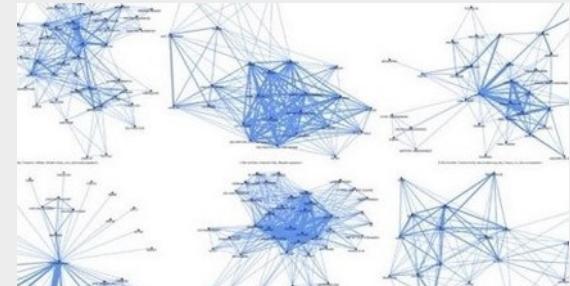
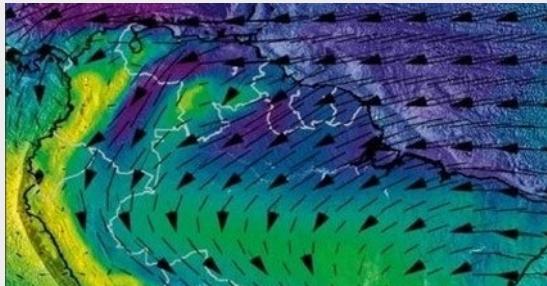
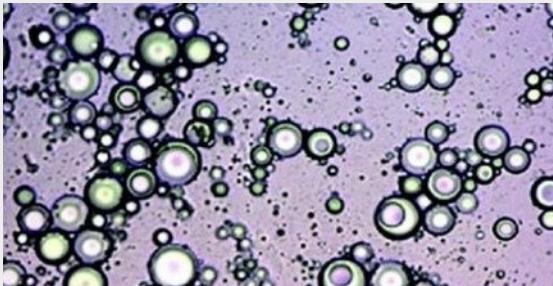


# Forschungsdaten richtig managen

FDM-Team von UB und ZIM  
Dr. Janine Straka

[www.uni-potsdam.de/forschungsdaten](http://www.uni-potsdam.de/forschungsdaten)



# Programm

## Einführung ins (praktische) Forschungsdatenmanagement

- 1) Digitale Forschungsdaten
  - Datenverlust, Datenskandale, Begriffe
- 2) Forschungsdaten richtig ablegen
  - Finden, speichern, ordnen, strukturieren
- 3) Forschungsdaten verständlich verfügbar machen
  - Dokumentieren, publizieren, archivieren

# Vorstellung: Dr. Janine Straka

- Studium: Biowissenschaften (B. Sc.) & Bioinformatik (M. Sc.)
- Doktor in Theoretische Biophysik
- 2017-2021 FDM-Projekte an der Fachhochschule Potsdam
- Seit 2022 Koordinatorin FDM & Fachreferentin Universitätsbibliothek Potsdam

# Session 1: Digitale Forschungsdaten

# 1.1. Datenverlust

# Daten

**Die weltweite Datenmenge wird von rund 33 Zettabyte (ZB) im Jahr 2018 auf 175 ZB im Jahr 2025 steigen – also jedes Jahr um circa 27 Prozent.**

- Zettabyte = Milliarde Terabyte
- ZB = ca. rund zwei Billionen 90-Minuten-Filme

**Die größten jährlichen Wachstumsraten für den Zeitraum von 2018 bis 2025 sagt die International Data Corporation mit 36 Prozent für den Gesundheitsbereich voraus. (E-Health)**

- Wie mit den Daten effektiv bewirtschaften? Künftig werden immer mehr Daten in zentralen Speichern (Clouds) liegen. Aktuell noch die meisten Daten dezentral gespeichert.

# Übung

Wo sind Ihre privaten und dienstlichen Daten gespeichert?

Sind Sie sicher? Gibt es Backups?

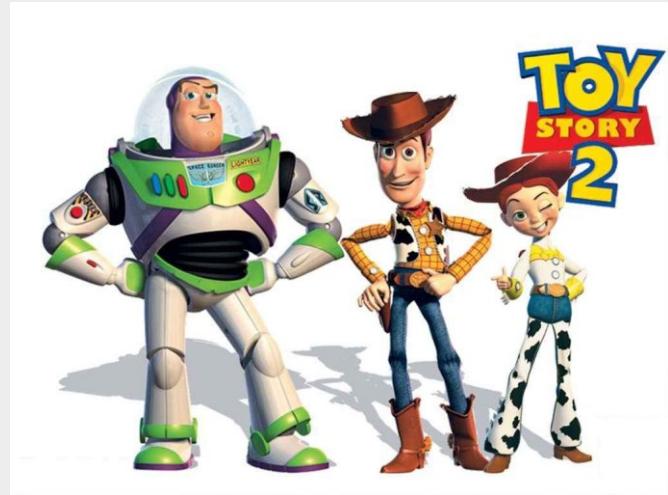
Sind diese gut beschriftet? Automatische Benennung  
z. B. bei Fotos?

Was könnten Sie verbessern?

(3 min. Brainstormen – Reflektion am Ende des Tages)

# Datenverlust bei Pixar

- versehentlich: Kommando *rm* \*
- das automatische Backup funktionierte nicht, 2 Monate Daten weg
- eine halb-private Arbeitskopie der technischen Direktorin
- Baby zuhause, Film auf ihren Computer transferiert für Homeoffice
- Selbst bei State of the Art Backups können Datenverluste auftreten



Pixy.org CC BY-NC-ND 4.0

# Bizarre und kuriose Fälle von Datenverlust

- Ein Kunde präsentiert eine Festplatte in einem tropfenden Plastikbeutel. Er erzählt, er habe im Internet gelesen, dass eine defekte Festplatte repariert werden könne, indem man sie ins Gefrierfach legt.
- Ein anderer Kunde hat beim Aufräumen seines Systems genau die Dateiordner gelöscht, die er eigentlich behalten wollte. Am Ende der Putzaktion leerte er noch fein säuberlich den „Papierkorb“ und startete dann die Defragmentierung der Festplatte.
- Der Mitarbeiter einer Medizintechnikfirma hatte in mehrtägiger Arbeit 1200 Rechnungseinträge in das System eingegeben. Als plötzlich die Trafostation neben dem Gebäude von einem Blitz getroffen wurde, kam es zum Stromausfall und kompletten Datenverlust.
- Während des Baus eines großen Bürogebäudes wurde das Notebook, auf dem die Konstruktionspläne gespeichert waren, von einem herabfallenden Stahlträger getroffen.

# Häufigste Ursachen für Datenverlust

- Technische Defekte
- Menschliche Fehler
- Malware, Virenbefall und Hackerangriffe
- Datendiebstahl
- Externe Einflüsse

# 3-2-1 Regel

- Zum sicheren Speichern von Daten

3 Orte

2 vers. Speichermedien

1 externer Ort

## 1.2. Datenskandale und wissenschaftliches Fehlverhalten

# Datenskandal bei Facebook

- „Cambridge Analytica steht unter dem Verdacht, Millionen von Wählern in den USA per Facebook gezielt mit unerlaubter Wahlwerbung für Donald Trump beeinflusst zu haben.“
- „Facebook beschuldigt den Wissenschaftler, Daten, die mithilfe einer Umfrage in dem Sozialen Netzwerk zu Forschungszwecken gesammelt wurden, unrechtmäßig an die Beratungsfirma Cambridge Analytica weitergegeben zu haben.“

# Beispiel für wissenschaftliches Fehlverhalten

„Eine Postdoktorandin der Universität Uppsala publizierte im Juni 2016 in *Science* gemeinsam mit dem Leiter ihrer Arbeitsgruppe eine weltweit beachtete Studie über den Zusammenhang von Mikroplastik im Ozean und Entwicklung von Fischlarven. Besucher der angeblich auf Gotland durchgeführten Experimente **bezweifelten** bereits unmittelbar nach der Publikation, dass die behaupteten **Versuchsansätze dort hatten durchgeführt werden können**. Kurz nach dieser Kritik wurde von den beiden Autoren der Studie **behauptet, die Rohdaten seien infolge des Diebstahl eines Laptops komplett verloren gegangen**. Eine Untersuchung durch die Universität Uppsala führte im Mai 2017 dazu, dass *Science* die Publikation für ungültig erklärte. Im Dezember 2017 erklärte die Universität Uppsala die Daten der Studie für erfunden („fabricated research results“).“

