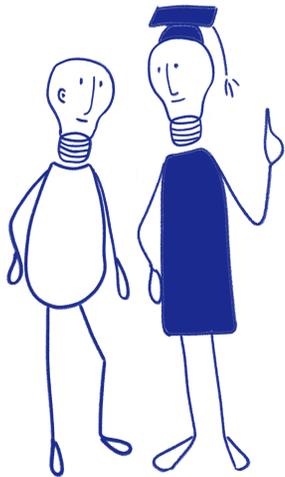


# Workshop für nachhaltiges Forschungsdatenmanagement



MIN-Graduiertenschule - Chemie  
Kathrin Thiemann, Hagen Peukert  
19.11.2019

# Themen

- Vorstellungsrunde
- Was sind Forschungsdaten?
- (Nachhaltiges) Datenmanagement
- Übung DMP ohne Hilfsmittel
- Datenmanagementplan
- Übung RDMO
- Policies und Anforderungen der Förderer
- Rechtliches und Open Access
- **Mittagspause**
- Forschungsdaten akquirieren
- Datenanalyse, -speicherung, Workflows & Zusammenarbeit
- Übung Datenbereinigung
- Datenspeicherung (Repositorien und Metadaten)
- Übung Metadaten

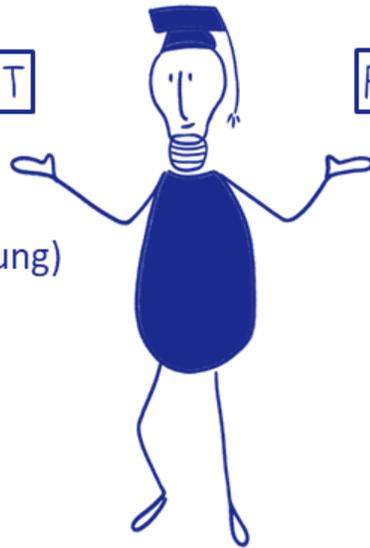


<https://legacy.gitbook.com/book/open-science-training-handbook/book/details>

# ZFDM-"SÄULEN"

## FORSCHUNGSDATENMANAGEMENT

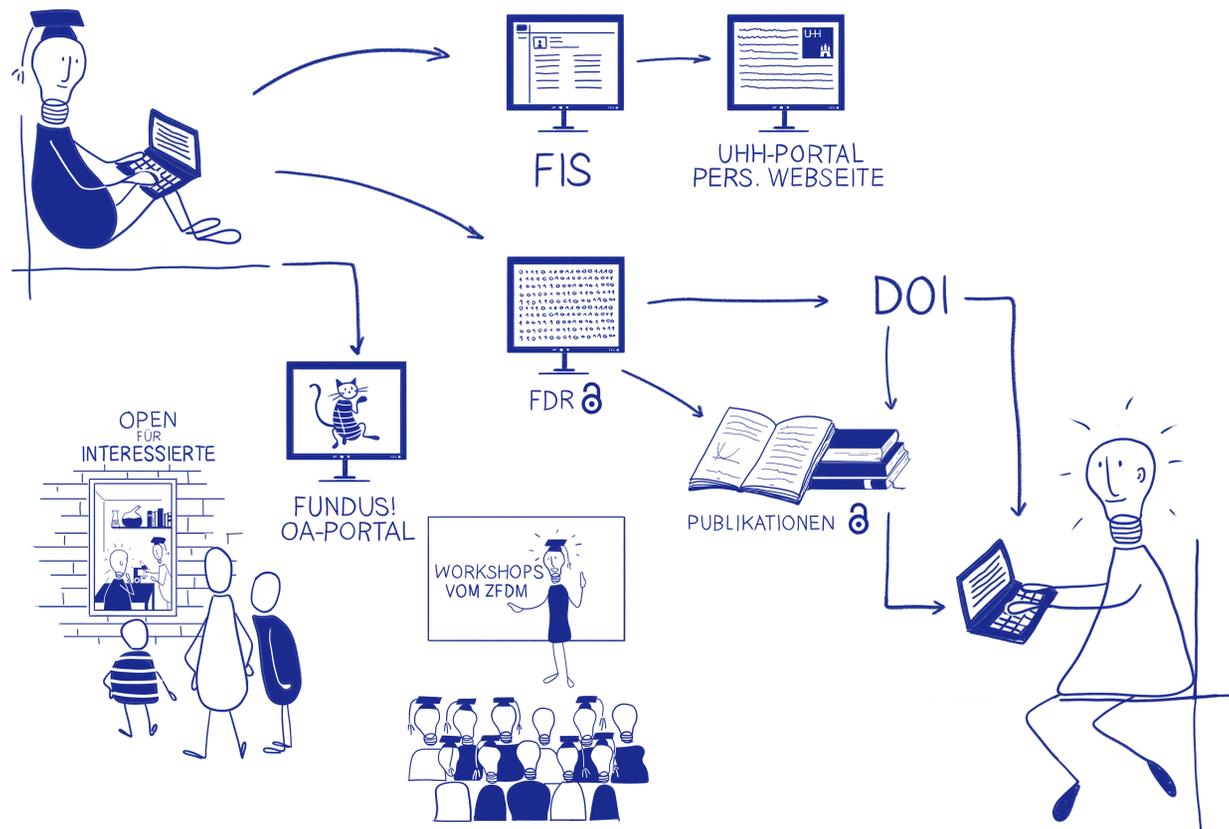
- Beratung (Projektantrag bis Archivierung)
- Service-Infrastruktur (Repositorien, Portale)
- Vernetzung (Technik-Treffen, RRZ)
- Schulung
- Strategische Planung, Policies
- Workshops zu FDM



## FORSCHUNGSMFORMATIONSSYSTEM

- Information zu Forschung an der UHH
- Publikationsverzeichnis
- Veröffentlichung im Web-CMS (Fiona)

# Services des ZFDM



# Hamburg Open Science (HOS)



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



[www.haw-hamburg.de](http://www.haw-hamburg.de)



[www.tuhh.de](http://www.tuhh.de)



[www.hcu-hamburg.de](http://www.hcu-hamburg.de)

**H F B K**

Hochschule für bildende  
Künste Hamburg

[www.hfbk-hamburg.de](http://www.hfbk-hamburg.de)



hochschule für musik und theater

[www.hfmt-hamburg.de](http://www.hfmt-hamburg.de)



HAMBURG

[www.uke.de](http://www.uke.de)



STAATS- UND UNIVERSITÄTS-  
BIBLIOTHEK  
HAMBURG  
CARL VON OSSIETZKY

[www.sub.uni-hamburg.de](http://www.sub.uni-hamburg.de)



Hamburg

Behörde für Wissenschaft,  
Forschung und Gleichstellung

[www.hamburg.de/bwfg](http://www.hamburg.de/bwfg)

Schaufenster [openscience.hamburg.de](http://openscience.hamburg.de)

# Wer sind Sie?

- Ihr Forschungsvorhaben?
- Ihre Daten?
- Ihre Erwartungen an den Workshop?
- ....

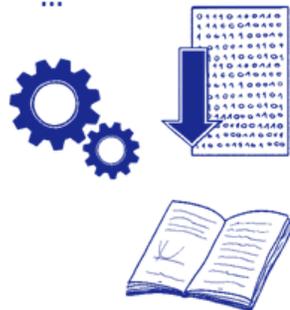
# 1. Was sind Forschungsdaten?

Forschungsdaten sind **alle Daten, die im Zuge eines wissenschaftlichen Vorhabens entstehen**. 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 (DFG 2015)

# Arten von Forschungsdaten

## METADATEN, z.B.:

- Name Autor
- Marke / Modell Kamera
- Kalibration / Einstellungen
- ...



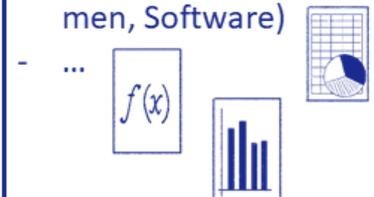
## PRIMÄRDATEN / ROHDATEN, z.B.:

- Messdaten, Manuskripte, Fotografien, Aufnahmen, ...
- Digitalisate
- ...



## SEKUNDÄRDATEN, z.B.:

- Aufbereitete Daten (Selektion, Korrektur, Aggregation, ...)
- Dokumentation (Methoden, Laborbücher, Versuchsprotokolle)
- Prozeduren (Algorithmen, Software)
- ...



## FIS

- Administrative Information, z.B. Mitarbeiter, Laufzeiten, Finanzierung von Projekten
- Förderanträge
- ...



# Herkunft von Forschungsdaten



Aust, Pamela, Helbig, Kerstin, Schenk, Ulrike, Zielke, Dennis, Rosenbaum, Anja, Schulze, Jörg (2016): Was sind Forschungsdaten? Video. Humboldt-Universität zu Berlin, Medien-Repositoryum.

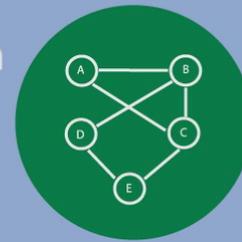
# Form von Forschungsdaten



statistische Daten  
Laborwerte



digitalisierte Karten  
Audio- oder  
Videoaufnahmen



Algorithmen  
Skripte  
Netzwerkanalysen



digitale Labor- oder  
Feldtagebücher  
Transkripte  
Annotationen  
Fragebögen



Proben  
Artefakte  
geräteabhängige  
Messwerte

Aust, Pamela, Helbig, Kerstin, Schenk, Ulrike, Zielke, Dennis, Rosenbaum, Anja, Schulze, Jörg (2016): Was sind Forschungsdaten? Video. Humboldt-Universität zu Berlin, Medien-Repositoryum.

# Forschungsdaten im Fokus der Wissenschaftsorganisationen

Deutsche  
Forschungsgemeinschaft

## Leitlinien zum Umgang mit Forschungsdaten

**HRK** Hochschulrektorenkonferenz  
Die Stimme der Hochschulen

Die HRK Mitglieder Themen Positionen Publikationen Presse

HRK / Positionen / Gesamtliste Beschlüsse / Position

Position



Positionen

Gesamtliste Beschlüsse

Beschlüsse nach Thema

Management von Forschungsdaten - eine zentrale strategische Herausforderung für Hochschulleitungen



## Die Ressource Information besser nutzbar machen!

Positionspapier zum Umgang mit Forschungsdaten in der Helmholtz-Gemeinschaft



Rat für  
Informations  
Infrastrukturen

## LEISTUNG AUS VIELFALT

Empfehlungen zu Strukturen, Prozessen und Finanzierung  
des Forschungsdatenmanagements in Deutschland

# Ihre Erfahrungen mit (Forschungs-)Daten?

## Haben Sie schon mal...

---

... Daten nachgenutzt?

... Daten „verloren“?

---

... Daten beschrieben und dokumentiert?

... Daten nicht wiedergefunden?

---

... Ergebnisse einer fremden wissenschaftlichen Arbeit überprüft?

... Daten nicht öffnen/lesen können?

---

... Daten veröffentlicht?

... eigene/fremde Daten nicht (mehr) nachvollziehen können?

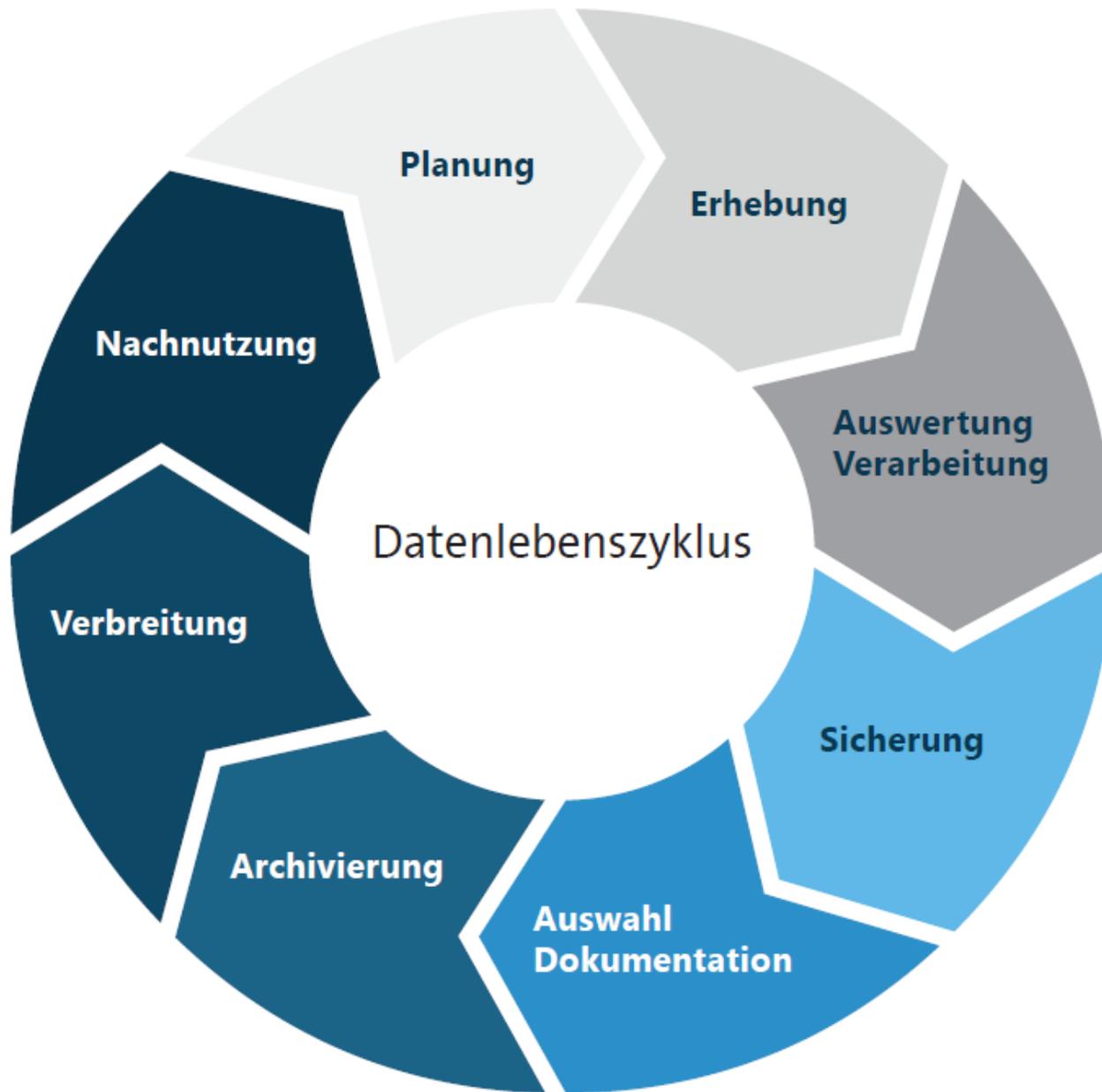
# 2. Nachhaltiges Forschungsdatenmanagement

# Was ist FDM?

Im Rahmen vom FDM werden Strategien entwickelt, um die Arbeitsprozesse, die die Erzeugung von und den Umgang mit Forschungsdaten betreffen, möglichst effizient zu organisieren und steuern. Damit begleitet FDM die Forschung von den ersten Planungen bis zur Archivierung, Nachnutzung oder Löschung der Daten.

