paenibacillus marchantiae* Isolated from the Liverwort *Marchantia polymorpha* subsp. *ruderalis* Ecotype BoGa

were predicted with Prokka (v1.14.5) (19). Relevant statistics for the assembly are listed in Table 1. Metabolic pathways were predicted with the KEGG Automatic Server (KAAS) (20), GO terms were calculated with eggNOG-mapper (v2) (21), secondary metabolite biosynthesis gene clusters were identified with antiSMASH

Phylogenetic relatedness to other strains was determined using the assembled genome (Strain) Genome Server (TYGS) (23) (Fig. 1). The closest species were *Paenibacillus* DSM 19942 (GenBank accession number GCA\_013359905.1) and *Paenibacillus* sp.

(GenBank accession number [GCA\\_001280595.1](#)), which were both isolated from

The KAAS analysis indicates that *Paenibacillus marchantiae* is unable to produce and sucrose but of can take up and degrade them. The gene cluster of pa identified and comprises all genes for bacillibactin synthesis. These siderophore ble for the extracellular conversion of  $Fe^{2+}$  and  $Fe^{3+}$  and thus enable the provi plants (9). The metabolic capabilities indicate a potential beneficial partner *Paenibacillus marchantiae* and BoGa.

**Data availability.** The genome assembly of *Paenibacillus marchantiae* has been deposited under the NCBI GenBank accession number [CP118270.1](#). The SRA, BioProject, and BioSample accession numbers are [SRP421713](#), [PRJNA933122](#), and [SAMN33224160](#), respectively.

assembly (v1.2.3) (<https://github.com/Concinn/Berkeley>) was run for overlap trimming. The assembly resulted in one circular contig (6,983,959 bp; GC content of 46.02%) (Table 1). Benchmarking Universal Single-Copy Orthologs (BUSCO) (v5.4.3; database "bacillales\_odb10") (17) and CheckM (v1.2.2) (18) were applied to check assembly quality resulting in 0.097% contamination and 99.84% assembly completeness. A total of 6,195 protein-coding genes