# Vertrauenskrise in der Wissenschaft

- Skandale
- Ob und wie Vertrauen der Bevölkerung in der Wissenschaft beeinträchtigt wird

# Formen von Betrug und Fälschung

- Erfundene oder gefälschte Daten:
  - komplett erfundene Daten
  - Vervielfachung der erhobenen Daten, um größere Stichproben vorzutäuschen
  - Manipulation von Abbildungen
  - Verheimlichen unliebsamer Daten
  - bei Regressionsverfahren das Entfernen von Ausreißern
- Falsche Aussagen oder Schlussfolgerungen

# Ursachen und Motivation

- Hoffen auf Ruhm, Ehre
- Publikationszwang („Publish or perish“)
- Bedarf an Fördermittel

# Vermeidung

DFG Sicherung guter wissenschaftlicher Praxis (2013)

DFG Leitlinien zum Umgang mit Forschungsdaten (2015)

DFG Leitlinien zur Sicherung guter wissenschaftlichen Praxis – Kodex  
(2019)

# Leitlinie 13: Herstellung von öffentlichem Zugang zu Forschungsergebnissen

„[...] 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. Selbst programmierte Software** wird unter Angabe des Quellcodes öffentlich zugänglich gemacht.“

“Erläuterungen: Aus Gründen der Nachvollziehbarkeit, **Anschlussfähigkeit** der Forschung und **Nachnutzbarkeit** hinterlegen Wissenschaftlerinnen und Wissenschaftler, wann immer möglich, die der Publikation zugrunde liegenden Forschungsdaten und zentralen Materialien – den **FAIR-Prinzipien** („**Findable, Accessible, Interoperable, Re-Usable**“) folgend – zugänglich in anerkannten Archiven und Repositorien.“

# 1.3. Definition Forschungsdaten und Forschungsdatenmanagement

# Forschungsdaten an der Universität Potsdam

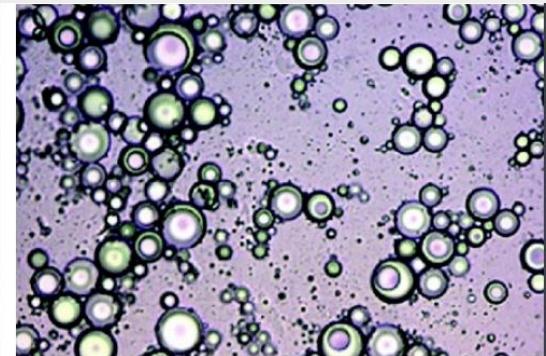
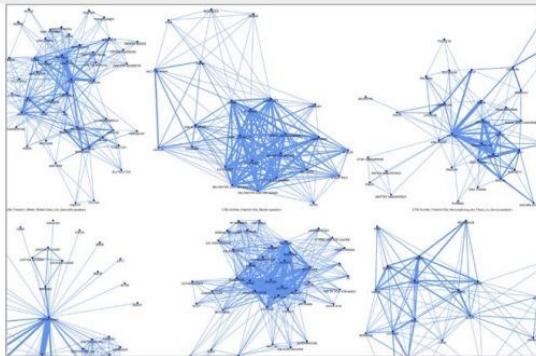
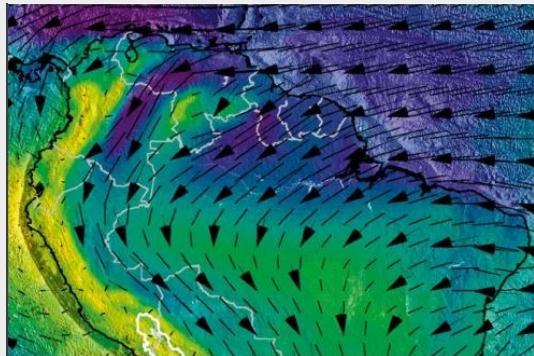
„Daten, die im Rahmen eines Forschungsvorhabens mit einer wissenschaftlichen Arbeitsmethode entstehen, nachgenutzt oder verarbeitet werden. Dies schließt als Daten vorliegende Operationalisierungen von Methoden und Testverfahren wie Fragebögen oder Software mit ein.“

## ➤ *Forschungsdaten-Policy der Universität Potsdam, Präambel*

- Umfasst unter anderem:
  - Rohdaten (z. B. aus Messungen, Beobachtungen)
  - Erstellte Daten (z. B. Annotationen, Simulationsmodell-Output)
  - Verarbeitete Daten (z. B. „bereinigte“ Rohdaten, Ergebnisse)
  - Werkzeuge als Daten (z. B. Fragebögen, Skripte, Software )
- Ausgeschlossen sind:
  - Daten aus Projektmanagement und Forschungsverwaltung

# Woran arbeiten Sie?

- Fachdisziplin
- Methoden
- Art der Daten
- Stand im Forschungsprozess



# Was ist Forschungsdatenmanagement?

Forschungsdatenmanagement (FDM) umfasst alle Aktivitäten, die mit

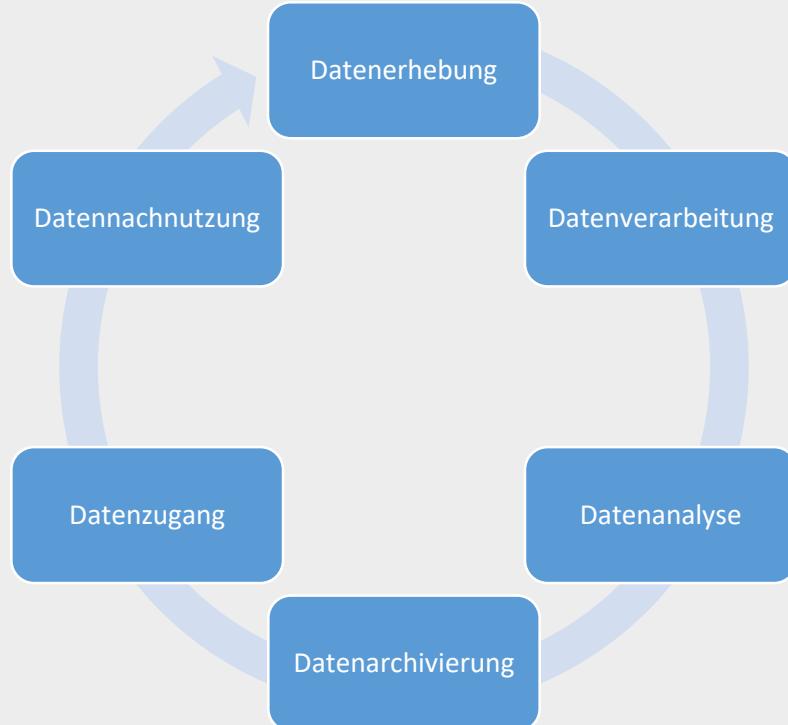
- Aufbereitung,
- Speicherung,
- Archivierung und
- Veröffentlichung

von Forschungsdaten verbunden sind.

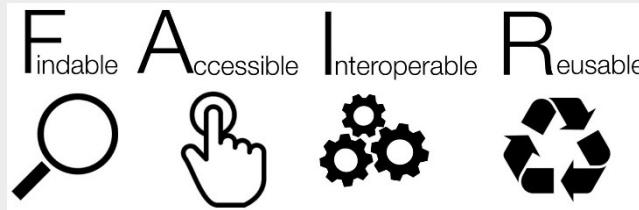
FDM begleitet den Forschungsprozess von den ersten Planungen bis zur Archivierung, Nachnutzung oder Löschung der Daten.

Quelle: Maxi Kindling, Peter Schirmbacher, Elena Simukovic: *Forschungsdatenmanagement an Hochschulen: das Beispiel der Humboldt-Universität zu Berlin*. LIBREAS. Library Ideas, 23 (2013). Online verfügbar unter: <http://libreas.eu/ausgabe23/07kindling/>

# Forschungsdaten-Lebenszyklen



Quelle: Forschungsdaten-Lebenszyklus nach dem UK Data Archive (Stand: Juni 2018).