Aspekte von FDM:

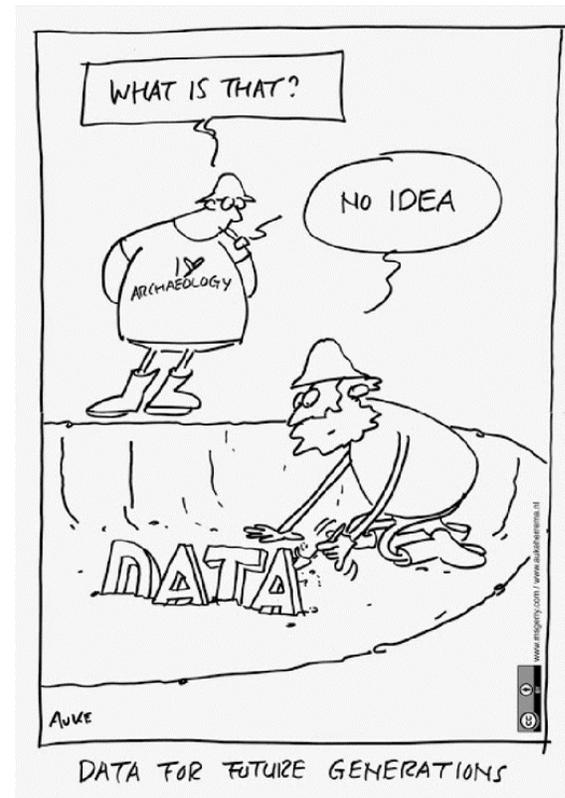
- Ordnung und Struktur
- Planung
- Speicherung und Backup
- Sicherheit und Zugriffsrechte
- Langzeitarchivierung
- Publikation von Forschungsdaten
- Rechtliche Aspekte



# Wozu FDM?

## Gutes FDM erleichtert die Nutzung

- Daten schneller finden (z.B. durch Benennung, Struktur)
- Datenverlust verhindern (Daten unabhängig von Personen)
- (halb-) automatische Verarbeitung ermöglichen (Metadaten)
- Ergebnisse langfristig verstehen (Dokumentation!)
- Daten langfristig nutzen
- Mittel optimal einsetzen (durch potentielle Nachnutzung)
- Auflagen von Geld-/Drittmittelgebern erfüllen



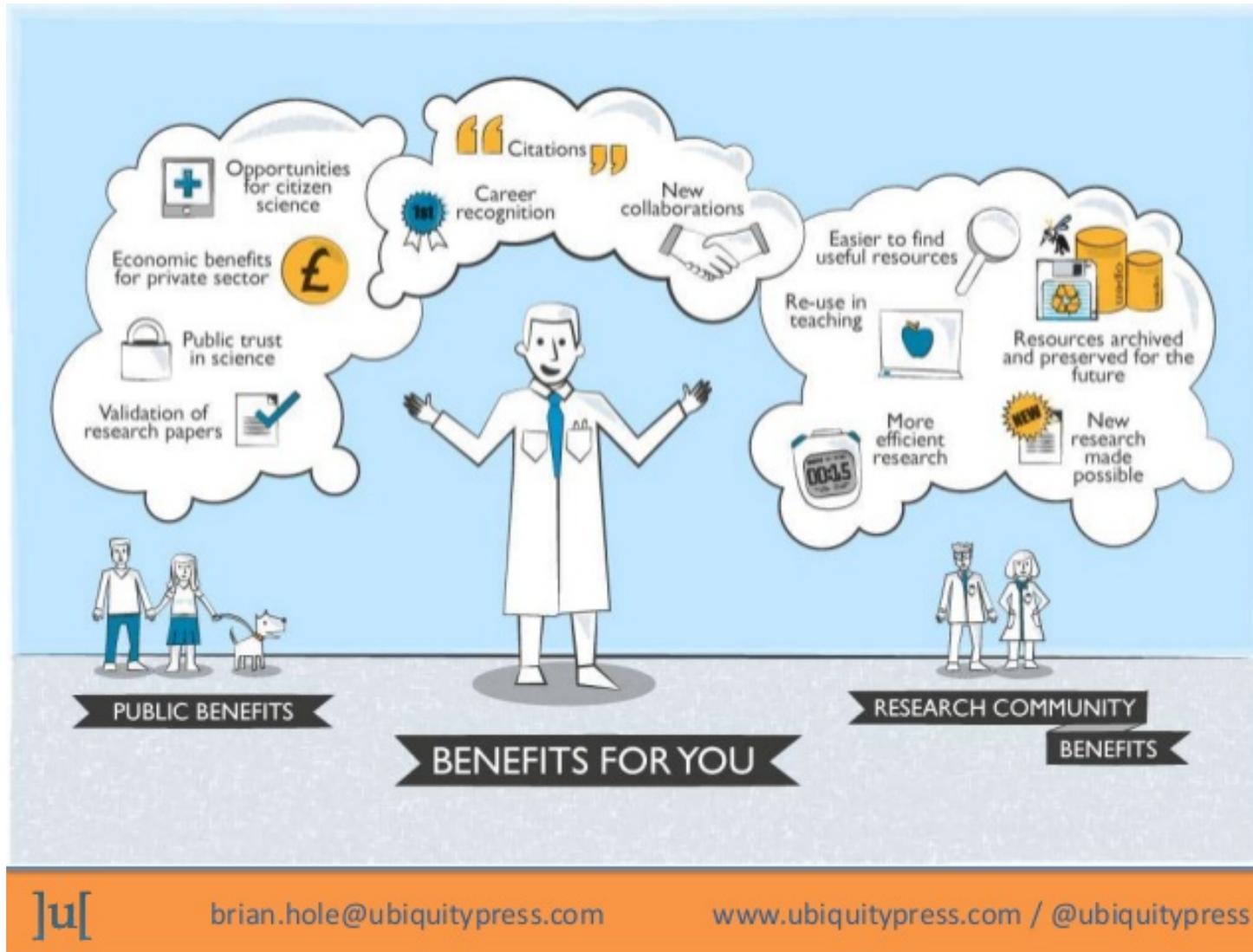
# ”Nachhaltiges” FDM

## Open Data

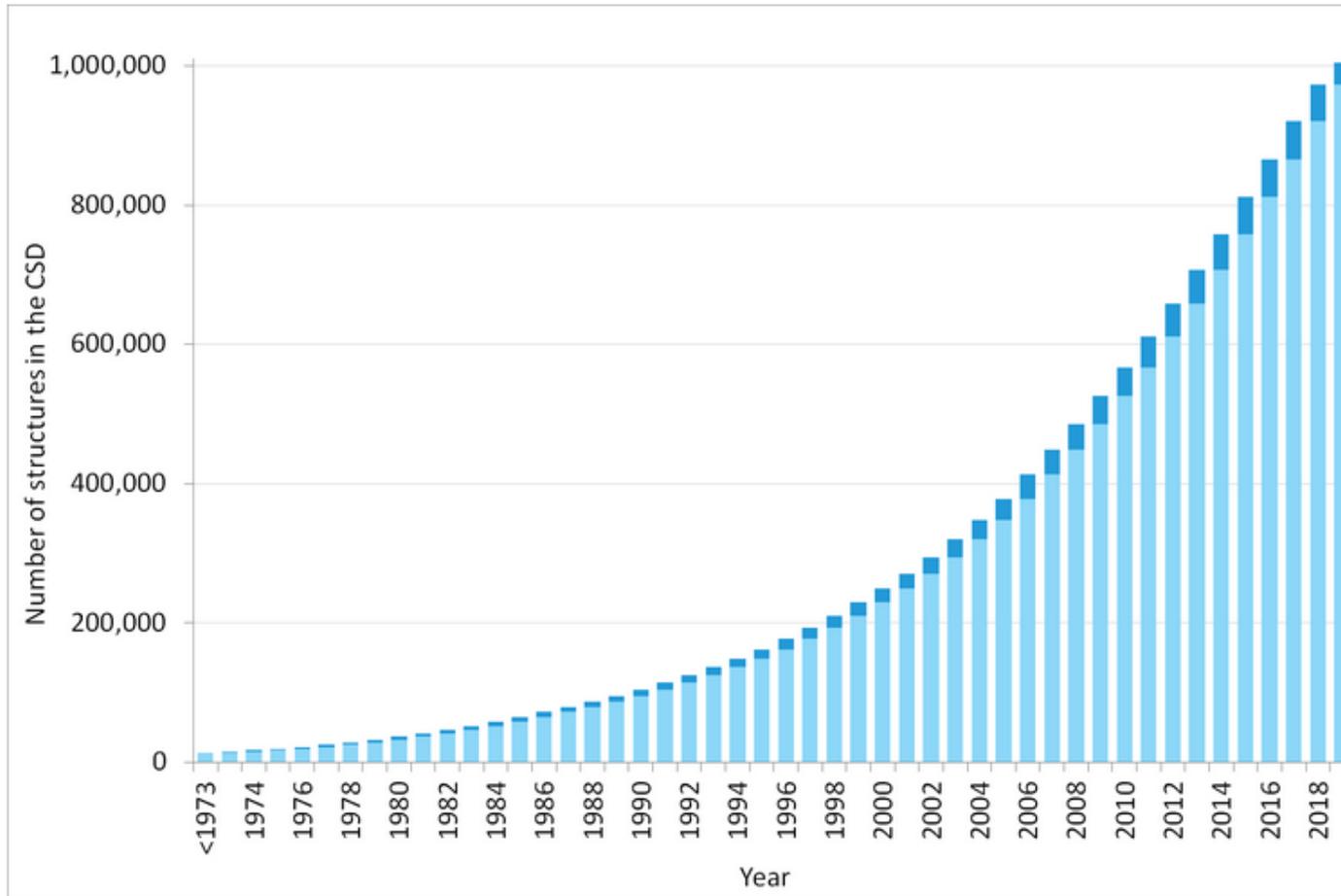
- Nutzbar machen, statt nur zu nutzen
- Freiheitsgedanke, der über kostenfrei hinausgeht
- Adaption und Aktualisierung
- Kollaboratives Arbeiten
- Gemeinschaft stärken
- Ideenvielfalt



# Warum sollte ich meine Daten überhaupt teilen?



# Datenwachstum Repositorien



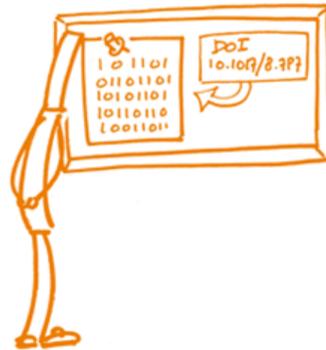
*Growth of the CSD since 1972, the red bar shows structures added annually.*

Cambridge Structural Database

**F**indable



**A**ccessible



**I**nteroperable

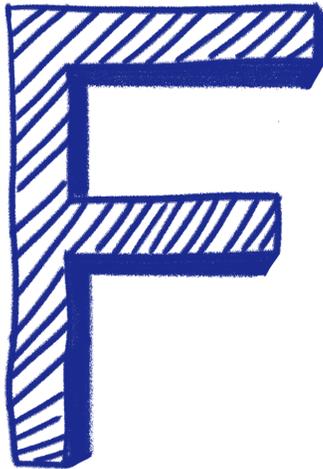


**R**eusable -Prinzipien



# Findable

Forschungsdaten müssen auffindbar sein



- Beschreibung durch Metadaten (maschinenlesbar)
- Zitierfähigkeit mit Hilfe von dauerhaft persistenter Identifier (PIDs)
- Sichtbarkeit (Registrierung der MD in Nachweissystemen und Katalogen)

re3data.org  
REGISTRY OF RESEARCH DATA REPOSITORIES

Home Search Browse Suggest FAQ About

## Browse by subject

Graphical Text

A. Humanities and Social Sciences

a. Humanities

I. Ancient Cultures

1. Prehistory
2. Classical Philology
3. Ancient History
4. Classical Archaeology
5. Egyptology and Ancient Near Eastern Studies

II. History

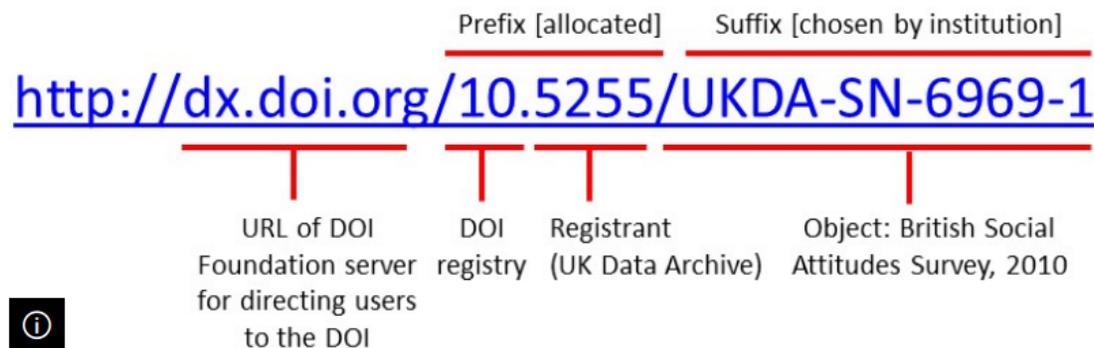
1. Medieval History
2. Early Modern History
3. Modern and Current History
4. History of Science

III. Fine Arts, Music, Theatre and Media Studies

# Was sind Persistente Identifier (PID)?

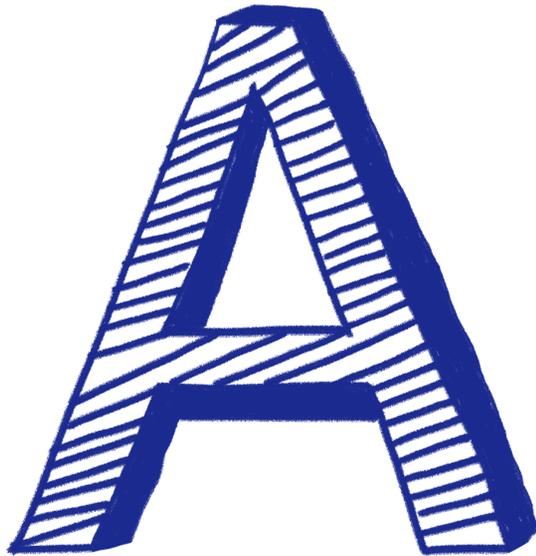
- dauerhafter, digitaler Identifikator, bestehend aus Ziffern und/oder alphanumerischen Zeichen, welcher einem digitalen Objekt zugeordnet wird und direkt auf diesen verweist
- für Forschungsdaten:
  - DOI (Digital Object Identifier)
  - URN (Uniform Resource Name)
  - Handle
  - ORCID (Personen)
- URLs problematisch, da sie nicht auf einen Inhalt, sondern auf einen Ort verweisen
- idealerweise ändert sich nichts an der PID, wenn sich der Ablageort des digitalen Objekts ändert
- wird das Objekt gelöscht, bleibt eine Landing-Page mit entsprechenden Metadaten erhalten

Durch PIDs sind digitale Objekte dauerhaft auffindbar, abrufbar und zitierbar.