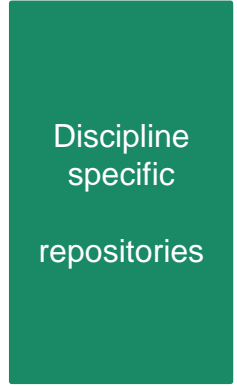
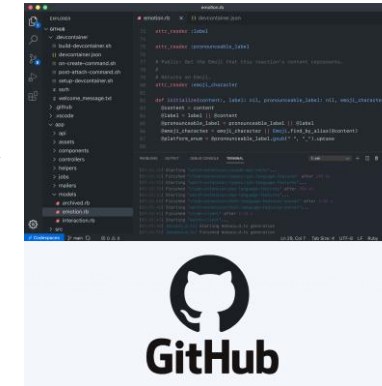
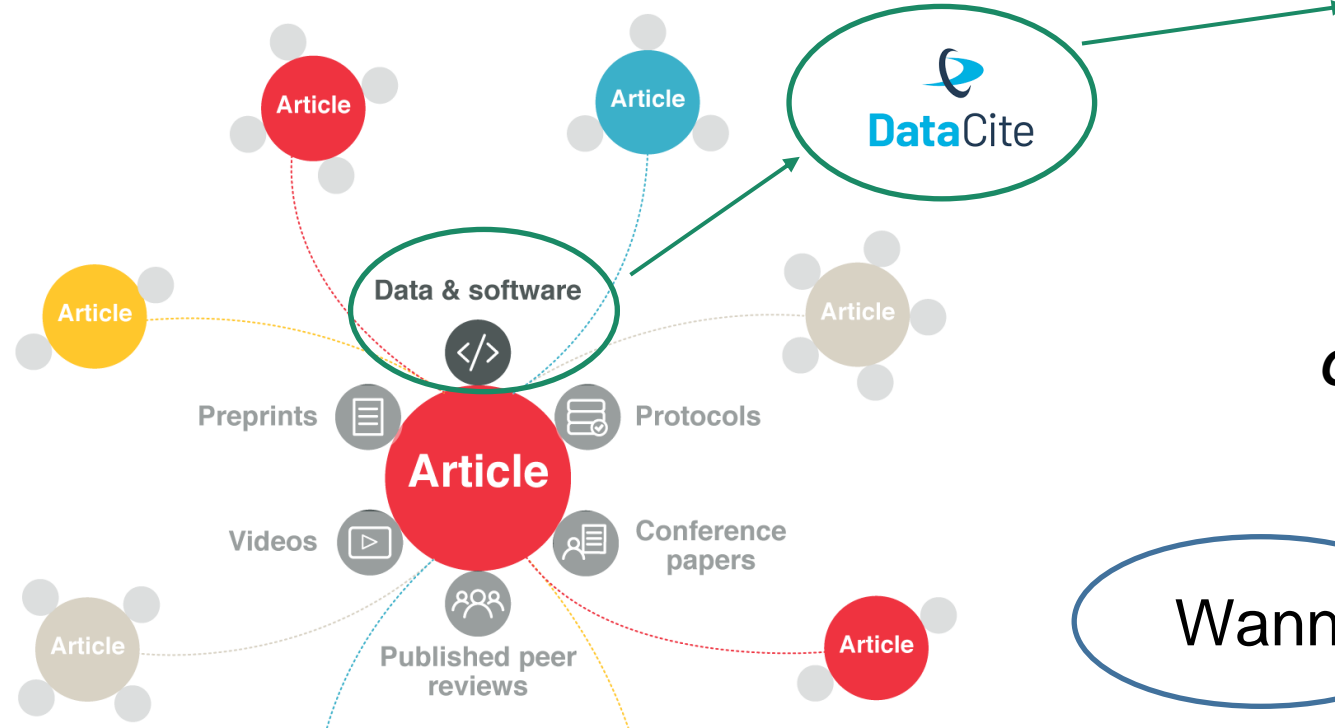
Address correspondence to Anja Meierhenrich  
anja.meierhenrich@uni-bielefeld.de.  
The authors declare no conflict of interest.  
**Received** 10 May 2023  
**Accepted** 9 June 2023  
**Published** 22 June 2023

# Eigene Daten auffindbar machen

Wo?

Persistente Referenzierung durch DOIs

z.B. <https://doi.org/10.4119/unibi/123456789>



Optionen für Veröffentlichung von Daten

Wann?

4.2 Project data will be published:

- ☒ Immediately upon collection
- ☐ Upon publication of findings.
- ☐ At the end of project.
- ☐ After embargo period.
- ☐ Upon request.
- ☒ Upon IP issues have been addressed.


Source: [crossref.org](https://crossref.org)