# FAIR Data Principles (Summary)

**Auffindbarkeit:** Die Daten sind ausreichend mit relevanten Metadaten beschrieben und werden durch ein eindeutiges persistentes Identifikationsmerkmal (z. B. einen DOI) referenziert.

**Verfügbarkeit:** Die Daten sind für Menschen und Maschinen lesbar und werden in einem vertrauenswürdigen Repository vorgehalten.

**Interoperabilität:** Daten und Metadaten nutzen ein formalisiertes, frei verfügbares, weit verbreitetes und inhaltlich passendes Vokabular zur Wissensrepräsentation.

**Nachnutzbarkeit:** Die Daten stehen unter einer eindeutigen Lizenz zur Verfügung, enthalten korrekte Angaben zur Provenienz und sind gut dokumentiert.

Abbildung: Pundir, Sangya. [https://commons.wikimedia.org/wiki/File:FAIR\\_data\\_principles.jpg](https://commons.wikimedia.org/wiki/File:FAIR_data_principles.jpg) [letzter Zugriff: 27.09.2018], CC-BY-SA-4.0 <https://creativecommons.org/licenses/by-sa/4.0/deed.en>.

# Warum ist FDM wichtig? (1)

1. Wissenschaftskultur (DFG-Kodex, Institutionelle Policies)
2. Drittmittel (Förderrichtlinien, DMP)
3. Daten publizieren (Zeitschriften-, Verlags- und Repository-Policies)
4. Daten nachnutzen (Fachspezifische Empfehlungen und Policies, Kooperative Forschung)

# Warum ist FDM wichtig? (2)

- schnellere Auffindbarkeit von Daten und Übersichtlichkeit
- Wissenserhalt – Daten sind unabhängig von einzelnen Menschen, Projekten oder Institutionen zugänglich
- Transfer der Daten in künftige Projekte
- Erleichterung der Zusammenarbeit
- Langfristige Nachvollziehbarkeit von Ergebnissen, statt neues Erzeugen
- beugt Datenverlust vor
- (halb-)automatische Verarbeitung wird durch Metadaten ermöglicht
- Weitergabe und Nachnutzung von Daten
- Optimierter Mitteleinsatz
- Forschungsdatenzitation
- Referenzierbarkeit
- Erhöhung der Relevanz der eigenen Arbeit durch bessere Sichtbarkeit

# Wo bekomme ich Unterstützung beim FDM?

- FDM-Team von UB und ZIM
  - FDM-Webseite: <https://www.uni-potsdam.de/de/forschungsdaten/>
  - Policy und Empfehlungen: <https://doi.org/10.25932/publishup-44437>
  - RDMO (Testbetrieb): <https://rdmo.uni-potsdam.de/>
- Projektleitung und Arbeitsgruppenleitung
- DFG-Fachkollegien und Fachgesellschaften
- Disziplinspezifische Dienste und Infrastrukturen
  - Sozialwissenschaften: [GESIS](#), [RatSWD](#) und [CESSDA Guide](#)
  - Empirische Bildungsforschung: [VerbundFDB](#)
  - Geisteswissenschaften: [CLARIN-D](#) bzw. [CLARIAH-DE](#)
  - Naturwissenschaften: Fachabhängig, siehe auch: <https://www.re3data.org/>
- Informationsplattform [www.forschungsdaten.info](http://www.forschungsdaten.info)

# Kurzes Feedback

In dieser Session haben wir folgende Themen behandelt:

- Datenverlust, Datenskandale
- Begriffe wie Forschungsdaten, Lebenszyklus der Forschungsdaten & Forschungsdatenmanagement (FDM)
  
- Gibt es Fragen zu den gerade behandelten Themen?
- Ist etwas unklar geblieben?
- Soll noch auf etwas eingegangen werden?

# Pause (5 Minuten)



# Session 2: Forschungsdaten finden und verwalten

## 2.1. Daten finden

# Forschungsdaten finden

- Direkt in Fachrepositorien und fachübergreifenden Repositorien
- Mittels Metasuchmaschinen (z. B. [B2FIND](#), [gesisDataSearch](#), [Mendeley Data](#), [DataCite](#), [OpenAIRE](#))
- Recherche in bibliothekarischen Suchmaschinen (z. B. [BASE](#))
- Google: Stichwort und „data set“ bzw. [Google Dataset Search](#)

# Forschungsdaten prüfen

- Lizenzrechte prüfen
- Passen Daten zu meiner Fragestellung?
- Ist die Erhebung der Daten genau protokolliert und nachvollziehbar?
- Sind die Variablen gut beschrieben?
- Wurden die Daten seit der Erhebung bearbeitet?
- Ist die Quelle vertrauenswürdig?

# Datenzitation

- Nach FORCE11-Empfehlung:

Autor(en) (Publikationsjahr): Titel der Forschungsdaten.  
Datenrepository oder Archiv. Version. Weltweit persistenter  
Identifikator (vorzugsweise als Link)

- Nach DataCite 2013:

Urheber (Veröffentlichungsdatum): Titel. Version.  
Publikationsagent. Genereller Ressourcentyp. Identifikator

# Datenzitation

- Risan, Patrick (2017): Accommodating Trauma in Police Interviews. An Exploration of Rapport in Investigative Interviews of Traumatized Victims, 2013. NSD - Norwegian Centre for Research Data.  
<https://doi.org/10.18712/NSD-NSD2407-V1>
- Parton, William J.; Gutmann, Myron P.; Hartman, Melannie D.; Merchant, Emily R.; Lutz, Susan M. (2012): Great Plains Population and Environment Data: Biogeochemical Modeling Data, 1860-2003 [United States]. Version V1. ICPSR - Interuniversity Consortium for Political and Social Research. Datensatz. <https://doi.org/10.3886/ICPSR31681.v1>

## 2.2. Daten speichern

# Kriterien für die Auswahl von Speicherorten

- Wie viel Speicher wird benötigt?
- Für wen und von wo muss der Zugriff möglich sein?
- Sollen Daten versioniert werden?
- Gibt es ein automatisches Backup?
- Welches Schutzniveau ist nötig?
  - Authentifizierung
  - Verschlüsselung
  - Hardwarezugang
- Wer betreibt den Speicher, welche Regeln gelten?

# Unsere Tipps

- Für mittelgroße Datenmengen ohne Versionierung:  
[Box.UP](#) und [B2Drop](#)
- Zur Versionierung: [Git.UP](#)
- Für große Datenmengen: Netzlaufwerke statt  
Datenträger
- Bei Nutzung von VMs und Servern: Gibt es ein  
Backup? (Durch lokale Admins oder ZIM)
- Passwortmanager benutzen!

# Beschreibung von empfohlenen Diensten

**Übersicht**

**Projekt planen**

**Daten speichern und teilen**

**Publizieren und archivieren**

**Daten finden und zitieren**

## Daten speichern und teilen

Datenverlust ist durch Einsatz geeigneter Speicherdiene bzw. Speichermedien und angemessenes Backup vorzubeugen. Die Nutzung von wissenschaftsgelehrten Speicherdiene wird empfohlen, von der Nutzung lokaler Speichermedien und kommerzieller Speicherdiene als Privatkundin oder Privatkunde wird abgeraten. In vielen Fällen ist die Nutzung eines Versionierungsdienstes sinnvoll. Das notwendige Maß an Datensicherheit ist durch geeignete technische und organisatorische Maßnahmen sicherzustellen, z. B. durch wirksame Zugangsbeschränkungen oder die Pseudonymisierung von personenbezogenen Daten, die nicht anonymisiert werden können.

Bitte lesen Sie auch die [Handlungsempfehlungen zum Umgang mit Forschungsdaten](#) der Universität Potsdam.