# Accessible

FD müssen (langfristig) zugänglich sein



- FD sind über ihren PID mit Hilfe standardisierter, offener Protokolle zugänglich (HTTP(S), FTP, SMTP)
- Das Protokoll unterstützt, wo notwendig, die Authentifizierung und Rechteverwaltung
- sichere Langzeit-Archivierung
- Metadaten bleiben verfügbar, auch wenn Ressourcen gelöscht wurden



# Was sind Metadaten?

- strukturierte Informationen über andere Daten bzw. Ressourcen und deren Merkmale → Daten über Daten
- können zusammen mit oder getrennt von den Daten, die sie näher beschreiben, abgespeichert werden
- Unterscheidung zwischen:
  - administrative Metadaten (Autor, Titel, Format, Größe...)
  - inhaltsbeschreibende bzw. fachlichen Metadaten (Messmethode, Probe, Interviewperson...)
- Metadatenstandards (z.B. Dublin Core, DataCite, Marc21) ermöglichen Interoperabilität und Harvesting von Metadaten
- Disziplinspezifische Metadatenstandards unter [www.dcc.ac.uk/resources/metadata-standards](http://www.dcc.ac.uk/resources/metadata-standards)

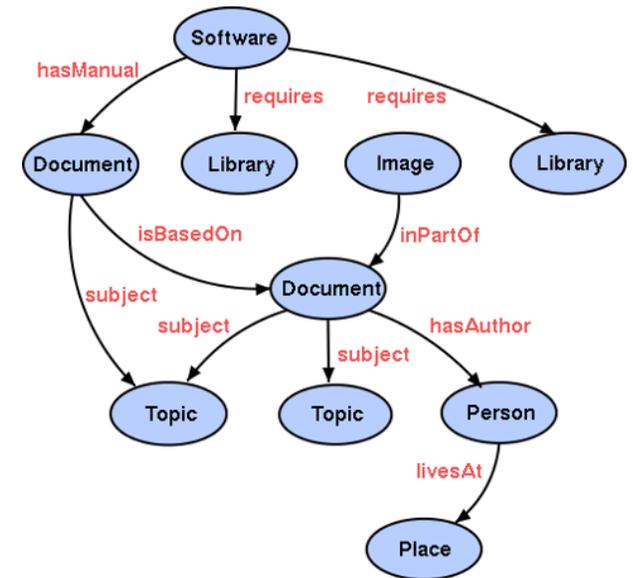
Standardisierung spielt insbesondere bei administrativen Metadaten eine wichtige Rolle.

# Interoperable

## FD müssen interoperable sein

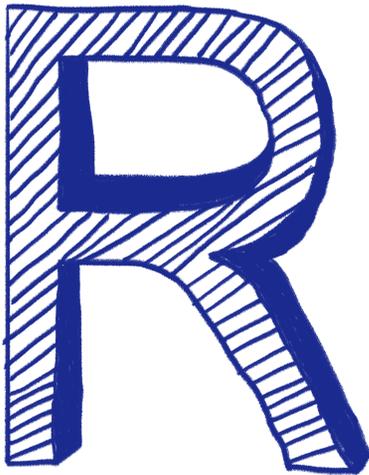


- standardisierte, maschinenlesbare Metadaten
- kontrollierte Vokabulare, Wertebereiche, Thesauri und Ontologien
- standardisierte Referenzen auf andere MD/ FD (z.B. ‚is new version of‘, ‚is supplement to‘, ‚relates to‘, )
- Verwendung von offenen, dokumentierten Formaten



# Reusable

FD müssen wiederverwendbar sein



- Präzise, standardisierte, detaillierte Beschreibung
- eindeutige, zugreifbare Nutzungslizenz
- klare, detaillierte Angaben zur Herkunft (Provenienzinformationen)
- FD und MD entsprechen den fachgebietsrelevanten Standards

# (Welche) Daten aufbewahren?

- Müssen die Daten aus rechtlichen Gründen aufbewahrt werden, wenn ja, wie lange?
- Können die Daten ohne Informationsverlust erhalten werden?
- Was bedeutet der Informationsverlust für die Daten, sind sie noch brauchbar?
- Gibt es Interessenten an den Daten?
- Sind die Daten Teil einer größeren Sammlung (z.B. Zeitreihe)?
- Können die Daten neu erhoben werden, wenn ja, mit welchem Aufwand?
- Ist eine Nachnutzung der Daten zu erwarten?
- Welche Art der Nachnutzung ist zu erwarten?
- Sind die Daten von kulturellem / gesellschaftlichem Wert?
- Sind die Daten von genügender Qualität für die Nachnutzung?
- Können die Daten für Lehrzwecke genutzt werden?

# Welche Daten aufbewahren?

Entscheidung sollte von potentieller Nachnutzung abhängen:

## 1. Weitere Publikation

→ referenzierte (=verarbeitete) Daten mit zusätzlicher Dokumentation

## 2. Verifikation

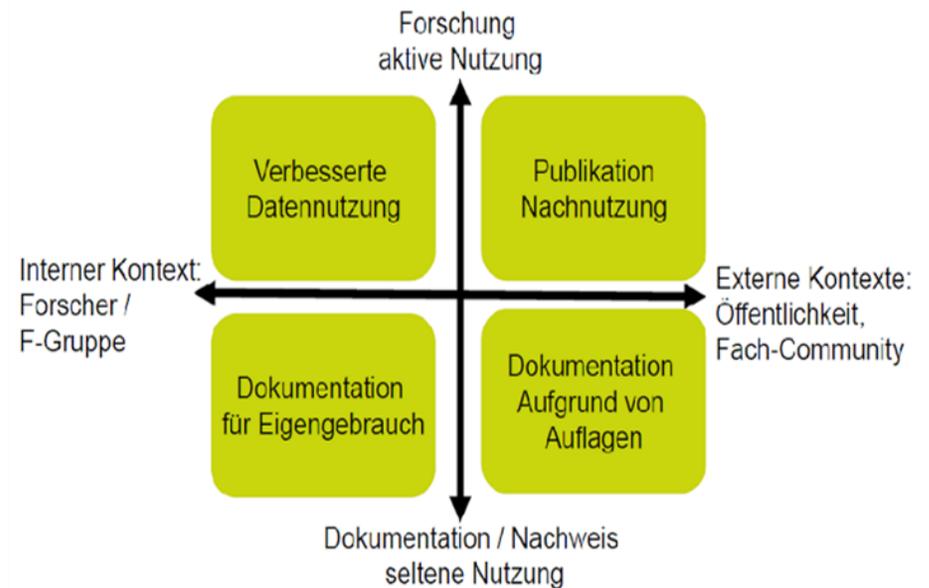
→ referenzierte Daten inklusive Analyseschritte

## 3. Weitere Analysen

→ alle Originaldaten

## 4. Lernen und Lehren

→ Proben von Originaldaten und zusammengestellten Daten inklusive Analyseschritte

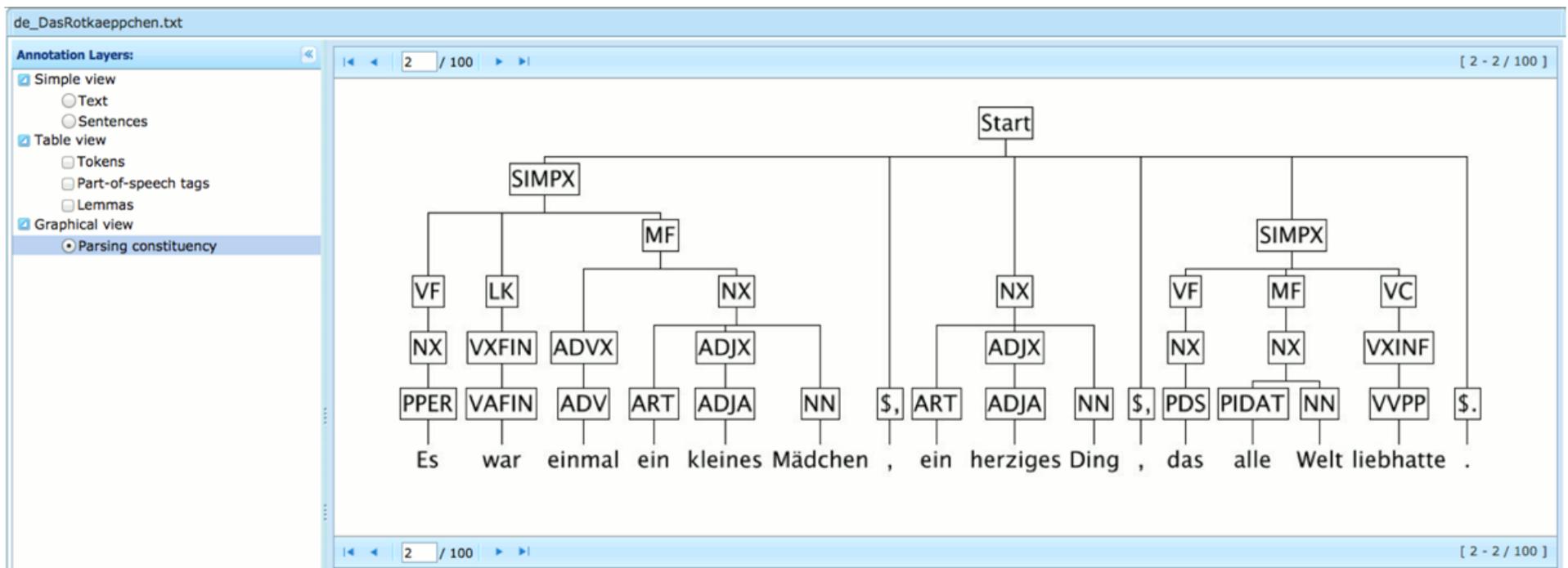


Jens Ludwig, Was sind Forschungsdaten? Nestor PERICLES School 2016

# Welche Daten aufbewahren?

Forschungsdaten müssen analysierbar und interpretierbar bleiben

- Bewahrung von Präsentations- und Analyseumgebungen
- Dokumentation von Annotations- und Analyseprozessen



# Was kann weg?

- technisch bedingte Strukturen und Daten
- leere Datensätze / Tabellen
- Duplikate
- unvollständige Datensätze (Pflichtfelder?)
- Versionen (welche?)

*„Poor data quality can be worse than missing data because it can waste resources and lead to faulty ideas and solutions, or at minimum challenges trust in the results and implications drawn from the data. Improvement in data quality can thus have significant benefits. (Curating for Quality P5)“*

# Gefahrenfaktoren für Forschungsdaten

- Wegfall der Verantwortlichen
- Daten und Backups „gehen verloren“
- Sicherheitsfragen
  - nicht aktualisierte Software-Versionen
  - Unzureichend gesicherte Zugänge (Passwörter)
- Technologische Weiterentwicklung
  - Dateiformate
  - Webanwendungen
- Änderung der Policies oder der Dienstleistungen der Universität
  - Wegfall von Service-Angeboten
  - Corporate Design



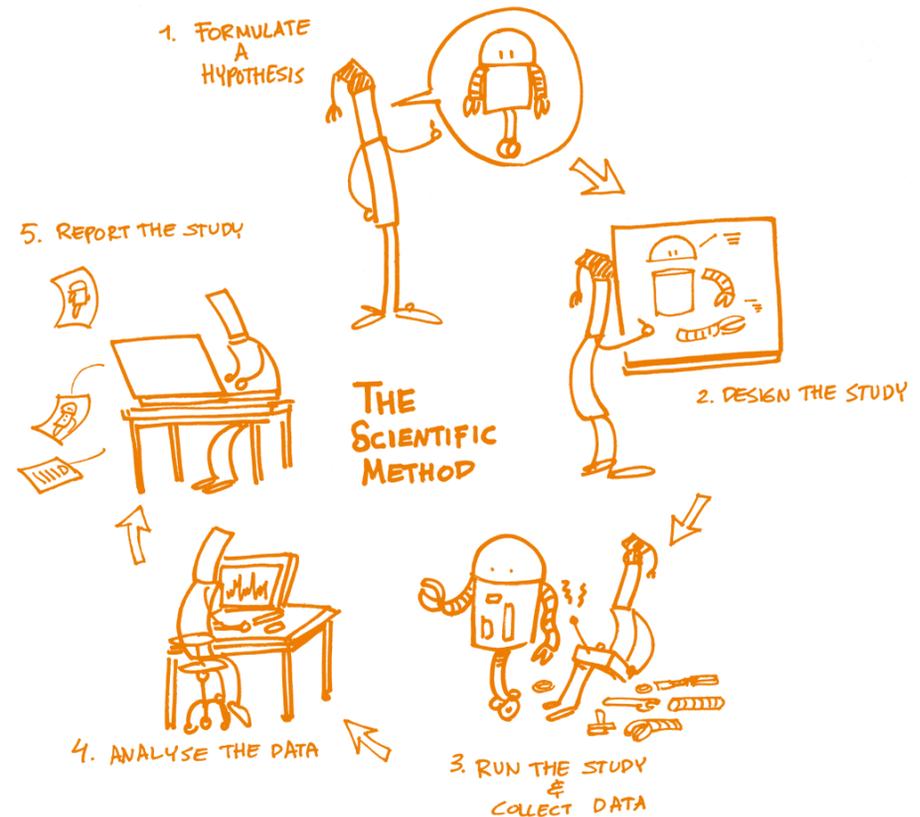
By Jørgen Stamp,  
<http://digitalbevaring.dk/>

# Datenmanagement

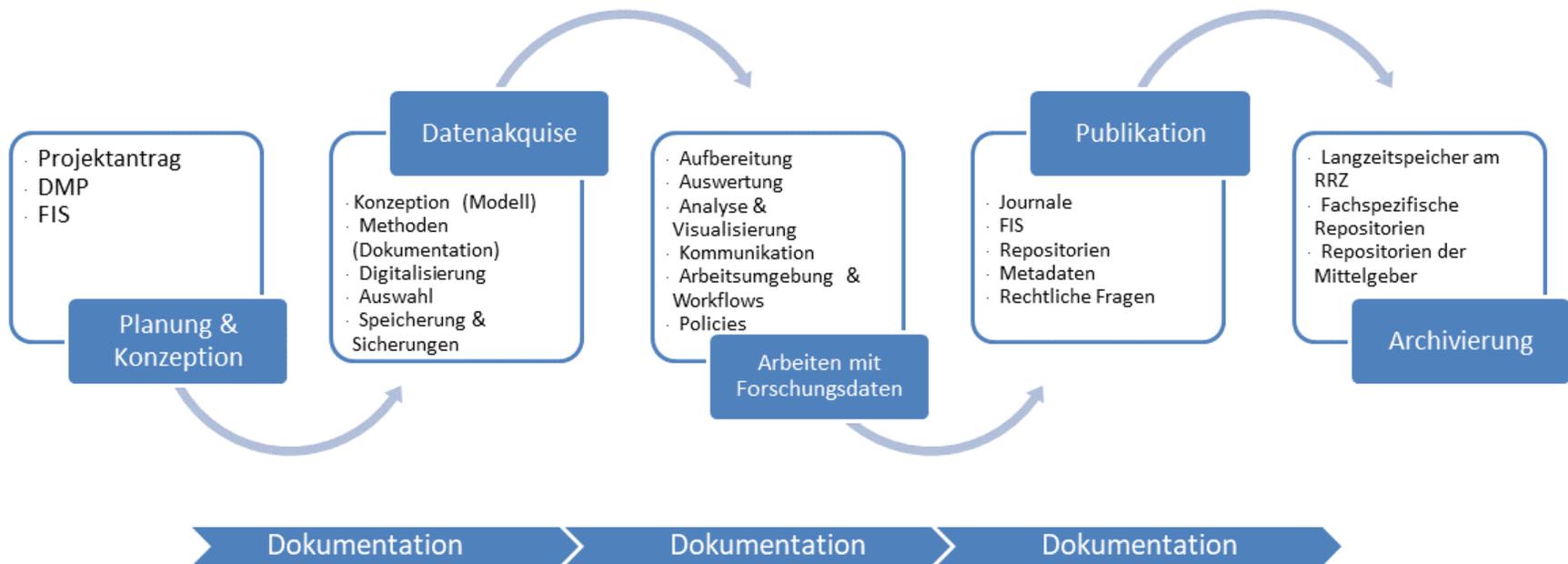
- Mit welchen Tools arbeiten Sie und Ihre Kollegen?
- In welchem Zusammenhang sind Sie mit Datenmanagementplänen in Berührung gekommen?
- Gibt es Instruktionen durch Betreuer?
- Gibt es Ratschläge für die Post-Doc-Zeit und Anträge?