# Daten teilen – PUB an der Universität Bielefeld

Publikationen an der  
Universität Bielefeld

PUB

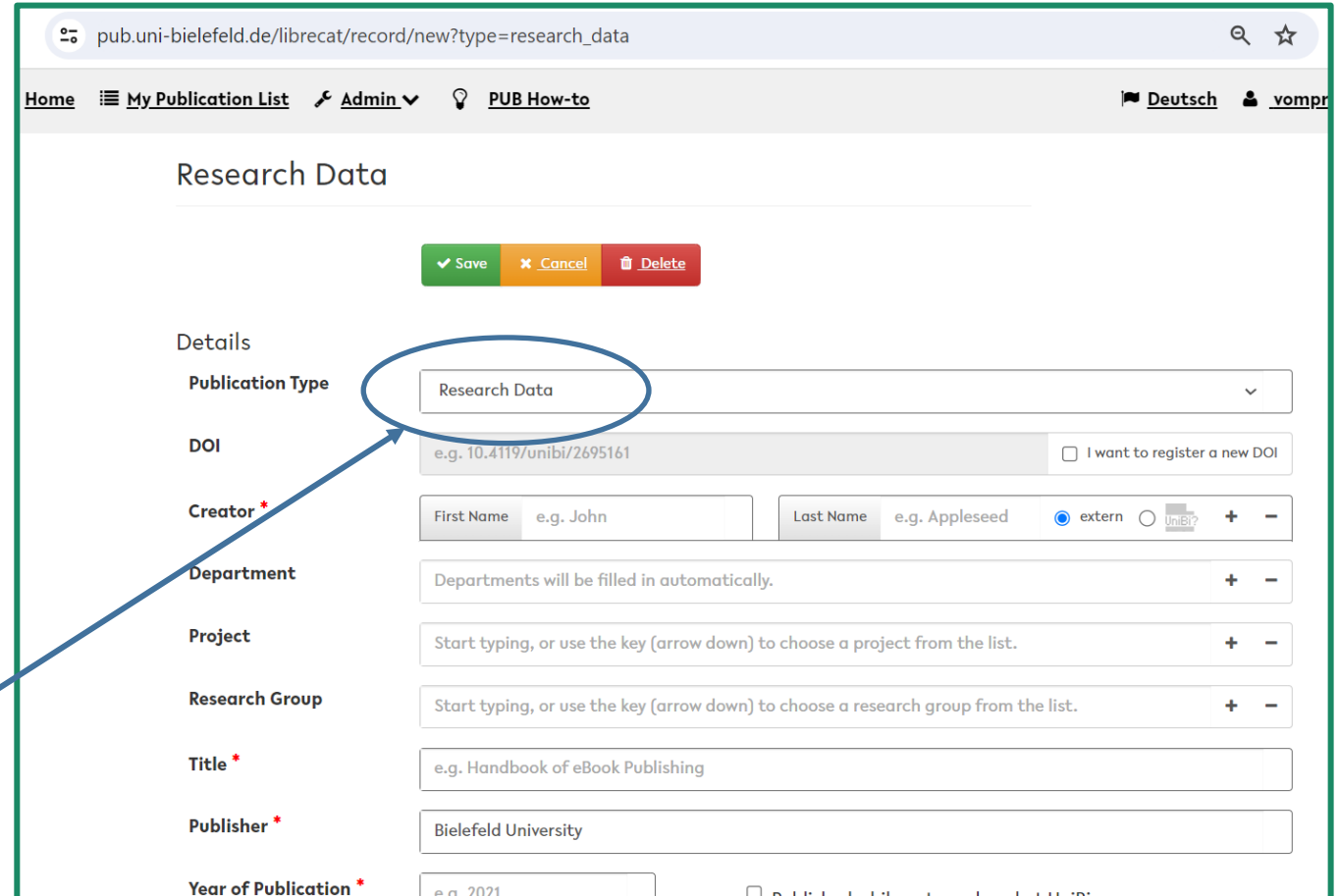
<https://pub.uni-bielefeld.de>

✓ **ORCID iD present!** 

 [Connect your ORCID iD](#)

 [ORCID iDs in PUB](#)

 [Add new Publication](#)



pub.uni-bielefeld.de/librecat/record/new?type=research\_data

Home My Publication List Admin PUB How-to Deutsch vompr

### Research Data

✓ Save ✕ Cancel 🗑 Delete

Details

Publication Type Research Data

DOI e.g. 10.4119/unibi/2695161 ☐ I want to register a new DOI

Creator \* First Name e.g. John Last Name e.g. Appleseed ☒ extern ☐ Unibi? + -

Department Departments will be filled in automatically. + -

Project Start typing, or use the key (arrow down) to choose a project from the list. + -

Research Group Start typing, or use the key (arrow down) to choose a research group from the list. + -

Title \* e.g. Handbook of eBook Publishing

Publisher \* Bielefeld University

Year of Publication \* e.g. 2021 ☐ Published online to open access at Unibi?