### Sync & Share-Dienste

Mit einem Sync & Share-Dienst können Sie Ihre Daten zwischen mehreren Endgeräten (z.B. Desktop im Büro, Desktop zu Hause, Laptop und Smartphone) synchronisieren. Außerdem werden die Daten in der Cloud des Anbieters gespeichert, der auch ein Backup sicherstellt. Sie können Daten mit anderen Nutzer\*innen teilen und diese gemeinsam bearbeiten.

**Box.UP** ist der Sync & Share-Dienst der Universität Potsdam. Alle UP-Mitglieder können Box.UP mit ihrem Uni-Account nutzen. Professuren und Arbeitsgruppen können zusätzliche Box.UP-Accounts und mehr Speicher in Box.UP beantragen.

Daten, die in Ordnern liegen, die mit einem individuellen Account angelegt wurden, sind nicht mehr verfügbar, wenn dieser Account deaktiviert wird (z.B. weil die Mitarbeiterin die Universität Potsdam verlässt). Daten, die für die gesamte Arbeitsgruppe relevant sind, sollten daher in Ordnern abgelegt werden, die von institutionellen Accounts angelegt wurden.

Weitere Informationen finden Sie auf den [Box.UP-Seiten](#). Direkt zum Box.UP-Login: <https://boxup.uni-potsdam.de/>

Zurzeit ist Box.UP leider noch nicht an Systeme zum einrichtungsübergreifenden Login angebunden. Soll Box.UP für einrichtungsübergreifende Kooperationen genutzt werden, müssen Sie für die externen Partner\*innen UP-Accounts beantragen. Alternativ empfehlen wir die Nutzung anderer wissenschaftsgelehrten Dienste, wie B2drop.

**B2drop** ist der Sync & Share-Dienst von EUDAT, einer von der EU geförderten Collaborative Data Infrastructure (CDI): <https://www.eudat.eu/services/b2drop>

Quelle: ZIM

Quelle: <https://www.uni-potsdam.de/de/forschungsdaten/dienste/speichern>, Zugriff 25.05.2021

39

# Kriterien für Backup in eigener Verantwortung

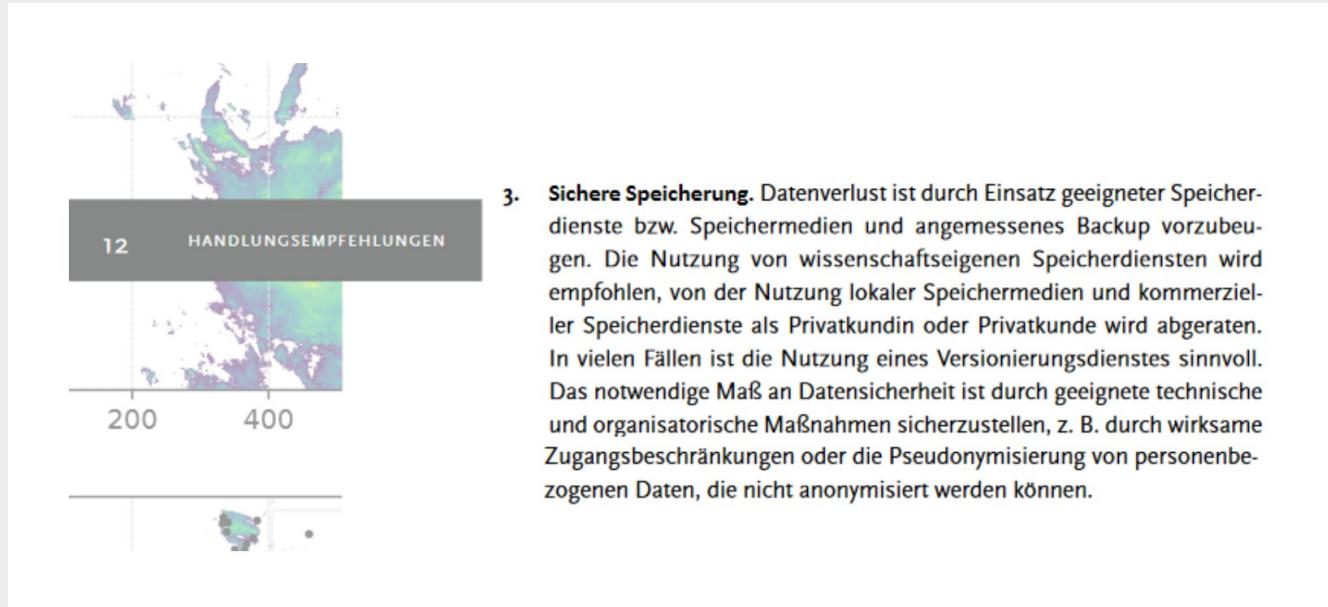
## 3-2-1 Regel

Testen Sie die Datenwiederherstellung zu Beginn sowie in regelmäßigen Abständen.

Schützen Sie Ihre (sensiblen) Daten:

- Hardware (bspw. separater abschließbarer Raum)
- Dateiverschlüsselung
- Sicherheit der Passwörter
- Mindestens zwei Personen sollten Zugang zu Ihren Daten haben

# Kriterien für Backup in eigener Verantwortung



Quelle: <https://publishup.uni-potsdam.de/frontdoor/index/index/docId/44437>, Zugriff 14.06.2023



WORLD  
BACKUP  
DAY //

MENU

DE ▾

# LASSEN SIE SICH NICHT IN DEN APRIL SCHICKEN.

Seien Sie vorbereitet. Sichern Sie Ihre Daten am **31. März**.



WAS IST EIN BACKUP? ↓

LEISTEN SIE DEN EID →

Quelle: <http://www.worldbackupday.com/de/>, Zugriff 25.05.2021

## DER "WORLD BACKUP DAY" EID

“Ich schwöre feierlich, am 31. März ein Backup meiner wichtigen Dokumente und wertvollen Erinnerungen zu machen.”

Ich werde auch meine Freunde und meine Familie über den "World Backup Day" informieren - Freunde lassen Freunde nicht ohne Backup steh'n.

---

LEISTEN SIE DEN EID AUF

---

FACEBOOK →

TWITTER →

## 2.3. Daten ordnen und strukturieren

# Übung

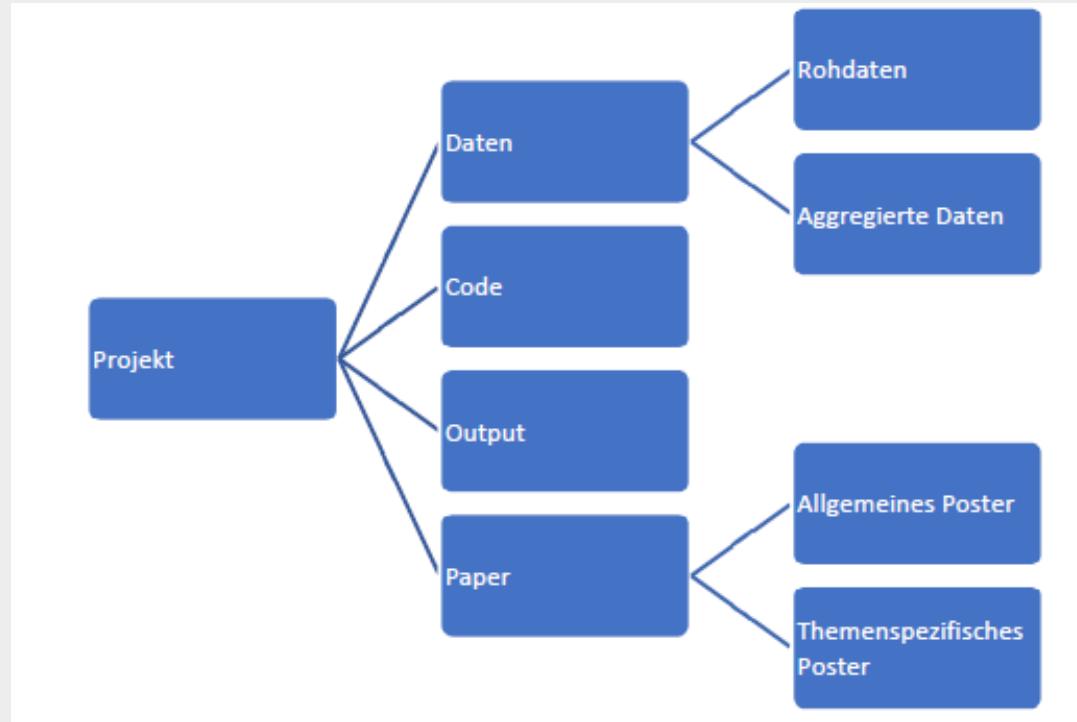
Wann habe ich schon mal Ordnung und/oder Struktur  
meiner Daten schmerzlich vermisst?