# Übung DMP ohne Hilfsmittel

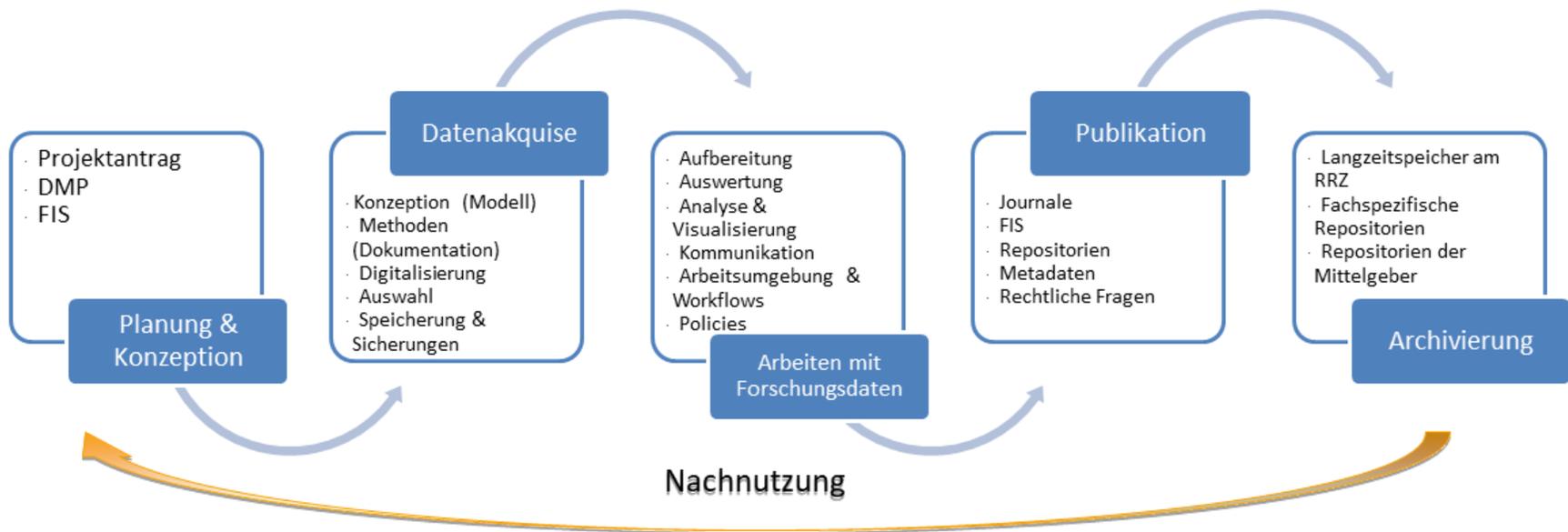
- Erstellen Sie in der Gruppe einen Datenlebenszyklus für ein (fiktives) Projekt
- Überlegen Sie und notieren Sie für die einzelnen Phasen Ihres Projektes, welche Schritte und Werkzeuge im Datenmanagement nötig sind (20min)
- Stellen Sie Ihr Poster kurz den anderen Seminarteilnehmern vor (5min)



## Phasen des Forschungsdatenmanagements



## Phasen des Forschungsdatenmanagements



# 3. Datenmanagementplan

# Datenmanagementplan – Definition

Ein Datenmanagementplan (DMP) bezeichnet ein Dokument zur Beschreibung des Lebenszyklus von Daten (insbesondere Forschungsdaten) von der Erhebung bis zur Archivierung, einschließlich aller Maßnahmen, die gewährleisten, dass die Daten verfügbar, nutzbar und nachvollziehbar bleiben.

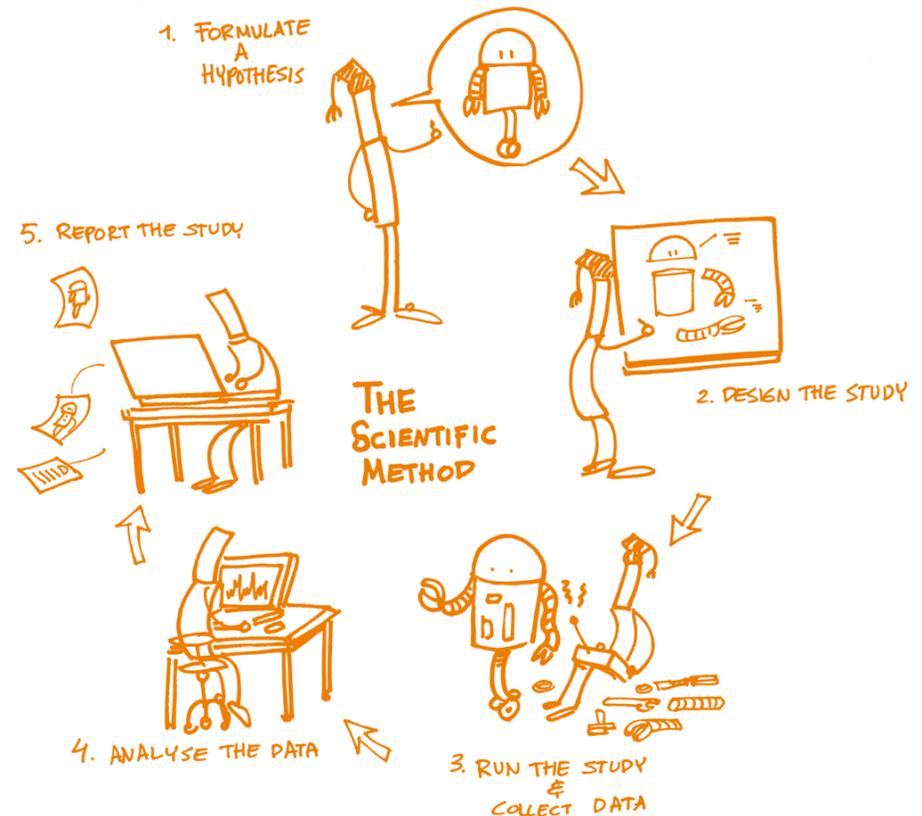


Tabelle 1: Vergleich der Anforderungen der Förderer bezüglich des Datenmanagementplans (Stand: 18.04.2018).

Förderer	Plan gefordert?	Abgabe bei Antrag?	Inhalt	Updates?
EC Horizon 2020	Datenmanagementplan	Nein, erster Plan innerhalb der ersten 6 Projektmonate	Inhalte des Horizon 2020 Template	Update, falls signifikante Änderungen auftreten sowie zum Projektende
DFG	Angaben zum Umgang mit Forschungsdaten	Ja	Inhalte der Leitlinie zum Umgang mit Forschungsdaten	Nein
BMBF	Plan manchmal erforderlich, abhängig vom Programm	Falls notwendig, ja	Inhalt hängt vom jeweiligen Programm ab	Nein
BMBF Bildungsforschung	Datenmanagementplan	Ja	Inhalte der Checkliste	Ja

Quelle: FDMentor

# Elemente eines DMP

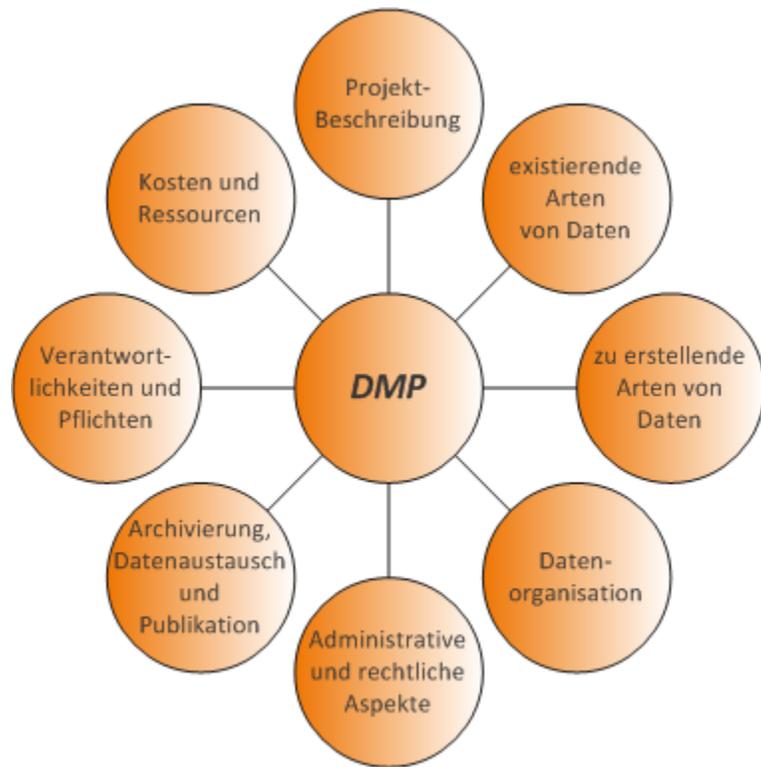


## 1. Projektbeschreibung

- Allgemeine Angaben zum Projektvorhaben, wie z.B. Ziele, Projektträger und Laufzeit
- relevante Policies

Quelle: Helbig, Kerstin. (2015). Hinweise und Checkliste zur Erstellung eines Datenmanagementplans. Version 1.0. Humboldt-Universität zu Berlin.

# Elemente eines DMP



## 2. Existierende Arten von Daten

- Beschreibung der existierenden Daten, die für das Projekt nachgenutzt werden können und ihre Integration in das Projektvorhaben.
- Begründung, wenn keine Daten nachgenutzt werden

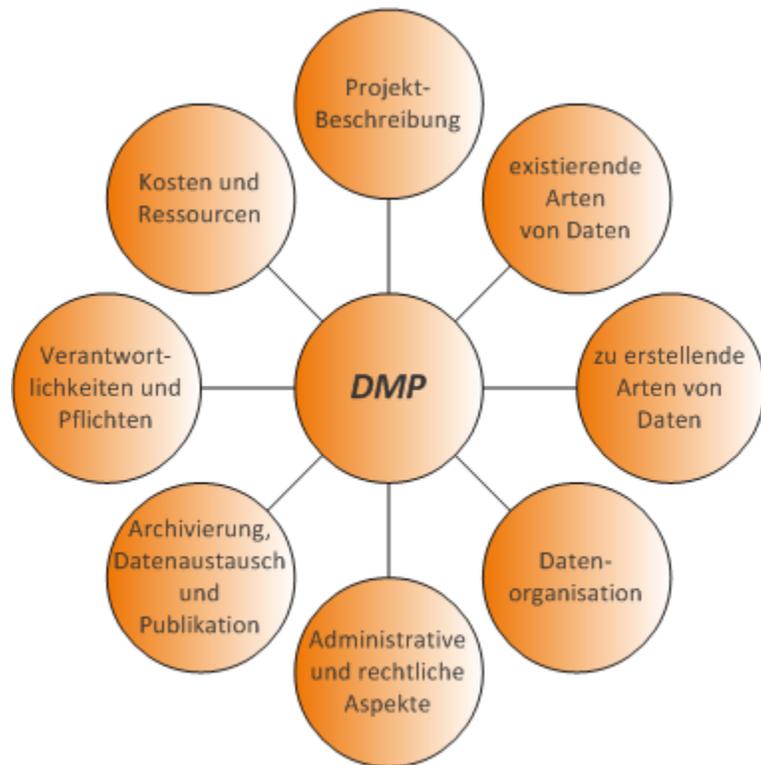
# Elemente eines DMP



## 3. Zu erstellende Arten von Daten

- Angaben zu Datenarten und Formaten
- geschätzte Datenmenge
- Informationen zum Prozess der Datengenerierung und Qualitätssicherung (wie z.B. mehrfache Stichprobenziehung, Validierung, Data Peer Review)
- Relevanz der Nachnutzung durch andere

# Elemente eines DMP



## 4. Datenorganisation

- Angaben zur einheitlichen Regelung der Datenverwaltung innerhalb des Projektes
- Dateispeicherung
- Dateibenennung, Ordnerstruktur
- Synchronisation, Versionierung
- kollaboratives Arbeiten
- Dokumentation
- Erstellung von Metadaten (Standards?)

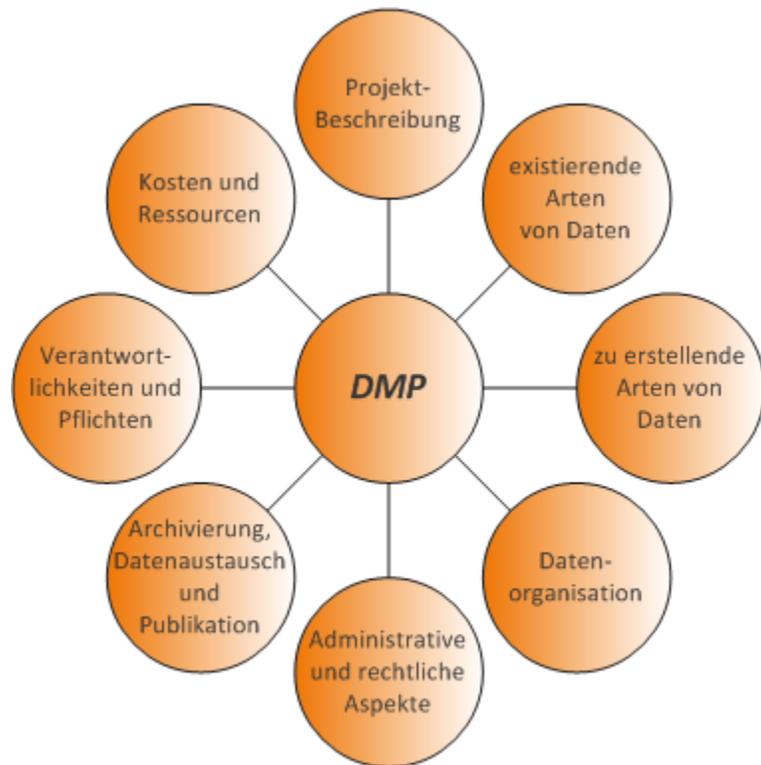
# Elemente eines DMP



## 5. Administrative und rechtliche Aspekte

- Urheberrecht / Rechteeigentümer der Daten
- Zugriff und Nutzung
- Datenschutz
- Datensicherung/Verschlüsselung und Backups

# Elemente eines DMP



## 6. Archivierung, Datenaustausch und Publikation

- Welche Datentypen werden veröffentlicht?
- Welches Repository/Archiv?
- Zugangsmöglichkeiten (frei, eingeschränkt etc.), Lizenz
- zeitlicher und finanzieller Rahmen der Datenaufbereitung, -bereitstellung bzw. -archivierung
- Gründe, wenn keine Publikation vorgesehen ist
- Angaben zur Verfahrensweise nach Ablauf der Aufbewahrungsfrist

# Elemente eines DMP



## 7. Verantwortlichkeiten und Pflichten

Wie sind die Verantwortlichkeiten für das Datenmanagement innerhalb des Projekts festgelegt und verteilt?

# Elemente eines DMP



## 8. Kosten und Ressourcen

Angaben zu den Kosten und dem Personalaufwand für die Einhaltung des Datenmanagementplans, sowie laufender Kosten für Datenkuration, Erstellung von Metadaten, Archivierung, etc.