# Weitere P

## Publikationsmöglich

- Als **Anhang** (supp) (Publikationsort: V Beispiele: *Integrat* oder *American Ec*
- Als **eigenständig** Beispiele: *PANGE* oder *SowiDataNe*
- **Data Journal:** Journal für "Daten in Forsch Wiederverwendbarkeit zu fördern. Beispiele: *Scientific Data* (Nature), *Biodiver*
- Sonstige generische Repositorien: *Zenodo*, *Open Science Framework*, *Figshare*

BROWSE PUBLISH ABOUT SEARCH

# PLOS BIOLOGY

advanced search

Introduction  
Minimal Data Set Definition  
Acceptable Data Sharing Methods  
Acceptable Data Access Restrictions  
Unacceptable Data Access Restrictions  
FAQs  
PLOS Data Advisory Board

## ASSOCIATED CONTENT

### Data Availability Statement

The data underlying this study are openly available in [Repository Name] at [Persistent Link to data in Repository, e.g., DOI, Accession Number].

### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/XXXX>.  
Experimental procedures and characterization (PDF)

☐ Data and code have been deposited in a trusted repository.

Please complete ONE option below:

[AEA Data and Code Repository](#)  
Project Number:   
☐ I have followed the AEA Repository [deposit instructions](#).

Other [trusted repository](#)  
Deposit DOI: <https://doi.org/>

under that policy.

PLOS believe  
> Validat

open-access journal  
ty  
urnal  
1314-2828 (online)

# Was ist zu beachten, wenn ich Daten teile?

- Alle Ressourcen für die **Nachnutzung verfügbar** und **zugänglich** machen!  
→ Vertrauenswürdige Repositorien nutzen | Persistente Identifikatoren verwenden | Offener oder eingeschränkter Zugang | Ethische und rechtliche Vorgaben einhalten
- **Daten** und **Datenverarbeitung** sauber dokumentieren!  
→ Metadaten | Verarbeitungsschritte ebenfalls beschreiben
- **Nachvollziehbarkeit** sichern: Setup, Code, Skripte, Trainingsdaten, Parameter, Outputs dokumentieren  
→ Standardisierte und offene Formate | Best Practices der Community
- Publikationen mit Daten **verlinken** – z. B. per DOI  
→ Zitationsformate | Daten sichtbar machen



# Programm

- Verständnis von Forschungsdaten
- Empfehlungen und Anforderungen
- Daten finden
- **Daten nachnutzen**
- Daten organisieren, dokumentieren und sichern
- Abschluss



# Daten nachnutzen - Lizenzen

Wenn Sie externe Daten in Ihre eigene Forschung einbauen möchten:

→ Achten Sie auf die Lizenz, unter der die Daten bereitgestellt werden!

Eine Lizenz ist eine erteilte **Erlaubnis/Recht**:  
Sie legt fest, was Sie mit den Daten **tun dürfen**  
(z. B. teilen, verändern, wiederverwenden)  
und auch **unter welchen Bedingungen**.



LICENSES	TERMS
	<b>Attribution</b> Others can copy, distribute, display, perform and remix your work if they credit your name as requested by you
	<b>No Derivative Works</b> Others can only copy, distribute, display or perform verbatim copies of your work
	<b>Share Alike</b> Others can distribute your work only under a license identical to the one you have chosen for your work
	<b>Non-Commercial</b> Others can copy, distribute, display, perform or remix your work but for non-commercial purposes only.

# Daten nachnutzen – richtig zitieren

Wenn Sie externe Daten in Ihre eigene Forschung einbinden oder nutzen möchten:

→ **Genutzte Quellen, Daten und Ressourcen zitieren!**

## Elemente einer Zitation von Daten/Statistiken:

- Autor(en)/Urheber
- Titel: “Daten aus ....”
- Publikationsdatum: Das Veröffentlichungs- bzw. Freigabedatum der Statistik/des Datensatzes (nicht das Erhebungs- oder Abdeckungsdatum)
- Herausgeber/Verlag: Datenzentrum/Repositorium
- **Sonstige Identifikatoren (inkl. Edition oder Version)**
- **Verfügbarkeit und Zugriff: URL / DOI oder andere Ortsangabe**