# Motivation für mehr Struktur

- Was, wie und weshalb getan wurde, bleibt auch nach Jahren nachvollziehbar
  - Andere können mit den Daten auch arbeiten
  - Daten schneller finden
  - Doppelte Arbeit wird vermieden
  - Datenverlust durch Überschreiben oder versehentliches Löschen wird vermieden
  - Aktuellen Stand erfassen
  - Maschinenlesbarkeit gewährleisten
- **Effizientes Arbeiten**

# Verzeichnisstruktur / Verzeichnisbaum



Quelle: Biernacka et al. 2021 S. 59

# Benennung von Dateien und Ordnern

- Kontext: inhaltsspezifische oder deskriptive Informationen, aussagekräftige Namen -> keine „Phantasienamen“
- Konsistenz: einheitliches Schema, logische Struktur
- Datumsangabe zur chronologischen Sortierung in folgender Form: JJJJMMTT
- Vermeidung von Leer- und Sonderzeichen
- Dokumentierte Namenskonventionen oder genutzte Abkürzungen, z. B.:  
[Sediment]\_[Probe]\_[Instrument]\_[JJJJMMTT].csv  
[Projekt]\_[Interview]\_[Ort]\_[Personen-ID]\_[JJJJMMTT].mp4

# Benennung von Dateien und Ordnern

- Inhalt
- Ersteller\*in
- Erstellungsdatum
- Bearbeitungsdatum
- Bezeichnung der Arbeitsgruppe
- Publikationsdatum
- Projektnummer
- Versionsnummer

# Beispiele

- 20180312\_h2oProbe1\_original.jpg
- 20180315\_h2oProbe1\_KDS\_Ausschnitt.jpg
- 20180324\_h2oProbe1\_KDS\_Ausschnitt\_bearbeitet\_bunt.jpg

# Beispiele

## interview02.docx

- Datentyp: Interview
- Der zweite Teil eines Interviews?
- Das zweite Interview?
- Interview mit Person Nr. 2?

## IP02R0120190731.docx

- I=Interview
- P02=Participant ID
- R01=Researcher ID
- 20190731=Datum des Interviews

# Übung

Welche der folgenden Beispiele folgen einer guten Benennungskonvention?



1. Olga\_170413\_probe17k
2. Naturepaper karl britta james fertig!
3. 647749157.pdf
4. 170413\_probe17k\_olga
5. Krst\_765\_spkt\_1203
6. Naturepaper+karl+britta+james fertig! überarbeitet
7. Kristall\_765\_spektr\_20161203
8. Nature\_karlbrittajames\_endendversion

# Werkzeuge für gleichzeitige Umbenennung

## Windows:

- Ant Renamer ([www.antp.be/software/renamer](http://www.antp.be/software/renamer))
- RenameIT ([sourceforge.net/projects/renameit](http://sourceforge.net/projects/renameit))
- Bulk Rename Utility ([www.bulkrenameutility.co.uk/](http://www.bulkrenameutility.co.uk/))
- Total Commander  
(<https://www.ghisler.com/deutsch.htm>)

# Werkzeuge für gleichzeitige Umbenennung

## Mac:

- Renamer 5 (for Mac) ([renamer.com/](http://renamer.com/))
- Name Changer ([mrrsoftware.com/namechanger/](http://mrrsoftware.com/namechanger/))
- ExifRenamer

## Linux:

- GNOME Commander  
(<http://mrrsoftware.com/namechanger>)
- GPRename

# Kontrolle der Dateiversion

Legen Sie obsolete Dateiversionen nach einem Backup separat ab, z. B. als eine separate ID (v1.0.0)

# Versionskontrolle

- Einfach Methode: vers. Versionen einer Datei anlegen und mit v1, v2, v1\_01, v2\_03 etc. versehen (nicht Bezeichnungen wie final)
- Software verwenden wie Subversion oder Git (z. B. Git.UP)
- Bei kollaborativen Dokumenten & Speicherorten meistens mit Versionierung und Änderungsverfolgung (z. B. Box.UP)



<https://download.logo.wine/logo/GitLab/GitLab-Logo.wine.png>

# Kurzes Feedback

In dieser Session haben wir folgende Themen behandelt:

- Daten in Repositorien finden
  - Sicherung und Backup
  - Daten benennen
- 
- Gibt es Fragen zu den gerade behandelten Themen?
  - Ist etwas unklar geblieben?
  - Soll noch auf etwas eingegangen werden?

# Pause (5 Minuten)



# 3. Forschungsdaten verständlich verfügbar machen

## 3.1. Daten dokumentieren

# Übung

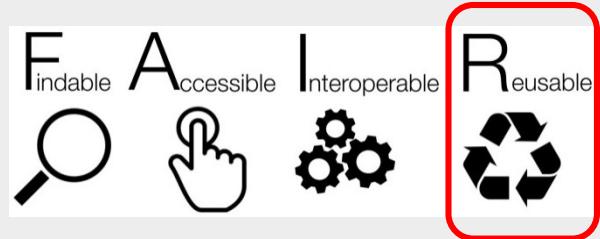
Wann hätten Sie schon mal eine gute Dokumentation oder Anleitung gebraucht?

Warum ist eine gute Dokumentation der Daten wichtig?

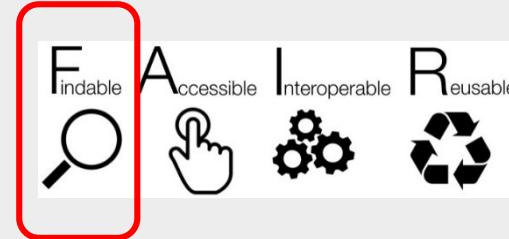
Was gehört in eine gute Dokumentation?

# Grundlegende Inhalte einer Dokumentation

- Beschreibung des Forschungsvorhabens
- Projektziele
- Hypothesen
- Informationen zur Erhebung der Daten (Methoden, Einheiten, Zeiträume, Orte, verwendete Technik)
- Maßnahmen zur Datenbereinigung
- Struktur der Daten und deren Beziehungen zueinander
- Erläuterung von Variablen, Labels und Codes
- Unterschiede zwischen verschiedenen Datensatz-Versionen
- Informationen zum Zugang und Nutzungsbedingungen



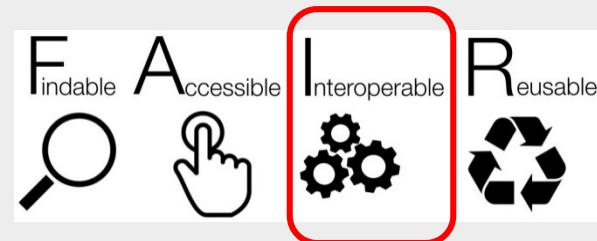
# Metadaten - Daten über Daten



- Inhaltliche Metadaten:
  - Titel
  - Beschreibung
  - Autor/in
  - Urheberrechts-Inhaber/in
  - Kontaktdaten
  - Lizenzangaben
  - Schlüsselwörter
- Technische Metadaten:
  - Aufnahmedatum
  - Brennweite
  - Blende
  - Belichtungsdauer
  - Geographische Koordinaten
  - und viele weitere

# Sie forschen über den russischen Schriftsteller....

- Alexander Issajewitsch Solschenizyn (D)
- Aleksandr Isayevich Solzhenitsyn (GB)
- Aleksandr İsayeviç Soljenitsin (T)
- Alexandre Issaïevitch Soljenitsyne (F)
- Aleksandr Isaevič Solženicyn (I)
- Aleksandr Isajevitsj Solzjenitsyn (N)
- Алексáндр Исаéвич Солжени́цын (R)
- (A) ألكسندر سولجنتسين
- (H) סולז'ניצין אלכסנדר



Virtual International Authority  
File - VIAF ID: 95215697  
(Person)  
Permalink:  
<http://viaf.org/viaf/95215697>

# Normdaten

Dienen der eindeutigen Identifikation von Personen, Institutionen, Forschungsförderer...