# FDM Budget planen

- Personalkosten (DatenmanagerIn, DatenaufbereiterIn, DatenkuratorIn)
- Infrastruktur (Ausstattung, Software, etc.)
- Dienstleistungskosten (Daten, Langzeitarchivierung, Publikation, Schulungen)

# Datenmanagementpläne nach RDMO (Wunschvorstellung...)

- Optimierung des Forschungsdatenmanagements vor Projektbeginn
- Leitfaden für Umgang mit Forschungsdaten über den gesamten Projektverlauf
- Planung des Verbleibs der Forschungsdaten nach Projektende
  - Veröffentlichung
  - Langzeitarchivierung
- Erhöhung von Qualität und Effizienz von wissenschaftlicher Arbeit
- Mehr als nur Anforderung der Förderorganisation
- Im Idealfall Gewinn für den Forschenden und die Institution

# Übung - RDMO (20min)

Research Data Management Organizer ist ein Tool zur Erstellung von Datenmanagementplänen mit verschiedenen Templates

<https://dmp.fdm.uni-hamburg.de/>

- Erstellen Sie mit Ihrem Nachbarn einen DMP für Ihr Projekt (oder ein fiktives Projekt) (20 min)
- Beantworten Sie die Fragen des RDMO Katalogs
- Fügen Sie Ihre Gruppenmitglieder als Mitglied zum Projekt hinzu
- Verteilen Sie Aufgaben
- Exportieren Sie Ihren DMP im Template für Horizon 2020

# 4. Policies und Anforderungen der Förderer

# Policy

- Interne Leit- bzw. Richtlinie, die formal durch das Unternehmen dokumentiert und über ihr Management verantwortet wird

# Interdisziplinäre Policies:

## Deutsche Forschungsgemeinschaft

DFG: Empfehlungen zur Sicherung guter wissenschaftlicher Praxis (Sept. 2019)



"Aus Gründen der Nachvollziehbarkeit, Anschlussfähigkeit der Forschung und Nachnutzbarkeit hinterlegen Wissenschaftler\*innen, wann immer möglich, die der Publikation zugrunde liegenden Forschungsdaten und zentralen Materialien – den **FAIR-Prinzipien** folgend – zugänglich in anerkannten Archiven und Repositorien."

„[...] werden die zugrunde liegenden Forschungsdaten (i. d. R. Rohdaten) – abhängig vom jeweiligen Fachgebiet – i. d. R. für einen Zeitraum von **zehn Jahren** zugänglich und nachvollziehbar in der Einrichtung, wo sie entstanden sind, oder in standortübergreifenden Repositorien aufbewahrt.“

Zielgruppe: Antragsteller bei der DFG. Verbindlich (keine Bewilligung ohne Stellungnahme bzw. DMP)

# Interdisziplinäre Policies:

## Europäische Kommission

ERC (2016): Guidelines on FAIR Data Management in Horizon 2020

“In general terms, your research data should be 'FAIR', that is findable, accessible, interoperable and re-usable.”

“as open as possible, as closed as necessary”

Zielgruppe: Antragsteller für Horizon 2020. Verbindlich (keine Bewilligung ohne Stellungnahme, Verpflichtung zu DMP zu Beginn und Aktualisierungen während der Projektlaufzeit).

Sehr konkrete Vorschläge zur Umsetzung.



# Disziplinäre Policies: Good Laboratory Practice (GLP)

OECD (1997): OECD Principles of Good Laboratory Practice

“...ensure that after completion (including termination) of the study, the study plan, the final report, raw data and supporting material are archived.

“Any change in the raw data should be made so as not to obscure the previous entry, should indicate the reason for change and should be dated and signed or initialled by the individual making the change.”

Zielgruppe: Forschungslabore; in vielen Ländern gesetzlich vorgeschrieben für nichtklinische experimentelle Prüfungen von Stoffen (REACH)



# Disziplinäre Policies:

Bermuda Principles (Human Genome Project, 1996)

“All human genomic sequence data generated by centers funded for large-scale human sequencing should be freely available and in the public domain to encourage research and development and to maximize the benefit to society.”

Selbstverpflichtung, mit Forschungsförderern abgestimmt und von diesen eingefordert.



# Journal Policies

E.g. Nature, Policy from 2017

**nature**

“Supporting data must be made available to editors and peer reviewers at the time of submission for the purposes of evaluating the manuscript. All manuscripts reporting original research published in Nature journals must include a data availability statement. Data availability statements should provide a statement about the availability of the **minimal dataset that would be necessary to interpret, replicate and build upon the methods or findings reported in the article** . ... The preferred way to share large data sets is via public repositories. Details about how to share some specific data sets can be found in the sections below.”

verpflichtend für alle, die einen Artikel in den Nature Journals veröffentlichen wollen

# Institutionelle Policies:

## Universität Hamburg

UHH: Satzung zur Sicherung Guter wissenschaftlicher Praxis und zur Vermeidung wissenschaftlichen Fehlverhaltens an der Universität Hamburg.



“Primärdaten als Grundlagen für Veröffentlichungen sind auf haltbaren und gesicherten Trägern in der Institution, in der sie entstanden sind, für zehn Jahre aufzubewahren, sofern nicht spezielle Regelungen eine längere Aufbewahrung vorsehen.”

DFG-Richtlinien für nicht-DFG-Projekte an der UHH. Betrachtet auch alle weiteren Aspekte. Unterstützung durch eine Geschäftsstelle, ein Ombudskollegium und einen Expertenausschuss.

# Open Access Policies der UHH

## Publikationen

"Für wissenschaftliche Veröffentlichungen empfiehlt die Universität Hamburg eine direkte Open-Access-Publikation („**goldener Weg**“), entweder durch ein an den Verlag zu zahlendes Entgelt oder unentgeltlich auf dem „Institutional Repository“ der Universität Hamburg."

"Bei Publikationen in kostenpflichtigen Verlagen sollte ein nicht-exklusives Nutzungsrecht ausgehandelt werden, um eine sofortige Zweitpublikation („**grüner Weg**“, „Self-archiving“) zu ermöglichen."

## Daten

"Die Universität Hamburg bekennt sich zur dauerhaften Erhaltung sowie zur zuverlässigen Bereitstellung und Nutzbarmachung von Publikationsdaten („Open Access“ im engeren Sinne), Forschungsdaten („Open Data“) sowie von Lehr- und Lerndaten („Open Educational Resources“) (...). Die Universität Hamburg stellt die zum Umgang mit wissenschaftlichen Daten notwendigen Dienstleistungen und Infrastrukturen bereit (...)."

# 5. Rechtliches und Open Access

<https://forschungsdaten.blogs.uni-hamburg.de/>

# Rechtliche Hürden vs. Open Data

## 1. Datenschutzgesetze

- Für Universitäten und andere öffentliche Stellen der Länder gelten die jeweiligen Landesdatenschutzgesetze und die DSGVO (nicht das BDSG!)
- nur bei personenbezogenen Daten
- Verarbeitung personenbezogener Daten: "Verbot mit Erlaubnisvorbehalt"
- Grundsätze der Zweckbindung und Datenminimierung

DSGVO: "Personenbezogenen Daten sind alle Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person beziehen; als identifizierbar wird eine natürliche Person angesehen, die direkt oder indirekt, insbesondere mittels Zuordnung einer Kennung wie einem Namen, zu einer Kennnummer, zu Standortdaten, zu einer Online-Kennung oder zu einem oder mehreren besonderen Merkmalen identifiziert werden kann, die Ausdruck der physischen, physiologischen, genetischen, psychischen, wirtschaftlichen, kulturellen oder sozialen Identität dieser natürlichen Person sind."

# Anonymisierung vs. Pseudonomisierung

- Durch Anonymisierung können aus personenbezogenen Daten Daten ohne Personenbezug werden
- Pseudonomisierung: Bezug zur Person, deren Daten betroffen sind, kann wiederhergestellt werden (z.B. durch einen getrennt aufbewahrten Schlüssel)
- Durch Löschung des Schlüssels kann eine Anonymisierung erfolgen
- pseudonomisierte Daten sind immer noch personenbezogene Daten



# Erlaubnistatbestände für Veröffentlichung

1. Einwilligungen der Personen liegen vor und erfüllen die Voraussetzungen (DSGVO):
  - Freiwilligkeit der Zustimmung
  - "informed consent"
  - Zweck und Art der Verarbeitung der Daten muss klar definiert sein
2. Veröffentlichung ist für die "Darstellung von Forschungsergebnissen über Ereignisse der Zeitgeschichte unerlässlich" (§ 11 Abs. 3 HmbDSG)

# Aufklärungsformular und Einwilligungserklärung

## 3 Beispielformulierungen für Aufklärungsformular und Einwilligungserklärung

### 3.1 Aufklärungsformular

„Das Forschungsvorhaben ..... hat zum Ziel ..... Ihre Aufgabe besteht darin, ..... Dabei können mitunter Situationen entstehen, die Sie als weniger angenehm empfinden...

Es werden auch persönliche Daten über Sie erhoben. Alle Informationen, die wir im Rahmen unserer Untersuchungen sammeln, werden vertraulich behandelt und werden nur von Mitarbeitern\_innen unseres Forschungsteams im Rahmen der laufenden Studie verwendet. Die Weitergabe, Speicherung und Auswertung dieser studienbezogenen Daten erfolgt nach gesetzlichen Bestimmungen ohne Namensnennung (das heißt, Ihre Daten werden kodiert). Sollten Sie sich nach der Untersuchung gegen eine Teilnahme und damit für eine Löschung der erhobenen Daten entscheiden, ist dies aus organisatorischen Gründen nur bis zum Abschluss der Studie möglich. Mit dem Abschluss der Studie sind wir verpflichtet, auch unsere Kodierliste der Personennamen zu löschen, damit ist die Zuordnung der erhobenen Daten zu individuellen Studienteilnehmern\_innen anschließend nicht mehr möglich.

Verantwortlich für die Analyse und Speicherung Ihrer Daten ist ....., Institut ....., Universität Hamburg.

Die Teilnahme an dem Forschungsvorhaben erfolgt auf gänzlich freiwilliger Basis. Sie können jederzeit und ohne Angabe von Gründen Ihr Einverständnis zur Teilnahme zurücknehmen, ohne dass Ihnen hieraus Nachteile entstehen.“

### 3.2 Einwilligungformular

„Ich bin über Wesen, Bedeutung und Tragweite der geplanten Untersuchungen aufgeklärt worden. Dazu wurde mir ein Exemplar der Teilnehmerinformation ausgehändigt. Zum Ablauf und den möglichen Risiken konnte ich Fragen stellen. Die Informationen, die mir erteilt wurden, habe ich inhaltlich verstanden.

# Rechtliche Hürden vs. Open Data

## 2. Urheberrecht

### Wem gehören Daten?

- Es gibt kein Eigentum an Daten
- urheberrechtlich schützenswert ist erst die Form der Darstellung
- Schutzvoraussetzung für Datenbanken ist eine "wesentliche Investition"
- Urheber ist derjenige, der diese Investition getätigt hat (Arbeitgeber (UHH) oder Forschende?)
- "Hochschullehrerprivileg" (Achtung bei Drittmittelgebern)
- "angereicherte Datensätze" (z.B. Bilder, Diagramme, Texte): Urheberrecht gilt



# Rechtliche Hürden vs. Open Data

## 2. Urheberrecht - Nutzungsrechte

- Dem Urheber stehen zunächst alle Urheberrechte/Leistungsschutzrechte zu (Ausnahme Software)
- Eine Datenbank, die das Kriterium der wesentlichen Investition erfüllt, ist davor geschützt, dass „wesentliche Teile“ der hierin enthaltenen Daten ohne Erlaubnis von Dritten entnommen (zum Beispiel kopiert) werden
- gemeinsame Urheberrechte mit Co-Autor\*innen
- (exklusive) Nutzungsrechte können durch Vertrag (auch mündlich) übertragen werden (Arbeitgeber, Verlag)
- "Zweitveröffentlichungsrecht", Preprints
- "Schrankenbestimmungen" (§§ 44a–63a UrhG) z.B. Zitatrecht
- Rechte an Datenbanken erlöschen 15 Jahre nach deren Veröffentlichung oder Herstellung (§ 87d UrhG)

# Rechtliche Hürden vs. Open Data

Datenschutz (Sicherheitsrelevante Daten, Artenschutz, Schutz von Kulturgütern)

- Es gilt das Gebot der Datensparsamkeit
- ethische Aspekte/ sensible Daten: im Zweifel Ethikkommission befragen

UH **Universität Hamburg**  
DER FORSCHUNG | DER LEHRE | DER BILDUNG

CeNAK

CeNAK Sammlungsportal Detailsuche Sammlungen Suchbegriff

UH Fundus / CeNAK Über CeNAK

Sammlung: Ichthyologie

Zmh Catalog No: ZMH 375  
Family: Cyprinidae  
Genus: *Rasbora*  
Species: *dorsiocellata*  
Author: Duncker, 1904  
Type: syntype  
Locality: Kuala Jelai, Malaysia  
Date: 10.02.1902



**Zoologisches Museum Hamburg**  
ZMH 375  
*Rasbora* syntype  
*dorsiocellata* DUNCKER, 1904

# Lizenzen

- Lizenzen regeln die Bedingungen der Nachnutzung von veröffentlichten Daten
- sog. offene Lizenzen stehen allen zum freien Einsatz zur Verfügung und müssen nur verlinkt werden
- Voraussetzung ist, dass Sie der Rechteinhaber sind
- Auswahl der Lizenz von der Art der Daten abhängig:
  - z.B. Creative Commons (CC) - Lizenzen für Artikel, Monografien, Bilder etc.
  - Open-Database-License (ODbL) für Datenbanken, oder CC ab Version 4
  - General Public Licence (GNU) für Software
- wird keine Lizenz vergeben gilt das strengere Urheberrecht

## Creative Commons Lizenzen

Namensnennung			Namensnennung Nicht kommerziell Share Alike
Namensnennung Share Alike			Namensnennung No Derivates
Namensnennung Nicht kommerziell			Namensnennung Nicht kommerziell No Derivates

nur CC-BY und CC0 entsprechen den FAIR-Prinzipien

# Ansprechpartner Rechtliches

- Datenschutzbeauftragter der UHH: Dirk-Andreas Hengst, [datenschutz@uni-hamburg.de](mailto:datenschutz@uni-hamburg.de)
- Multimedia Kontor Hamburg <https://www.hh-datenschutz.de/>
- Literatur: Leitfaden “Rechtsfragen bei Open Science” von Till Kreutzer und Henning Lahmann <https://blogs.sub.uni-hamburg.de/hup/products-page/publikationen/169/>

# Quiz

[www.kahoot.it](http://www.kahoot.it)

Game PIN eingeben

# 6. Forschungsdaten akquirieren

# Datenakquise: Vorbereitung

1. Methoden und Modelle zur Datenverarbeitung festlegen
2. Datenmanagement planen (Formate, Speicherorte etc.)
3. ggfs. bereits vorhandene Daten lokalisieren und analysieren
4. rechtliche Aspekte der Nutzung und Veröffentlichung klären

# Arten der Akquise von Forschungsdaten

Art der Akquise	Vorteile	Nachteile
Erhebung / Erfassung eigener Daten	volle Kontrolle über Art und Umfang	verbunden mit erheblichen Kosten und Risiken
Nachnutzung vorhandener Daten	Kosten / Aufwand der Erhebung wird überwiegend von anderen getragen potenziell größere Reichweite (Metastudien) standardisierte Methoden	Umfang und Format der Daten ist vorgegeben u.U. erheblicher Einarbeitungsaufwand

# Mögliche Anlaufstellen

- Fachdatenbanken
  - disziplinäre Forschungsdatenrepositorien
  - Portale oder Index-Services
  - Forschungsdatenzentren
  - Institutionen, z.B. Behörden...
- WWW
  - Webseiten (Inhalt, Domäne, Sprache)
  - Social media (twitter, facebook, blogs...)...
- Assoziierte Projekte
  - Datenträger
  - Zugang zu Shares...