# Daten nachnutzen – Beispiele von Zitationen

## Zitationen (externer) Daten / Datenbanken / Ressourcen:

Bureau of Transportation Statistics, Bureau of Transportation Statistics (2021-07-25). 15 Minute or More Delay: 15 Minute or More Delay | , 01/01/2010 - 02/28/2021. Data Planet™ Statistical Datasets: A SAGE Publishing Resource. **(Dataset). Dataset-ID: 007-001-004**

Food and Agriculture Organization of the United Nations. (2021). Production of Bananas: top 10 producers. **FAOSTAT**. Rome, Italy: FAO. **Retrieved August 30, 2021** from <http://www.fao.org/faostat/en/#home>

Vobejda, C. (2025). **Data: Stability of heart rate at maximal lactate steady state in recreational sportsmen**. Bielefeld University. **DOI: <https://doi.org/10.4119/unibi/3000793>**

# Daten nachnutzen – Zitationsstile

Am besten Empfehlungen der **Datenportale** oder **Repositorien** beachten ...



## Citation requirements

The Core Scientific Team of the ESS requests that references to ESS data and the Data Documentation Reports should use the form of words listed below.

To ensure that such source citations must appear in the

### Citation of data

### Citation of documentation

### Citation requirements for

## Citation of data

- ESS Round 11: European Social Survey European Research Infrastructure (ESS ERIC) (2024) ESS11 - integrated file, edition 2.0 [Data set]. Sikt - Norwegian Agency for Shared Services in Education and Research. <https://doi.org/10.21338/ess11-2023>

- ESS Round 10: European Social Survey (2023): ESS-10 2020 Documentation Report. Edition 3.0. Bergen, European Social Survey Data Archive, Sikt - Norwegian Agency for Shared Services in Education and Research, Norway for ESS ERIC. [doi:10.21338/NSD-ESS10-2020](https://doi.org/10.21338/NSD-ESS10-2020).