- Gemeinsame Normdatei (GND)
- International Standard Name Identifier (ISNI)
- Virtual International Authority File (VIAF)
- Open Funder Registry

- Open Researcher and Contributor ID (ORCID)



11 FREUNDE vor 12 Stunden

Superstark!

Die Tageszeitung »Hürriyet« schrieb über den neuen Entwicklungsminister Gerd Müller folgendes:

»Geboren 1945, früher Star der deutschen Nationalmannschaft und von Bayern München. Torschützenkönig der Nationalmannschaft.«

(Danke für den Hinweis geht an User Emrah K.)



The image is a screenshot of a Facebook post. At the top, it shows the profile picture of the poster and the text '11 FREUNDE vor 12 Stunden'. Below this is the post content: 'Superstark! Die Tageszeitung »Hürriyet« schrieb über den neuen Entwicklungsminister Gerd Müller folgendes: »Geboren 1945, früher Star der deutschen Nationalmannschaft und von Bayern München. Torschützenkönig der Nationalmannschaft.« (Danke für den Hinweis geht an User Emrah K.)'. Below the text is a composite image. On the left is a photo of Gerd Müller in a suit and tie, with vertical text on the left side of the photo reading 'KÜLE A Nrd er U'. On the right is a close-up photo of Gerd Müller's face, smiling, wearing glasses and a beard. At the bottom of the post are standard Facebook interaction buttons: 'Gefällt mir · Kommentieren · Teilen' and '1.732'. Below these buttons is a 'like' icon and the text 'Christian Gesslein, Jan-Niklas Löw, Tim My und 5.932 anderen gefällt das.'

Quelle: <https://www.augsburger-allgemeine.de/incoming/Gerd-Mueller-jpg-id28156617.html?aid=28164767>  
Abgerufen am 09.06.2023

# Kontrolliertes Vokabular

## Thesaurus

= eine natürlich-sprachliche geordnete Sammlung von Begriffen und deren Beziehungen zueinander.

Suche nach fachspezifischen Thesauri:  
<http://www.bartoc.org/en>

# Kontrolliertes Vokabular

## Klassifikation

Dient der Zuordnung von Objekten in (meist hierarchisch strukturierten) Klassen. Die Klassen sind durch bestimmte Merkmale charakterisiert.

- Beispiel: [Mathematics Subject Classification](#)

# Übung Metadaten (Teil1)

Welche Metadaten könnten in ihrem Fachgebiet vorkommen? (3 min. Brainstormen)

# Übung Metadatenstandards (Teil2)

Suchen Sie nach einem Metadatenstandard passend zu  
ihrem Fachgebiet.

<https://rdamsc.bath.ac.uk/>

# Metadatenstandards - Beispiele

- Fachübergreifende Metadatenstandards:
  - Dublin Core ([Dublin Core Generator](#))
  - MARC21
- Geowissenschaften:
  - ISO 19115
- Biodiversität:
  - Darwin Core
- Geisteswissenschaften:
  - Text Encoding Initiative (TEI)
- Naturwissenschaften:
  - ICAT Schema
  - Crystallographic Information Framework
- Sozialwissenschaften:
  - Data Documentation Initiative

# Elektronische Laborbücher (ELB)

- Für Dokumentation der Konzeptionierung, Durchführung und Auswertung von wissenschaftlichen Experimenten, Beobachtungen oder Versuchen und den Forschungsdaten
- Oft disziplinspezifisch
- Beispiele:
  - [Chemotion](#) (Open Source)
  - [eLabFTW](#) (Open Source)
  - [Labfolder](#) (kommerziell)
  - [openBIS](#) (Open Source)
  - [Rspace ELN](#) (kommerziell)

# README.txt

= Textdatei, die ein Projekt vorstellt und erklärt. Sie enthält Informationen, die in der Regel erforderlich sind, um zu verstehen, worum es in dem Projekt geht.

- **Warum?:** „*Because no one can read your mind (yet)*“
- **Wer?:** Jeder, der an einem Programmierprojekt arbeitet.
- **Wann?:** Bevor Sie ein Projekt anderen Personen zeigen oder es veröffentlichen, ggf. zur Gewohnheit machen, die Datei als erste in einem neuen Projekt zu erstellen.
- **Wo?:** Im obersten Verzeichnis des Projekts. GitHub, GitLab, etc. suchen nach einer README.txt

# README.txt - Vorlage

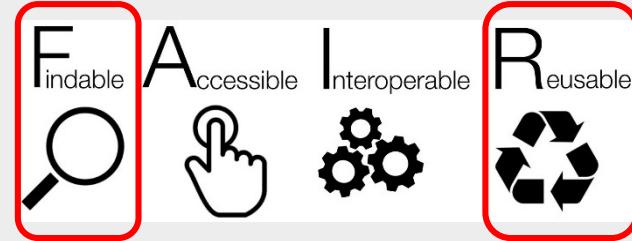
- Creator(s):
- Reason(s) for data analysis:
- Creation data of file(s):
- Used method(s):
- Used software (incl. Version and add-on) and tools:
- Data (file name and content):
- Code (filename and content):
- Additional files:
- License:
- Notes:

# Dokumentation und Metadaten...FAIR

- **Findability:** Die Daten sind mit ausreichend Metadaten versehen und durch persistenten Identifikator referenziert
- **Accessibility:** Die Daten sind sowohl für Menschen als auch Maschinen lesbar und in einem vertrauenswürdigen Repository abgelegt
- **Interoperability:** Daten und Metadaten nutzen ein formalisiertes, frei verfügbares, weit verbreitetes und inhaltlich passendes Vokabular zur Wissensrepräsentation
- **Reusability:** Die Daten stehen unter einer eindeutigen Lizenz zur Verfügung, enthalten korrekte Angaben zur Provenienz und sind gut dokumentiert

## 3.2. Daten publizieren

# Publikationswege für Daten



1. Ergänzung zu einem begutachteten Artikel („enhanced publication“)
2. Eigenständiges Informationsobjekt in einem Forschungsdaten-Repositorium
  - Als Ergänzung zu einer bestimmten Textpublikation
  - Unabhängig von einer bestimmten Textpublikation
3. Data Journals und Data Reports