# Suche nach passenden Repositorien

- [Re3data.org](https://re3data.org/) (Karlsruhe Institute of Technology)
- [OpenDOAR UK](https://open.doar.uk/)
- [Research Pipeline](https://researchpipeline.org/) Mediawiki mit Links zu Open Data
- [Repository Finder](https://re3data.org/repository-finder/) Suche in Re3Data nach Repositorien, die den Enabling FAIR Data Projekt Prinzipien entsprechen

# Suche nach passenden *Daten* in Repositorien

- [DataCite.org](#) NGO / DOI provider
- [EU Open Data Portal](#)
- [B2FIND](#) Recherche in EUDAT FD Sammlungen
- [Research Data Australia](#)
- ....

# Nachnutzung von Daten: Voraussetzungen

- Qualität der Daten muss für das Forschungsvorhaben ausreichend sein
- Provenienz und Rechte müssen geklärt sein
- Vollständigkeit muss ausreichend sein für das Projektziel
- Interpretierbarkeit (Dokumentation vorhanden?)

# Datennachnutzung

## Übung (Excel-Tabelle-Datenqualität)

### **Provenienz/ Dokumentation**

---

? Worum geht es bei den Daten

---

? Woher stammen die Daten

---

? Wann wurden sie erhoben

---

? In welchem Format liegen die Daten vor

---

? Wurden Änderungen versioniert

### **Bewerten Sie die Datenqualität:**

---

Bewerten Sie die Datenqualität:

---

? Ist der Datensatz vollständig

---

? Sind die Daten konsistent

---

? Sind die Daten interpretierbar

---

? Wurden ggfs. Lücken gekennzeichnet

# Mögliche Quellen für eigene Erhebungen

- Umfragen
- Aufzeichnungen
- Materialsammlung / Digitalisierung ...
-

# Datenakquise: Daten erfassen

1. Methoden und Modelle zur Datenverarbeitung festlegen
2. ggfs. bereits vorhandene Daten lokalisieren
3. Datenmanagement planen (Formate, Speicherorte etc.)
4. rechtliche Aspekte der Nutzung und Veröffentlichung klären
5. Daten erfassen / erheben
6. Metadaten erstellen / Dokumentation

# Modellierung - Datenqualität

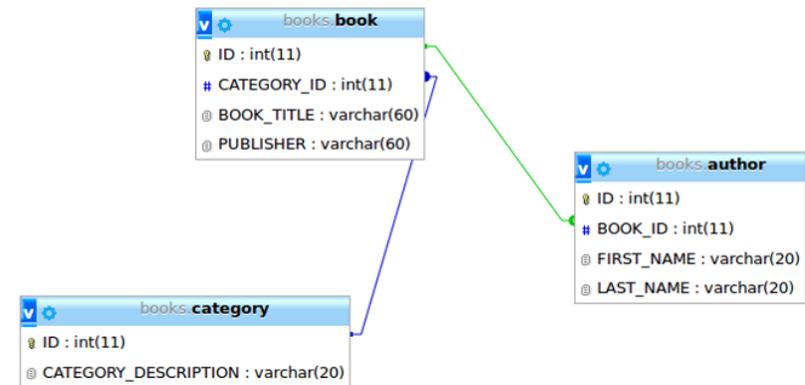
- Erkundigen Sie sich nach möglichen Standards oder Best Practices in Ihrem Feld
- unterstützen Sie die Eingabe von validen Daten durch entsprechende Eingabehilfen
- gestalten Sie die Eingabe nur so flexibel wie nötig (Validierung)

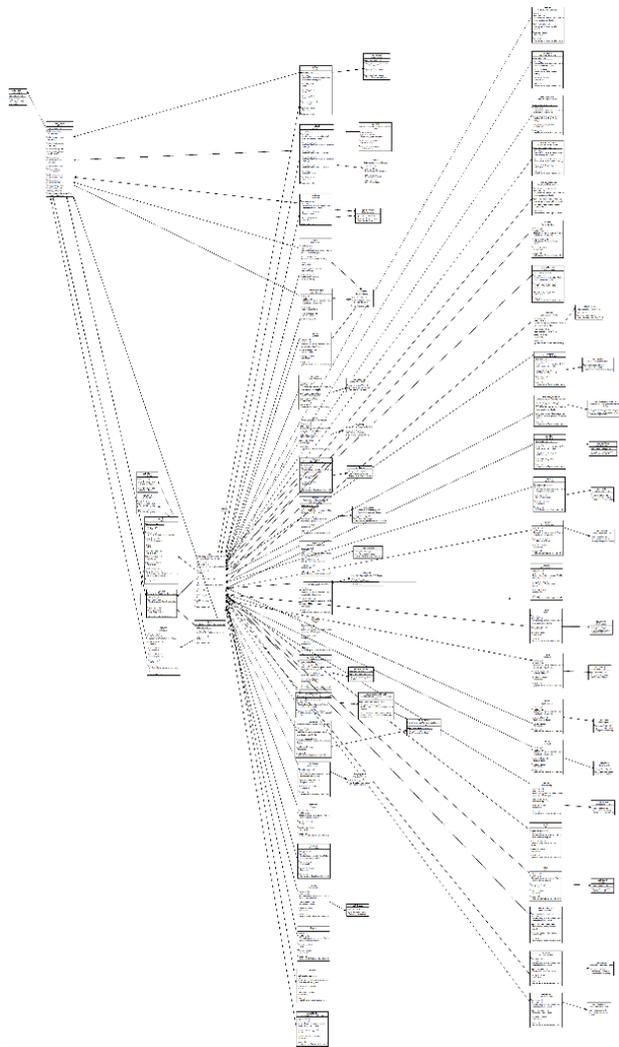
# Fokus Datenqualität

- Wertebereiche / Wertelisten helfen bei der Datenqualität
- kontrollierte Vokabulare und Ontologien bzw. Thesauri
- vordefinierte Wertebereiche, Format-vorgaben (z.B. Datum YYYY-MM-DD)
- Validierung (bei Eingabe/Import)
- Aussagekräftige Bezeichnungen
- Überprüfung / Peer Review / 4-Augen-Prinzip

# Modellierung von Daten

- Formale Abbildung der Objekte mittels Attributen und Beziehungen
- Eindeutige Spezifikation für die Verwaltung in einem Informationssystem
- Datenmodelle überleben in der Regel damit verbundene Funktionen und Prozesse
- Klare Trennung von Daten und Anwendung wichtig
- erkundigen Sie sich nach Standards in Ihrem Feld





# Daten dokumentieren

- Kontext: Projekthistorie, Absicht/ Zielsetzung, Hypothesen ...
- Methoden: Sampling, Umstände der Erhebung, technische Rahmenbedingungen ...
- Datenstrukturen, Beziehungen zwischen Objekten
- Wertebereiche, Qualitätskriterien, Gültigkeit
- Änderungen (!) im Projektverlauf, Versionierung
- Rechtliche Fragen / Zugriff

*Publikationen ersetzen nicht die Dokumentation der Daten*

# Technische Dokumentation

- Namen, Bezeichnungen für Variablen und ihre Werte
- Erklärungen für Codes / Klassifikationsschemata
- Erläuterung der verwendeten Terminologie, ggfs. Definitionen
- Kodierung fehlender Werte / Gründe für fehlende Werte
- Abgeleitete Daten, verwendete Algorithmen, Gewichtungen..
- ..

The image shows a screenshot of a data entry form. The form has several fields with labels on the left and input areas on the right. The labels are: Ehepartner\*in, Religion, Emigrantenstatus, Stationen / Orte, Tätigkeit, and Genre. The input areas are: a text box with a dashed border, a text box containing 'im Feld Ehepartner', a dropdown menu, a dropdown menu with a blue highlight, a list of characters (0, 1, 2, D, E, I, K, N, S, T, U, V, Z), and a text box. The dropdown menu for 'Emigrantenstatus' is open, showing a list of characters: 0, 1, 2, D, E, I, K, N, S, T, U, V, Z. The dropdown menu for 'Religion' is also visible but not open.

# ALGORITHMS BY COMPLEXITY

MORE COMPLEX →

LEFTPAD QUICKSORT

GIT  
MERGE

SELF-  
DRIVING  
CAR

GOOGLE  
SEARCH  
BACKEND

SPRAWLING EXCEL SPREADSHEET  
BUILT UP OVER 20 YEARS BY A  
CHURCH GROUP IN NEBRASKA TO  
COORDINATE THEIR SCHEDULING

--- Pause ---

# Daten speichern: Dateiformate

- = die Art und Weise, wie Information digital gespeichert ist
- Dateiendungen (z.B. *.txt*, *.docx*, *.jpg*) geben Hinweise auf Dateiformate und somit die darin enthaltenen Daten

*Win und Mac Systeme zeigen Datei-Endungen standardmäßig nicht an. Die Anzeige lässt sich über die Systemeinstellung aktivieren.*

- Der Internet-Media-Type / MIME (Multipurpose Internet Mail Extension) sorgt im WWW dafür, dass den Dateiformaten die richtige Anwendung zugeordnet wird

# MIME-Typen und File-Extensions

MIME-Typ	Dateiendung(en)	Bedeutung
application/acad	*.dwg	AutoCAD-Dateien (nach NCSA)
application/applefile		AppleFile-Dateien
application/astound	*.asd *.asn	Astound-Dateien
application/dsptype	*.tsp	TSP-Dateien
application/dxf	*.dxf	AutoCAD-Dateien (nach CERN)
application/force-download	*.reg	Registrierungsdateien
application/futuresplash	*.spl	Flash Futuresplash-Dateien
application/gzip	*.gz	GNU Zip-Dateien
application/javascript	*.js	serverseitige JavaScript-Dateien
application/json	*.json	enthält einen String in <a href="#">JavaScript-Objekt-Notation</a>
application/listenup	*.ptlk	Listenup-Dateien
application/mac-binhex40	*.hqx	Macintosh Binärdateien
application/mbedlet	*.mbd	Mbedlet-Dateien
application/mif	*.mif	FrameMaker Interchange Format Dateien
application/msexcel	*.xls *.xla	Microsoft Excel Dateien
application/mshelp	*.hlp *.chm	Microsoft Windows Hilfe Dateien
application/mspowerpoint	*.ppt *.ppz *.pps *.pot	Microsoft Powerpoint Dateien
application/msword	*.doc *.dot	Microsoft Word Dateien
application/octet-stream	*.bin *.file *.com *.class *.ini	Nicht näher spezifizierte Daten, z.B. ausführbare Dateien
application/oda	*.oda	Oda-Dateien

# Proprietäre Datei-Formate

- sind meist abhängig von
  - spezifischer Software
  - entsprechenden Software-Lizenzen
  - spezifischen Plattformen / Betriebssystemen
  - sind gefährdet durch 'Software Obsolescence'
  - rasche technologische Weiterentwicklung
  - Kompatibilität zwischen Versionen nicht gewährleistet
  - Marktabhängigkeit
  - Konvertierung / Export in offene Formate meist verlustbehaftet



# Dateiformate – offene Formate

Für die Langzeitarchivierung sollten Dateien

- unverschlüsselt sein
- nicht komprimiert sein
- nicht firmeneigen/patentiert sein
- einen offenen/dokumentierten Standard nutzen

# Tabelle Dateitypen

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

# Beispiel CSV

```
ID, Sammlungsname, Band, Röhrchen, Familie, Name lateinisch und Autor, Name deutsch, Name englisch, Date
1, Bredemann & Nieser, 01, 01, Fabaceae, Anthyllis vulneraria L., Wundklee, woundwort, 0001_Anthyllis_vu
2, Bredemann & Nieser, 01, 02, Fabaceae, Astragalus boeticus L., Kaffee-Traganth, swedish coffee, 0002_A
3, Bredemann & Nieser, 01, 03, Fabaceae, Astragalus falcatus Lam., Sicheltraganth, russian milk vetch, 0
4, Bredemann & Nieser, 01, 04, Fabaceae, Astragalus glycyphyllos L., Süßer Traganth, sweet milk vetch, 00
5, Bredemann & Nieser, 01, 05, Fabaceae, Astragalus hamosus L., Angelstragel, yellow milk vetch, 0005_As
6, Bredemann & Nieser, 01, 05a, Fabaceae, Cajanus cajan (L.) Millsp., Taubenerbse, pigeonpea, 0006_Cajan
7, Bredemann & Nieser, 01, 06, Fabaceae, Senna tora (L.) Roxb., Gemüse-Kassie, sicklepod, 0007_Senna_tor
8, Bredemann & Nieser, 01, 07, Fabaceae, Cicer arietinum L., Kichererbse, chickpea, 0008_Cicer_arietinum
9, Bredemann & Nieser, 01, 08, Fabaceae, Coronilla scorpioides (L.) W.D.J.Koch, Skorpionspeitsche, annu
10, Bredemann & Nieser, 01, 09, Fabaceae, Securigera varia (L.) Lassen, Bunte Kronwicke, Crown-vetch, 00
11, Bredemann & Nieser, 01, 10, Fabaceae, Cyamopsis tetragonoloba (L.) Taub., Guar, Guar, 0011_Cyamopsis
12, Bredemann & Nieser, 01, 10a, Fabaceae, Macrotyloma uniflorum (Lam.) Verdc. var. uniflorum, Pferdeh
13, Bredemann & Nieser, 01, 11, Fabaceae, Galega officinalis L., Echte Geisraute, goat's-rue, 0013_Galeg
14, Bredemann & Nieser, 01, 12, Fabaceae, Genista anglica L., Englischer Ginster, petty whin, 0014_Genis
15, Bredemann & Nieser, 01, 13, Fabaceae, Genista germanica L., Deutscher Ginster, german greenweed, 001
16, Bredemann & Nieser, 01, 14, Fabaceae, Genista sagittalis L., Flügelginster, winged greenweed, 0016_G
17, Bredemann & Nieser, 01, 15, Fabaceae, Genista tinctoria L., Färber-Ginster, dyer's greenweed, 0017_G
18, Bredemann & Nieser, 01, 16, Fabaceae, Glycine max (L.) Merr., "Sojabohne, Soja", "soybean, Soya", 00
19, Bredemann & Nieser, 01, 17, Fabaceae, Hedysarum coronarium L., Kronen-Süßklee, French-honeysuckle, 0
```

## Beispiel XML

```
<?xml version="1.0" encoding="utf-8"?>
<mycoreobject xmlns:xsi = "http://www.w3.org/2001/XMLSchema-instance"
  xmlns:xlink = "http://www.w3.org/1999/xlink"
  xsi:noNamespaceSchemaLocation = "datamodel-document.xsd"
  ID="ngmcp_document_4376" label="The 4376 NGMCP document">

  <!-- Strukturinformationen fuer MyCoRe-Interna -->
  <structure />

  <!-- Beschreibungsdaten zum Dokument -->
  <metadata>

    <!-- 01 - Titel -->
    <titles class="MCRMetaLangText">
      <title>
        A&#7779;&#7789;avarga
      </title>
    </titles>

    <!-- 02 - Material -->
    <materials class="MCRMetaClassification">
      <material classid="ngmcp_class_00000011" categid="XX" />
    </materials>

    <!-- 03 - Language -->
    <languages class="MCRMetaClassification">
      <language classid="ngmcp_class_00000012" categid="S" />
    </languages>

    <!-- 04 - Script -->
    <scripts class="MCRMetaClassification">
      <script classid="ngmcp_class_00000013" categid="D" />
    </scripts>

    <!-- 06 - Author -->
    <authors class="MCRMetalinkID" >
      <author xlink:type="locator" xlink:href="ngmcp_author_00000002"
        xlink:label="author" xlink:title=""/>
    </authors>

    <!-- 07 - Subject -->
    <subjects class="MCRMetaClassification" >
      <subject classid="ngmcp_class_00000014" categid="J" />
    </subjects>
    [...]
  </xml>
```