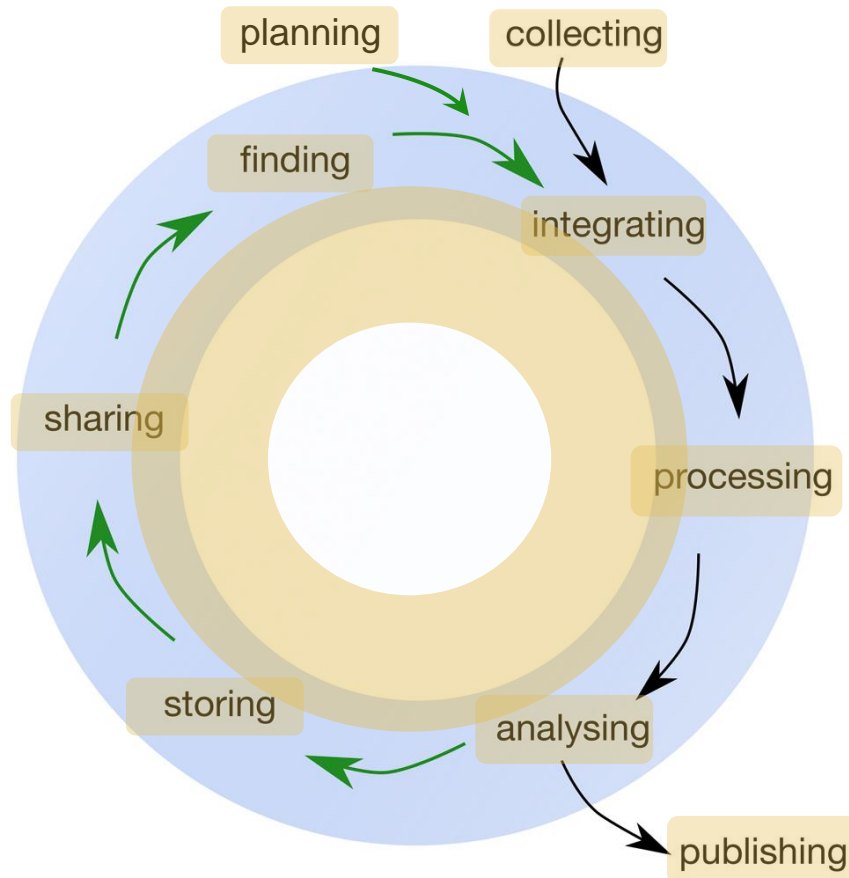
## Citation of documentation

- European Social Survey European Research Infrastructure (ESS ERIC) (2024) ESS round 11 - 2023. Social inequalities in health, Gender in contemporary Europe. Sikt - Norwegian Agency for Shared Services in Education and Research. [doi:10.21338/ess11-2023](https://doi.org/10.21338/ess11-2023).
- ESS Round 10: European Social Survey (2023): ESS-10 2020 Documentation Report. Edition 3.0. Bergen, European Social Survey Data Archive, Sikt - Norwegian Agency for Shared Services in Education and Research, Norway for ESS ERIC. [doi:10.21338/NSD-ESS10-2020](https://doi.org/10.21338/NSD-ESS10-2020).

# Programm

- Verständnis von Forschungsdaten
- Empfehlungen und Anforderungen
- Daten finden
- Daten nachnutzen
- **Daten organisieren, dokumentieren und sichern**
- Abschluss

# Datenorganisation

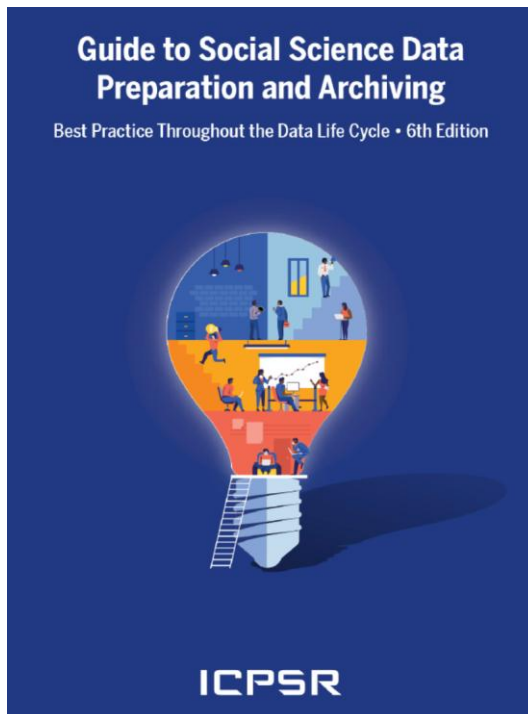


Aspekte der Datenorganisation:

Am besten im Rahmen eines Data Management Plans (DMP) in Vorfeld klären/festlegen:

- Daten – und Dateistruktur
- Ordnerstruktur / Gruppierung
- Benennung von Dateien
- Benennung Variablen, etc.
- Versionierung

# Datenorganisation



## Namenskonvention

Minimalanforderung:

- Sollten korrekte Informationen vermitteln,
- Benennung sollte eindeutig sein (Codebuch erstellen)
- Länge kompatibel mit technischen Anforderungen.



Variablennamen:  
ausgelegt auf eine  
langfristige Nutzung  
durch ein **breites Publikum**

# Datenorganisation

Systeme für die Konstruktion von Variablennamen.