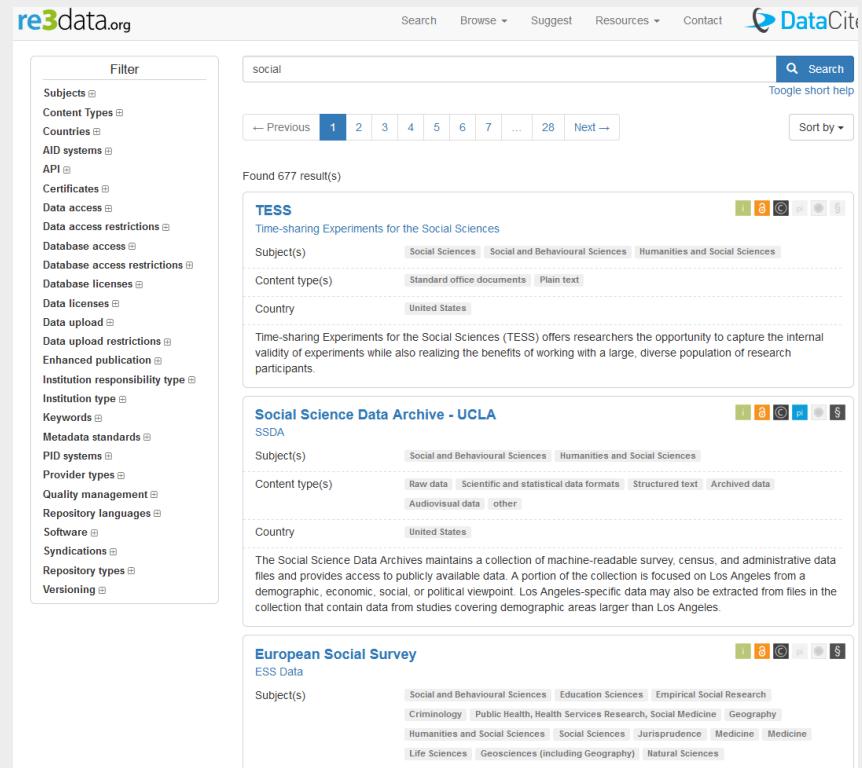
# Publikationswege für Daten

## 2. Eigenständiges Informationsobjekt in einem Forschungsdaten-Repositorium

- Disziplinspezifische Repositorien, z. B. Datorium, Pangaea bieten oft den besten Service, z.B. disziplinspezifische Metadaten und Qualitätssicherung der Datendokumentation
- Institutionelle Repositorien auch zur nicht-öffentlichen Archivierung
- Disziplinübergreifende Repositorien, z. B. ZENODO

# Repositorien finden: [re3data.org](http://re3data.org)

- Sammlung von Repositorien
- Weltweit
- Verschiedene Disziplinen
- Forschende, Förderer, Verlage und Institutionen



The screenshot shows the re3data.org search interface. The search bar at the top contains the word "social". Below the search bar is a navigation menu with links for Search, Browse, Suggest, Resources, Contact, and DataCite. The main content area displays search results for "social". The first result is "TESS: Time-sharing Experiments for the Social Sciences", which is a dataset from the Social Science Data Archive at UCLA. The second result is "Social Science Data Archive - UCLA SSDA", and the third result is "European Social Survey ESS Data". Each result card includes a brief description, subject categories, content type, and a link to the dataset.

re3data.org

Search Browse Suggest Resources Contact DataCite

social

1 2 3 4 5 6 7 ... 28 Next

Found 677 result(s)

**TESS**  
Time-sharing Experiments for the Social Sciences

Subject(s): Social Sciences, Social and Behavioural Sciences, Humanities and Social Sciences

Content type(s): Standard office documents, Plain text

Country: United States

Time-sharing Experiments for the Social Sciences (TESS) offers researchers the opportunity to capture the internal validity of experiments while also realizing the benefits of working with a large, diverse population of research participants.

**Social Science Data Archive - UCLA SSDA**

Subject(s): Social and Behavioural Sciences, Humanities and Social Sciences

Content type(s): Raw data, Scientific and statistical data formats, Structured text, Archived data, Audiovisual data, other

Country: United States

The Social Science Data Archives maintains a collection of machine-readable survey, census, and administrative data files and provides access to publicly available data. A portion of the collection is focused on Los Angeles from a demographic, economic, social, or political viewpoint. Los Angeles-specific data may also be extracted from files in the collection that contain data from studies covering demographic areas larger than Los Angeles.

**European Social Survey ESS Data**

Subject(s): Social and Behavioural Sciences, Education Sciences, Empirical Social Research, Criminology, Public Health, Health Services Research, Social Medicine, Geography, Humanities and Social Sciences, Social Sciences, Jurisprudence, Medicine, Medicine, Life Sciences, Geosciences (including Geography), Natural Sciences

Quelle: re3data About. <http://service.re3data.org/about>. Zugriff 16.03.2018

# Publikationswege für Daten

## 3. Data Journals und Data Reports

→ ausführliche Beschreibung der Daten  
→ Daten liegen im Repository



Quelle: <https://www.earth-system-science-data.net>, Zugriff 6.11.18



Quelle: <https://www.journals.elsevier.com/data-in-brief>, Zugriff 6.11.18

# Diskussion

Veröffentlichen Sie Ihre Forschungsdaten zurzeit?

Wenn ja, wo werden diese veröffentlicht?

# Forschungsdatenzentren (FDZ)

- FDZ bieten Dienste über Repositorium hinaus
  - Methodenberatung
  - Prüfung der Anonymisierung
  - Unterstützung bei der Datendokumentation
- Wichtigste Verbünde in Deutschland
  - FDZ im Rat für Wirtschafts- und Sozialdaten:  
<https://www.konsortswd.de/datenzentren/alle-datenzentren/>
  - Dienstekatalog der AG Datenzentren Digital Humanities:  
<https://dhd-ag-datenzentren-dienstekatalog.github.io/>

# Lizenzen – Nutzungsrechte einräumen

## Problem:

- Für die meisten Daten gilt das Urheberrecht nicht
- Unklare Nutzungsrechte behindern Nachnutzung trotzdem

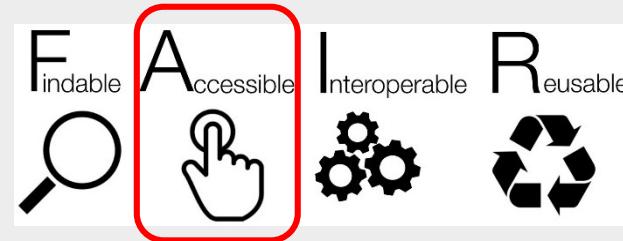
## Lösung:

- Standardisierte Lizenzen
- Permissive („freie“) Lizenzen

## Ausnahme:

- Daten dürfen nur eingeschränkt weitergegeben werden  
 („Scientific Use“)

# Lizenzvergabe konkret



- Creative Commons Lizenzen (für Daten/Inhalte)  
*UP empfiehlt i. d. R. CC0*
- Open Data Commons (insb. für Datenbanken)
- Freie Software-Lizenzen (für Software)  
*UP empfiehlt Nutzung von <http://choosealicense.com>*

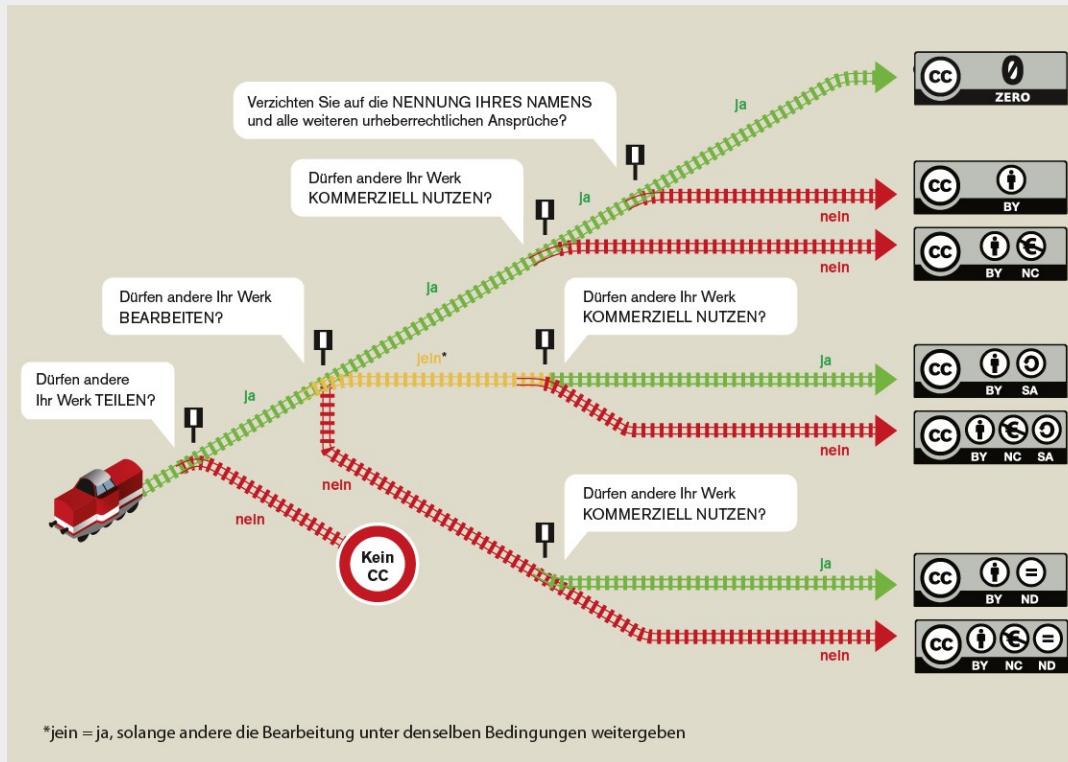
# Brainstorming

# Lizenzfrei

≠

# freie Lizenz

# Lizenzvergabe



Quelle: Barbara Klute und Jöran Muuß-Merholz [https://commons.wikimedia.org/wiki/File:Infografik\\_Auswahl\\_cc\\_lizenz.jpg](https://commons.wikimedia.org/wiki/File:Infografik_Auswahl_cc_lizenz.jpg), CC BY-SA 3.0, Zugriff 25.05.2021