# Beispiel JSON

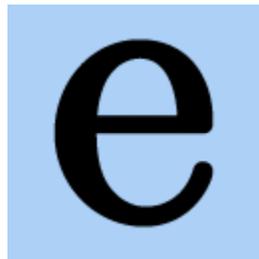
```
{
  "id": 50283,
  "catalogno": "ZMH 22258",
  "collection": "entomology",
  "created_at": "2018-08-16T13:35:45.955Z",
  "updated_at": "2018-09-21T06:04:51.480Z",
  "details": {
    "tax_class": "Insecta",
    "tax_order": "Coleoptera",
    "odesc_name": "Tefflus, meyerlei, Fabricius 1801",
    "tax_domain": "Holometabola",
    "tax_family": "Carabidae",
    "geo_locality": "Côte d'Ivoire, Dimbroko",
    "geo_continent": "Africa",
    "znh_catalog_no": "ZMH 22258",
    "identification_stage": "Imago",
    "identification_specimen": "pinned"
  },
  "pix": [
    "meyerlei_tefflus_zmh_22258_50mm_1x2_5.6f.jpg",
    "meyerlei_tefflus_zmh_22258_label.jpg"
  ]
}
```

# Datenformate - Empfehlungen

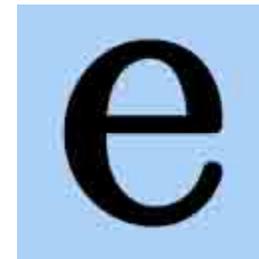
- Multimediale Aufnahmen
  - Hohe Qualität
  - Keine verlustbehaftete Kompression
- Texte
  - Unicode-basierte Kodierung
  - Offene und dokumentierte Formate
  - De Facto Standards (UTF-8, XML, TEI, JSON-LD, RDF...)
- Allgemein Empfehlungen
  - Keine proprietären oder exotischen Formate
  - Hohe Qualität
  - Kompression: verbreitete Formate wählen (.zip, .tar.gz, .tar.bz2)

# Verlustfreie versus verlustbehaftete Bildkompression

PNG: 1007 Byte



JPEG: 4273 Byte



Von Phrood , <https://commons.wikimedia.org/w/index.php?curid=14906788>

Alles zu spät?!

# Datenbereinigung

# Werkzeuge zur Bereinigung von Daten

- OpenRefine (ehemals GoogleRefine)
- Python Pandas
- Knime

The screenshot shows the OpenRefine web interface. The browser address bar displays the URL `127.0.0.1:3333/project?project=1586603301677`. The page title is "OpenRefine 2017 01 26 Botanik LokiSchmidt DB mit ID csv". The interface includes a "Facet / Filter" sidebar on the left with a "Using facets and filters" help box. The main area shows a table with 1124 rows. The columns are: All, ID, Sammlungsname, Band, Röhrchen, Familie, Name lateinisch und Autor, Name deutsch, Name englisch, Dateinamen in Reihenfolge, and Dateinamen ergänzende Bilder. A context menu is open over the "Röhrchen" column, listing options like Facet, Text filter, Edit cells, Edit column, Transpose, Sort..., View, and Reconcile. The table data includes columns for ID, Sammlungsname, Band, Name deutsch, Name englisch, and Dateinamen in Reihenfolge.

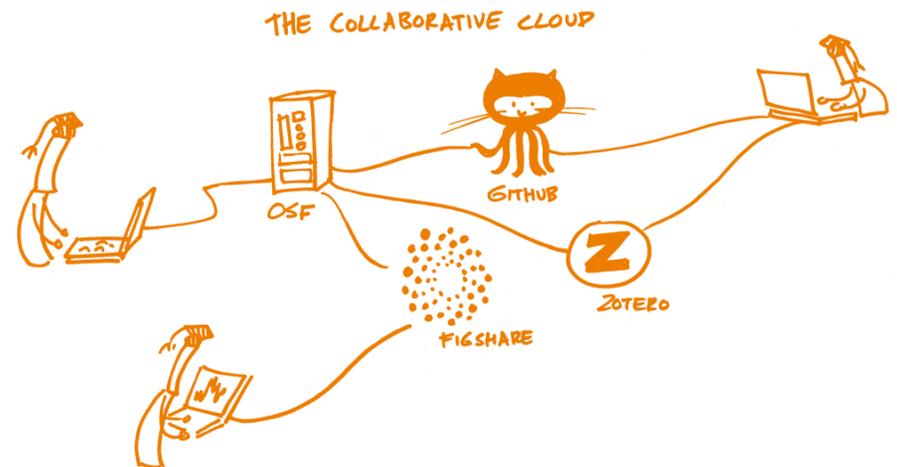
All	ID	Sammlungsname	Band	Röhrchen	Familie	Name lateinisch und Autor	Name deutsch	Name englisch	Dateinamen in Reihenfolge	Dateinamen ergänzende Bilder
1	1	Bredemann & Nieser	01				Wundklee	woundwort	0001_Anthyllis_vulneraria	
2	2	Bredemann & Nieser	01				Kaffee-Traganth	swedish coffee	0002_Astragalus_boeticus	
3	3	Bredemann & Nieser	01				Sicheltraganth	russian milk vetch	0003_Astragalus_falcatus	
4	4	Bredemann & Nieser	01				Süßer Tragant	sweet milk vetch	0004_Astragalus_glycyphylus	
5	5	Bredemann & Nieser	01				Angelstragel	yellow milk vetch	0005_Astragalus_hamosus	
6	6	Bredemann & Nieser	01				Taubenerbse	pigeonpea	0006_Cajanus_cajan	
7	7	Bredemann & Nieser	01				Gemüse-Kassie	sicklepod	0007_Senna_tora	
8	8	Bredemann & Nieser	01				Kichererbse	chickpea	0008_Cicer_arietinum	
9	9	Bredemann & Nieser	01		ae	Coronilla scorpioides (L.) W.D.J.Koch	Skorpionspeitsche	annual scorpion-vetch	0009_Coronilla_scorpioides	0953_Coronilla_scorpioides_Huelsen
10	10	Bredemann & Nieser	01		ae	Securigera varia (L.) Lassen	Bunte Kronwicke	Crown-vetch	0010_Securigera_varia	0954_Securigera_varia_Huelsen

# Herausforderungen

- Redundanz (Dopplungen von Inhalten / Datenstrukturen / Code)
- Lücken (unvollständige Datensätze)
- Inkonsistenzen in den Daten (z.B. Transliteration, Kodierung, Abkürzungen)
- Intransparenz (Dokumentation / Interpretierbarkeit von Werten)
- unklare Herkunft (Expertise der Datenersteller unklar / fehlende Versionierung)

# Datenorganisation

- Datenmenge wächst sukzessiv während des Projektverlaufs
- Gute Organisation hilft Zeit und Nerven zu sparen
- Konventionen und Abläufe klären für
  - Ablageort (Netzlaufwerken, Shares)
  - Synchronisierung
  - Benennung
  - Versionierung
  - Lebensdauer
  - Sicherung



# Daten speichern & teilen ?

- Speichermöglichkeiten – Dropbox, Google, Amazon, Microsoft, Apple usw.
- Google Docs
- Google (u.a. für Suchmaschinen)

# Warum nicht all die tollen Tools im Netz?

- Nachhaltigkeit (wie lange wird es den Dienst Gebühren)
- Wie lange werden Daten aufbewahrt (AGB)
- Wird der Dienst irgendwann kostenpflichtig
- Wer hat Zugriff / Rechte an meinen Daten?
- In welchem Format werden meine Daten abgespeichert
- Kann ich sie wieder löschen?

# Beispiel Google Drive

Stand 2017: Google erhält das Recht Daten weltweit

- zu verwenden
- zu hosten
- zu speichern
- zu vervielfältigen
- zu verändern
- abgeleitete Werke daraus zu erstellen
- zu kommunizieren
- zu veröffentlichen
- zu verteilen

<https://www.heise.de/download/blog/Cloud-Anbieter-Wie-steht-es-um-den-Datenschutz-3658164>

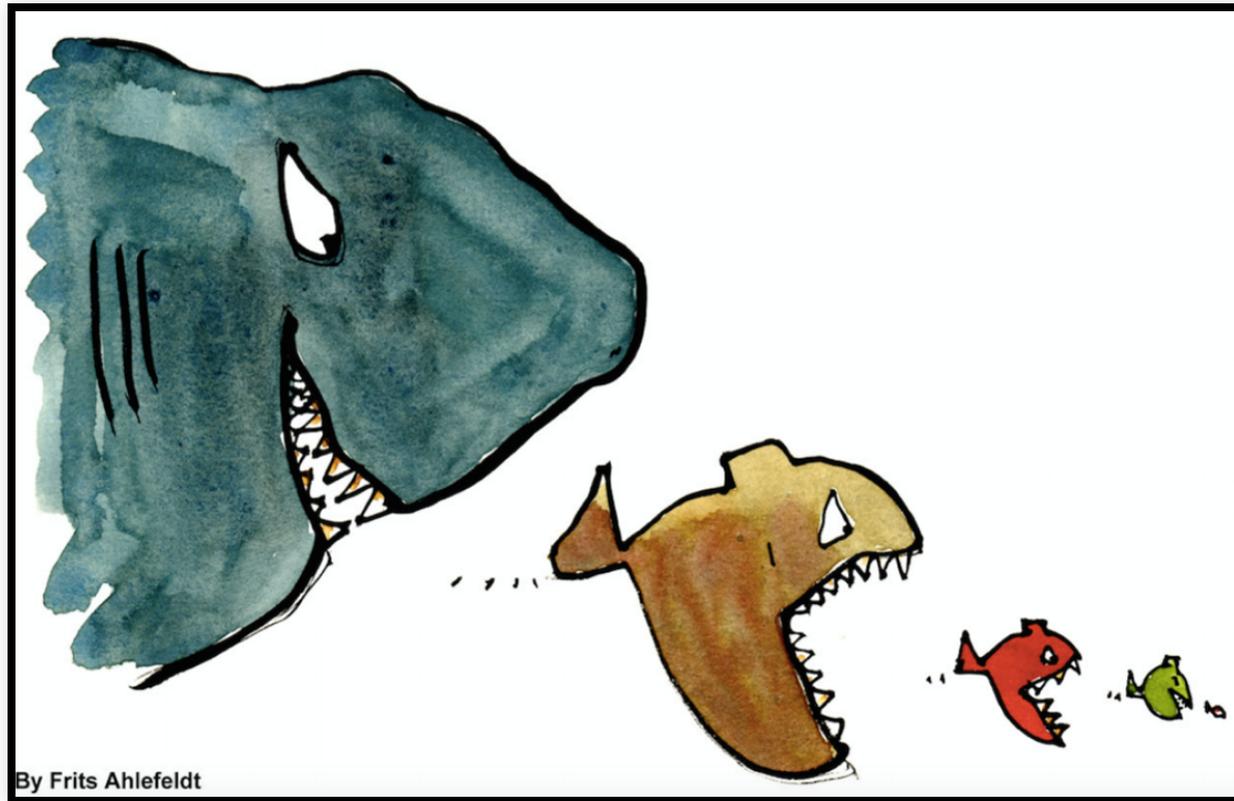
Stand heute: Vervielfältigung, Weitergabe, Veröffentlichung nur mit Einverständnis

# Warum benutzen wir nicht all die tollen Tools im Netz?

Stichwort Nachhaltigkeit

- Online-Anbieter
- Wie lange wird es den Dienst geben?
- Gibt es Pflichten zur Datenaufbewahrung?
- Gibt es (in Zukunft) Kosten?
- Was passiert mit meinen Daten?
- Gibt es Nachfolger- bzw. Vertretungsregeln?
- Werden proprietäre Dateiformate verwendet?

# Und nicht zuletzt...



...gestalten wir unsere Welt mit

# Dienste am RRZ

## **UHH-Share**

Ordner und Dokumente teilen

## **UHH Disk**

Ordner und Dokumente teilen

## **OWA- Office Exchange**

Kalender und Funktionsmailadressen

## **Sharepoint**

Dateien, Office-Anwendungen, Listen

## **gitlab**

Quellcode / Texte, Wikis, Aufgabenlisten

# Wahl der Mittel

- welche Funktion muss der Dienst erfüllen?
- welche Personen(gruppen) sollen darauf Zugriff haben?
- wie sicher ist der Dienst (Backups, Redundanz, Verfügbarkeit)?
- was passiert, wenn ich den Dienst wechseln will / muss?
- welche Nutzungsbedingungen sind an die Nutzung geknüpft?

# Datenorganisation - Versionierung

## Mit Datei- und Ordnernamen

- Suffix mit laufender Nummer (V1, V2..., final, copy...)
- Verkettung einheitlicher Bestandteile in festgelegter Reihenfolge
- Vermeidung von Sonder- und Leerzeichen
- Nach Möglichkeit jede Änderung kennzeichnen
- Entscheiden Sie sich, wie viele/welche Versionen sie behalten wollen

```
20180301_BohrkernSibirien_ABC_Original
```

```
20180301_Mikrofilm1997_XYZ_005
```

# Bleiben Sie konsistent

L\_12\_cop\_22\_09.php

L\_12\_cop\_27\_10.php

L\_12\_cop\_28\_09.php

L\_12\_cop\_30\_09.php

L\_12.php

L\_13\_cop\_16\_12.php

L\_13\_cop\_20\_12.php

L\_13\_cop\_mittag.php

L\_13.php

L\_13\_zum\_wegnehmen.php

L\_6\_new.php

L\_6.php

L\_7-kopie.php

# Datenorganisation - Versionierung

Viele Anwendungen bieten integrierte Versionierung

- Versionierung in Office-Programmen, Bildverarbeitung etc. aktivieren
- Persistenz über Sitzungen überprüfen
- Achten Sie dennoch auf regelmäßige Backups
  - automatisiert (TSM-Backup des RRZ)
  - auf Servern
  - auf externen Speichermedien

# Versionskontrolle Software / Textbasierte Formate

- Kostenlose Versionskontrollsysteme aus dem Bereich der Softwareentwicklung bieten die Möglichkeit der parallelen Arbeit und kontrollierten Zusammenführung durch Versionierung
- geeignet für alle textbasierten Dateiformate
- *git* ist derzeit de-Facto-Standard, verteiltes Arbeiten
- gitlab ist eine OpenSource Serverlösung für Softwareprojekte

# Was bietet Gitlab?

- Intuitive responsive Web-Oberfläche für git-Server
- Datei-Versionierung auf Zeilenebene
- Integriertes 'Tickettool', *kanban*
- Wiki
- Automatische Code-Übersetzung (CI)
- Integration von Kommunikationskanälen

Universität Hamburg | Projects | Groups | Activity | Milestones | Snippets | Search or jump to... | 16 | 20

F fundus

- Project
- Repository
- Issues 33**
  - List
  - Board**
  - Labels
  - Milestones
- Merge Requests 0
- CI / CD
- Operations
- Registry
- Wiki
- Settings

<< Collapse sidebar

fdm > fundus > Issue Boards

Search or filter results... Add list

**Open** 23 +

- OAI-PMH Schnittstelle**  
feature NTH #173
- Daten Ornithologie**  
data-import #64
- Hilfestellung zur Suche**  
#202
- Höhe / Breite der Bilder**  
data-import #162
- Herpetologie / Daten**  
discuss feature content #108
- Import extrahieren**  
server-administration