Einige Beispiele:

1. **Zahlenkombinationen:** Variable 1 ... n: **V0001, V0002, ...**
2. **Fragenummern:** Variable 1 ... n: **Q1, Q2a, Q2b, ...**
3. **Mnemonicische Namen:** z.B. **B\_EKOM** (*Einkommen des Befragten*)
4. **Präfix, Stamm und Suffix:** z.B. Stamm: **ED** für Bildung  
Höchster Bildungsstand der Mutter: **MOED**, Vater: **FAED**  
**Suffix** für ergänzende Informationen

Beispiele: **A)** Welle: **MOED\_W3** oder **MOED3**  
**B)** Landesspezifität: **MOED\_DE**



# Datenorganisation



Bildquelle: <https://www.forschungsdaten.uni-bonn.de/>

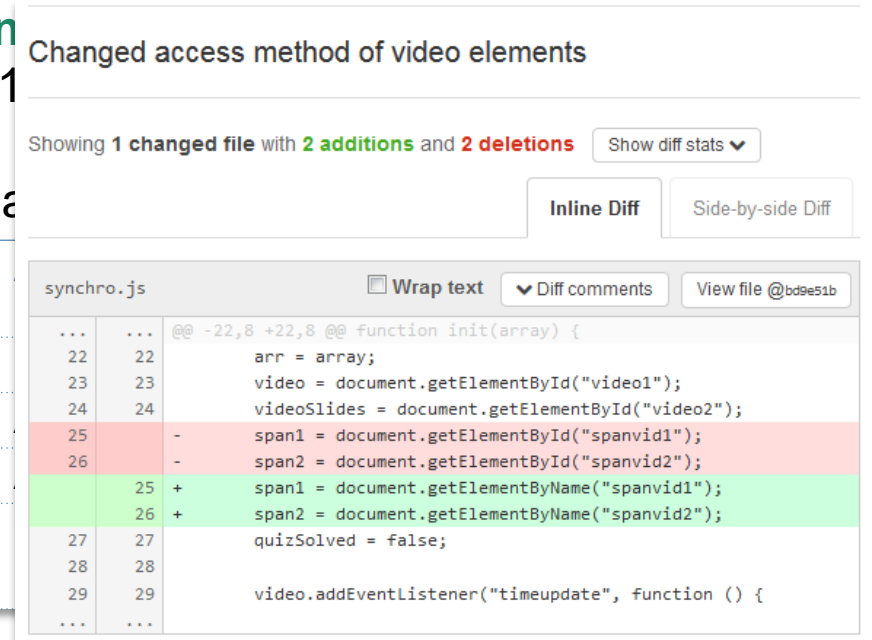
## Anwendung von Namenskonventionen auf Dateien und Ordner

1. Verschiedene **Ordnung**
  - Teilprojekte (TN1)
  - Zeit: YYYY-MM
  - Inhaltsbasiert: Da

### Historie / Version

2. Dateiverzeichnis (größere Beispielsätze)

1.0. (10.3.2021)
1.1. (17.3.2021)
1.2. (19.3.2021)
2.0. (19.6.2017)



Weitere Informationen:

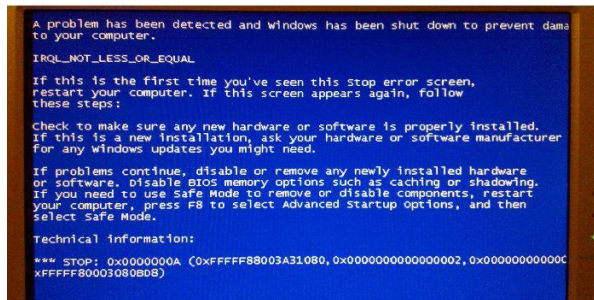
<https://www.forschungsdaten-bildung.de/dateien-benennen#Dateiorganisation-Ordnerstruktur>

Data Crunch handout "DIY: File naming": <https://zenodo.org/records/10275946>

# Datensicherung und Backup

## Warum Backups?

- Datenverlust kann **jeden** treffen!
- Ihre Daten und Dokumente sind **das Gold** Ihrer Forschung!