# Archivierung – Abgrenzung zu Backup

## Backup

- Regelmäßig während Projektlaufzeit
- Automatische Sicherung **aller** Daten, um Datenverlust vorzubeugen (technisch oder menschlich)
- Alle Versionen

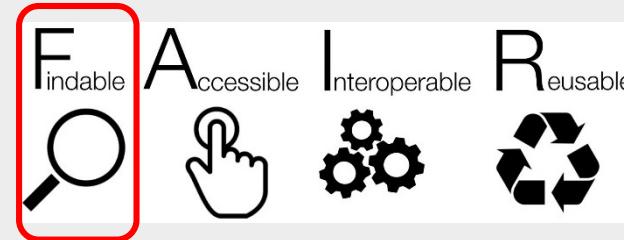
## Archivierung

- Zu bestimmten Zeitpunkten
- Sicherung **ausgewählter** Daten, um diese langfristig aufzubewahren
- Nur endgültige Versionen
- Integritätssicherung
- Langzeitspeicherung
- Durchsuchbarkeit

# Daten auswählen

- Meilenstein-Version
- Daten, die einer Publikation zugrunde liegen
- Datenqualität
- Rechte
- Einzigartigkeit der Daten
- Zeit/Kosten/Nutzen
- Technischer Aufwand
- Geeignete Dateiformate

# Persistente Identifier



- Ziel → langfristige Auffindbarkeit
  - „tote“ Links
  - Namensänderungen
- Digital Object Identifier (DOI)
- Open Researcher and Contributor ID (ORCID)

### 3.3. Daten langzeitarchivieren

# Wie lange ist Langzeit?



CD: 5-10 Jahre



Festplatte: 3-10 Jahre



Diskette: 10-20 Jahre

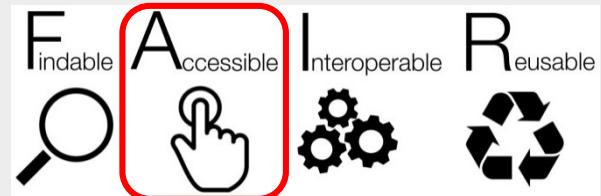
Quelle: pixabay.com.  
Dieses Werk ist lizenziert unter einer [CC0 1.0 Universal \(CC0 1.0\) Public Domain Dedication](https://creativecommons.org/publicdomain/zero/1.0/deed.de).  
<https://creativecommons.org/publicdomain/zero/1.0/deed.de>

# Nachhaltige Dateiformate

Dateiformat	Empfehlung	Vermeiden
Tabellen	CSV, TSV, SPSS portable, XLSX	XLS, SPSS
Text	TXT, HTML, RTF, PDF/A, DOCX	DOC, PDF
Multimedia	Container: MPEG4, MKV Codec: Theora, Dirac, FLAC	QuickTime, Flash
Bilder/Grafiken	TIFF, JPEG2000, PNG	GIF, JPEG

# Anforderungen an Langzeitarchive

- Technische Anforderungen
- Siegel für vertrauenswürdige Langzeitarchive  
(z. B. CoreTrustSeal, nestor-Siegel, DIN 31644)
- Kosten
- Zugänglichmachung der Daten
- Langlebigkeit des Dienstleisters



# Kosten für FDM

- Personalkosten
  - Materialkosten (z. B. Software, Server, etc.)
  - Dienstleistungskosten (z. B. Gebühren für nachgenutzte Daten, für die Archivierung, für die Publikation, etc.)
- [forschungsdaten.info: Budgetplanung](https://forschungsdaten.info: Budgetplanung)

# Kurzes Feedback

In dieser Session haben wir folgende Themen behandelt:

- Metadaten
  - Publikation
  - Archivierung
- 
- Gibt es Fragen zu den gerade behandelten Themen?
  - Ist etwas unklar geblieben?
  - Soll noch auf etwas eingegangen werden?

# Übung Inventur

## Fragen vom Anfang anschauen:

- Wo sind Ihre privaten und dienstlichen Daten gespeichert?
- Sind Sie sicher? Gibt es Backups?
- Sind diese gut beschriftet? Automatische Metadatengenerierung
- z. B. bei Fotos?
- Was könnten Sie verbessern?

Was würden Sie jetzt anders machen? Konnten Sie heute etwas für sich mitnehmen?

Teil 2 im November:

Rechtliche Aspekte, Lizenzen, Datenmanagementpläne, ...

Offene Fragen? Feedback?

Herzlichen Dank für Ihre  
Aufmerksamkeit!

# Wo bekomme ich Unterstützung beim FDM?

- FDM-Team von UB und ZIM [forschungsdaten@uni-potsdam.de](mailto:forschungsdaten@uni-potsdam.de)
  - FDM-Webseite: <https://www.uni-potsdam.de/de/forschungsdaten/>
  - Policy und Empfehlungen: <https://doi.org/10.25932/publishup-44437>
  - RDMO (Testbetrieb): <https://rdmo.uni-potsdam.de/>
- Projektleitung und Arbeitsgruppenleitung
- DFG-Fachkollegien und Fachgesellschaften
- Disziplinspezifische Dienste und Infrastrukturen
  - Sozialwissenschaften: [GESIS](#), [RatSWD](#) und [CESSDA Guide](#)
  - Empirische Bildungsforschung: [VerbundFDB](#)
  - Geisteswissenschaften: [CLARIN-D](#) bzw. [CLARIAH-DE](#)
  - Naturwissenschaften: Fachabhängig, siehe auch: <https://www.re3data.org/>
- Informationsplattform [www.forschungsdaten.info](http://www.forschungsdaten.info)

# Nachnutzung und Bildnachweise

Konzept und Folien des Workshops basieren auf dem im BMBF-Verbundprojekt FD-Mentor entwickelten **Train-the-Trainer Konzept zum Thema Forschungsdatenmanagement** (Version 4.0), 2021

- Online auf Zenodo: <https://doi.org/10.5281/zenodo.5773203>
- Autor\*innen: Biernacka, Katarzyna, Buchholz, Petra, Danker, Sarah Ann, Dolzycka, Dominika, Engelhardt, Claudia, Helbig, Kerstin, Jacob, Julianne, Neumann, Janna, Odebrecht, Carolin, Petersen, Britta, Slowig, Benjamin, Trautwein-Brunn, Ute, Wiljes, Cord, & Wuttke, Ulrike.
- Lizenziert unter: [Creative Commons Attribution 4.0 International](#)



Bildnachweis “Band Forschungsdaten”  
v.l.n.r.:

- Netzwerkgraphen zu deutschsprachigen Dramen aus dem DLINA Korpus (1731–1929). Fischer et al., Figshare (2015). <http://doi.org/drkfCCBY4.0>
- Mikrofotografien von Janus-Tröpfchen. Raju et al., RSC Advances 9 (2019), 19271, Abb. 5. <http://doi.org/drj9CCBY3.0>
- Mittlere Windgeschwindigkeit an der 850 hPa Druckfläche über Südamerika im Südsommer. Ziarani et al., Atmosphere 10(2019), 379, Abb. 1. <http://doi.org/dpjsCCBY4.0>