## Bestandteile von Backups

- Backup-Strategie
- Dedizierte Backup-Hardware (Speicherorte)
- Backup-Software

# Datensicherung und Backup

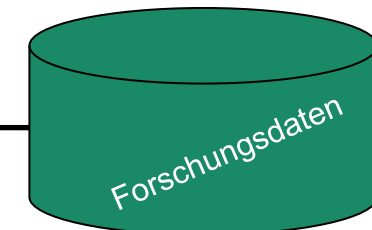
Ersetzbarkeit / Wiederherstellbarkeit von Daten

Leicht	Mittel/Schwer	Unmöglich
Betriebssystem, kostenfreie Anwendungen	Anwendungseinstellungen / Betriebssystemeinstellungen	Selbst gemachte Fotos, Videos
Fotos, die von anderen gemacht wurden	Adressdatenbanken	(Abschluss-)Arbeiten, Steuerunterlagen
gekaufte digitale Inhalte wie Musik, eBooks, Filme; privat genutzte Software	Verlauf Kommunikation (Messengerdienste); Email-Kommunikation	generell: Erzeugnisse eigener Arbeit

Lizenzen für beruflich genutzte Software oder alles was frei im Internet zur Verfügung steht (Publikationen, Daten)

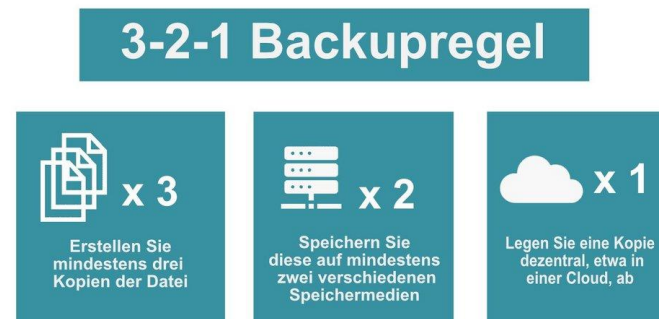
**Vor allem kritisch:**

- mühselig erhobene Daten, Experimentaldaten, ...
- Weitere Hilfsdokumente (Dokumentationen, Skripte), ...
- Drafts von Publikationen, usw.



# Datensicherung und Backup

- Backup machen! Backup jetzt erstellen!
- Anfangen – z. B. mit einer Kopie wichtiger Ordner und Dateien auf einem USB-Stick / externen Laufwerk
- Die 3-2-1-Regel: 3 Kopien auf 2 Datenträgern, davon 1 außer Haus



**Beispiel:** Eine Kopie auf der lokalen Platte, zwei auf zwei USB-Datenträgern, eine davon bei Verwandten, Eltern, Arbeitsplatz zusätzlich in der Cloud

- Kontrolle des Backups – Restore-Fall testen!

# Programm

- Verständnis von Forschungsdaten
- Empfehlungen und Anforderungen
- Daten finden
- Daten nachnutzen
- Daten organisieren, dokumentieren und sichern
- **Abschluss**

# Was wir Ihnen mitgeben möchten



Denken Sie frühzeitig an **das Datenmanagement!**  
Nutzen Sie die **Beratungsangebote** der Universität Bielefeld

Ein gutes Datenmanagement ist ein **Qualitätssiegel**  
Ihrer Forschung. Es hilft dabei, die **offene Wissenschaft** zu  
leben (Open Science)

Es legt auch die Weichen für:

- Ermöglichung der Weitergabe und **Wiederverwendung** von Daten
- Gute wissenschaftliche Praxis: **Überprüfung** von wissenschaftlichen Ergebnissen möglich!
- Effizienzsteigerung: **Vermeidung von doppelter Arbeit**, (z.B. schnelles Onboarding neuer KollegInnen)



*Vielen Dank!*



Markus Rump (BITS & UB), Johanna Vompras (UB), Anne Gärtner (UB)

## **Kompetenzzentrum Forschungsdaten**

Tel.: 0521-106-3967

E-Mail: [data@uni-bielefeld.de](mailto:data@uni-bielefeld.de)

<https://data.uni-bielefeld.de>

*Sprechen Sie uns an!*