**To Do** 6 +

- Datenübertragung**  
To Do #211
- Report: Anzahl der DS mit mehreren Bildern**  
To Do wanze #164
- Persistente IDs für Objekte**  
To Do high\_prio #185
- Rails cache bei Deployment leeren**  
To Do #212
- Labels hinzufügen**  
To Do #217
- Gesteine - Mineralogie**

**Doing** 2 +

- Content Objects hinzufügen und löschen**  
Doing #46
- Teilsammlung beim Import berücksichtigen**  
Doing discuss #87

stag

# Gründe für individuell gehostete Werkzeuge zur Projektorganisation

## Pro

- Erweiterung des begrenzten Angebot an Lösungen
- Nutzerverwaltung in Eigenregie
- z.B. Cloud-Server (Owncloud, Nextcloud..)
- Datensicherheit, Datenschutz

## Contra

- Funktionalität und Support bei kostenlosen Versionen i.d.R. begrenzt
- Updates in Eigenverantwortung
- Datensicherheit, Datenschutz in Eigenverantwortung

# Virtuelle Maschinen am RRZ

- komplette Datenhoheit
- Administrativer Aufwand
- Sicherheit in Eigenverantwortung
- insbes. Eigenentwicklungen ggfs. aufwendig in Administration und Betrieb
- Nachhaltigkeit, Sicherungen, Redundanz
- Standort am RRZ

# Kommerzielle Werkzeuge zur Projektorganisation

- meist bieten diese ein Rundum-Sorglos-Paket
- viele OpenSource-Projekte bieten einen Hosted Service an
- teilweise spezialisieren sich Firmen auch auf den Wissenschaftsbereich
- Datensicherheit, Datenschutz, Nutzungsbedingungen beachten
- viele Features (u.a. Workflows, Speichern und Austausch von Daten, Kommunikation, Versionskontrolle, paralleles Arbeiten)
- Zukunftsfähigkeit und Standort der Firma prüfen (Server)

# Checkliste virtuelle Forschungsumgebungen

- Speicherung / Austausch von Dateien mit allen Beteiligten ?
- Zugriffskontrolle (Authentifizierung & Authorisierung), wer bekommt Berechtigungen und wie gut lassen sich diese steuern?
- Versionskontrolle - wie oft wird ein backup gemacht?
- Daten - gemeinsames Arbeiten konfliktfrei möglich?
- Kommunikation - welche integrierten Wege gibt es?
- Nachhaltigkeit - welche Exportmöglichkeiten gibt es?
- Transparenz - wo werden die Daten gespeichert / gesichert?
- Import / Export - alle Daten vollständig und korrekt?
- Sicherheit - verfügbar auch kurz vor der Deadline?

# Dokumentation - Tools

- Wiki-Software (z.B. Mediawiki, Wiki in gitlab), z.B. für Nutzer-Dokumentation, nicht-technische Dokumentation
- gitbooks (TUHH - Hamburg Open Online University)
- Readme-Dateien (z.B. [Readme.md](#) in gitlab)
- Software-Dokumentationstools abhängig von der genutzten Technologie
- ggfs. auch kostenlose hosting-Angebote (z.B. read-the-docs) für Software-Doku (Nachhaltigkeit?)

*Es gibt deutliche Unterschiede in der Mächtigkeit der verwendeten Markdown-Sprachen*

# Öffentliche Darstellung und Kommunikation - Angebote

- Fiona CMS als Teil der UHH (Antrag RRZ)
  - Schulungen am RRZ
  - gekoppelt an B-Kennung / Uni-Zugehörigkeit
  - Art der Seite vorstrukturiert (Projekt, Konferenz, Institution...)
- WordPress (Antrag Unikolleg, lokale Kennungen möglich)
  - Webseiten, Design relativ frei
  - Blog posting, Kommentare etc. möglich
  - Hostname [xy.blogs.uni-hamburg.de](https://xy.blogs.uni-hamburg.de)
- Hostnamen (Antrag Abt. Kommunikation und Öffentlichkeitsarbeit /RRZ)
- Mattermost Chat (TUHH)

# Empfehlungen Datenorganisation

- Informieren Sie sich eingehend über Dateiformate & Software
- Versuchen Sie sich an Standards und verbreiteten Formaten zu halten
- Strukturieren und dokumentieren Sie Ihre Daten und Dateien so, dass Sie auch ohne Ihr Wissen verständlich sind
- Nutzen Sie zentrale Backup-Speichersysteme und Versionierung zur Datensicherung
- Wählen Sie Werkzeuge passend zu Ihren Anforderungen
- Löschen sie nicht mehr benötigte Daten

# Sicherheit

myspace 360 Mio (2008)

Yahoo 500 Mio (2014) 1 Mrd (2013)

LinkedIn 164 Mio (2016)

Adobe 152 Mio (2013)

twitter 32 Mio (2016)

Dropbox 69 Mio (2012)

ebay (145 Mio (2014)

DHL 3 Mio (2016)

# Geschäftsmodell Datenklau

## About us

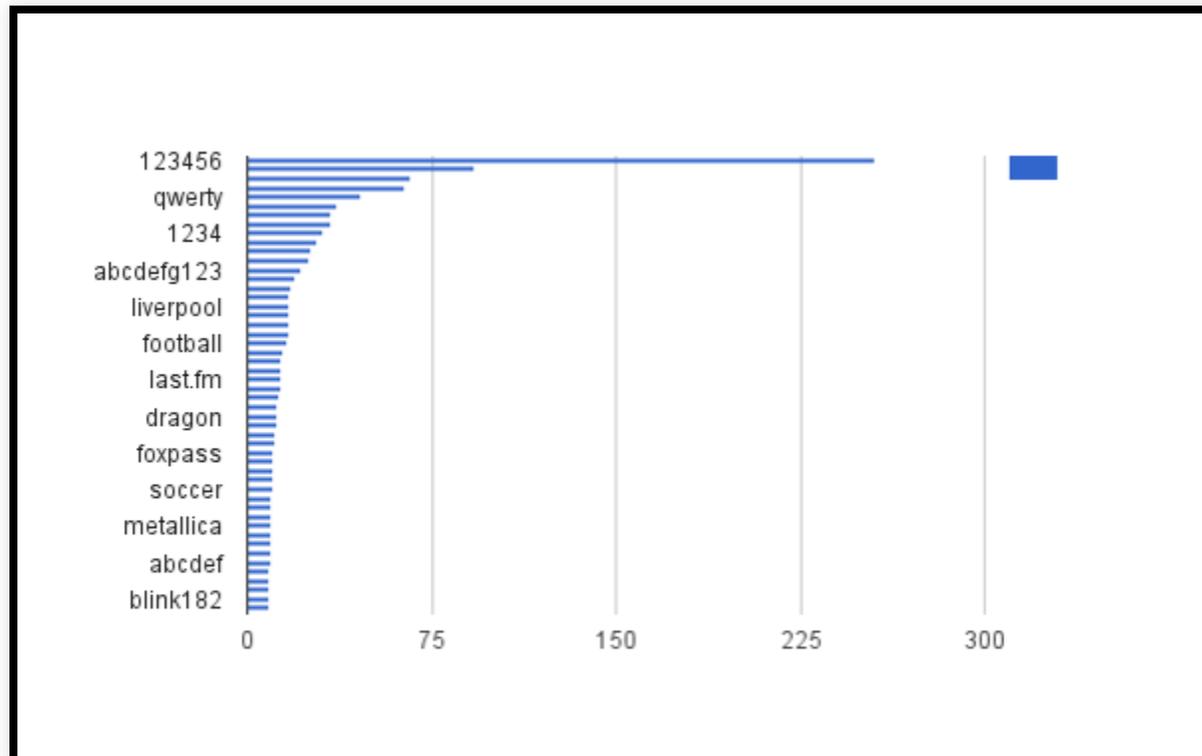
LeakedSource has [exposed](#) every single mega breach of 2016 including LinkedIn, MySpace, and VK.com but because we are the most effective breach notification service in the world, we're back with more.

We have already helped multiple companies, many of them worth billions of dollars, to secure their users [with our API](#) which provides cracked plaintext passwords from all 2 billion leaked records. Businesses are then easily able to compare leaked passwords against their own users and if a match is found, they forcibly change the passwords for affected users. If you're interested in using our services, contact us about an API key.

Any journalists that want to get notified about breaches, [DM us on Twitter with your email address](#)

One final note about the state of affairs on the internet: We have so many databases waiting to be added that if we were to add one per day it would still take multiple years to finish them all. We are also currently processing multiple more mega breaches so stay tuned to our [Twitter](#).

# Häufigste Passwörter (lastfm)



# Passwortsicherheit

- *wichtige* Passwörter nicht wiederverwenden
  - lassen Sie Passwörter automatisch generieren
  - denken Sie sich ein System aus, Varianten von einem Passwort zu erstellen
  - Verwenden Sie einen Passwort-Manager (z.B. LastPass, KeePass)
- Sichere Passwörter enthalten
  - keine Umlaute und Sonderzeichen
  - keine Begriffe aus Wörterbüchern
  - viele Buchstaben / Zahlen / Zeichen
- Achten Sie darauf, wie Ihr Passwort übertragen / gespeichert wird
  - 2-Faktor-Authentifizierung, Verschlüsselung
  - bei Übertragung auf sichere Verbindung (https 😞 /) achten

# Gefahren

- SPAM, Phishing, Trojaner-Mails
- Verkauf im Darknet (insbesondere Streaming-Dienste)
- Einkauf & Umleitung auf Packstation...
- schalten von betrügerischen Anzeigen über Verkaufsplattformen (eBay, Amazon)
- **Verlust Ihrer Daten / Ihrer Arbeit**

# Sicherheit Benutzerkonten

- Überprüfen Sie die Sicherheit Ihrer Konten
  - Unter <https://haveibeenpwned.com/> können Sie herausfinden, ob Ihre Daten offen im Netz kursieren
  - Ändern Sie Dubletten der Login/Passwort-Kombi anderer Konten
- Geizen Sie mit persönlicher Information, persönlichen Details bei Sicherheitsfragen
- Verwenden Sie unterschiedliche E-Mail-Aliase für verschiedenen Konten
- Trennen Sie berufliche und private Nutzung

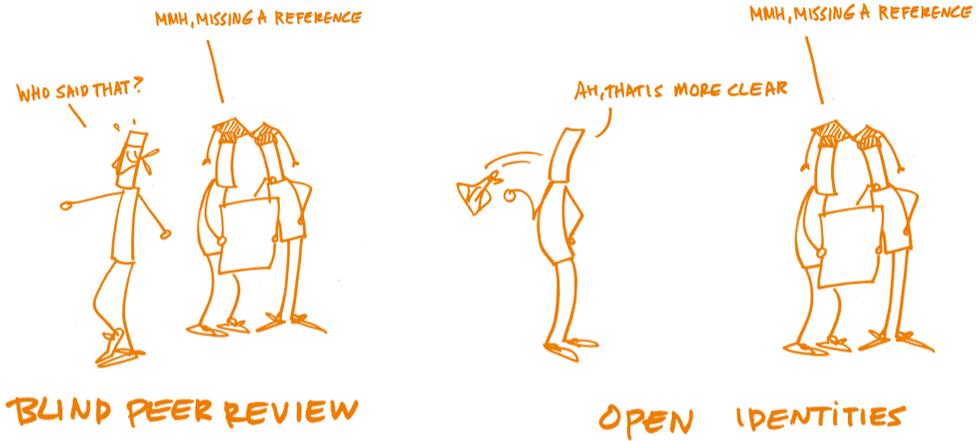
# Bleiben Sie auf dem Laufenden

<p>UNCOMMON (NON-GIBBERISH) BASE WORD</p> <p>ORDER UNKNOWN</p> <p>Tr0ub4dor&amp;3</p> <p>CAPS? COMMON SUBSTITUTIONS NUMERAL PUNCTUATION</p> <p>(YOU CAN ADD A FEW MORE BITS TO ACCOUNT FOR THE FACT THAT THIS IS ONLY ONE OF A FEW COMMON FORMATS.)</p>	<p>~28 BITS OF ENTROPY</p> <p><math>2^{28} = 3 \text{ DAYS AT } 1000 \text{ GUESSES/SEC}</math></p> <p>(PLAUSIBLE ATTACK ON A WEAK REMOTE WEB SERVICE: YES, CRACKING A STOKEN HIGH IS FASTER, BUT IT'S NOT WHAT THE AVERAGE USER SHOULD WORRY ABOUT.)</p> <p>DIFFICULTY TO GUESS: <b>EASY</b></p>	<p>WAS IT TROMBONE? NO, TROUBADOR. AND ONE OF THE 0s WAS A ZERO?</p> <p>AND THERE WAS SOME SYMBOL...</p> <p>DIFFICULTY TO REMEMBER: <b>HARD</b></p>
<p>correct horse battery staple</p> <p>FOUR RANDOM COMMON WORDS</p>	<p>~44 BITS OF ENTROPY</p> <p><math>2^{44} = 550 \text{ YEARS AT } 1000 \text{ GUESSES/SEC}</math></p> <p>DIFFICULTY TO GUESS: <b>HARD</b></p>	<p>THAT'S A BATTERY STAPLE.</p> <p>CORRECT!</p> <p>DIFFICULTY TO REMEMBER: YOU'VE ALREADY MEMORIZED IT</p>

THROUGH 20 YEARS OF EFFORT, WE'VE SUCCESSFULLY TRAINED EVERYONE TO USE PASSWORDS THAT ARE HARD FOR HUMANS TO REMEMBER, BUT EASY FOR COMPUTERS TO GUESS.

# 6. Datenpublikation

## MODES OF PEER REVIEW:



# Was sind Repositorien?

*“Ein Repitorium (lateinisch repositorium, Lager‘) ist ein verwalteter Ort zur Aufbewahrung geordneter Dokumente, die öffentlich oder einem beschränkten Nutzerkreis zugänglich sind.*

*In einem Archiv (lat. Archivum, Aktenschrank‘) werden dagegen ausschließlich historische Dokumente verwaltet.”*

*"Digitale Forschungsdaten-Repositorien sind Informationsinfrastrukturen, die digitale Forschungsdaten möglichst dauerhaft... speichern und organisieren, um die Auffindbarkeit und Zugänglichkeit der Daten zu sichern..."*

Quelle: Pampel (2014), Vortrag „Ausgewählte Aspekte digitaler Informationsversorgung“, HU Berlin

# Repositorien...

Disziplinspezifische Repositorien sind in der Regel der beste Ort für Ihre Daten:

The screenshot shows the homepage of the datorium repository. At the top, there is a navigation bar with the 'gesis' logo and a language selector. Below this is the 'datorium' logo. The main content area is divided into several sections:

- Suche:** A search bar with a 'Los' button.
- Herzlich willkommen bei datorium:** A welcome message explaining the service for social and economic scientists to document and publish their research data.
- Kontakt:** Contact information for the datorium team, including a phone number and email address.
- Replikationen ermöglichen:** A section explaining that datorium supports the replication of research data to increase empirical research.
- Icons:** Four icons representing the core services: 'Beschreiben und Hochladen' (Describe and Upload), 'Sicherung und Review' (Secure and Review), 'Veröffentlichung' (Publication), and 'Suche und Download' (Search and Download).
- Neue Daten:** A list of recent data uploads with titles and dates.

Code/Syntax: Fördert Partizipation am Arbeitsplatz die Entwicklung des politischen Interesses und der politischen Beteiligung?	2019-02-11
RightwingLinks 1.1	2019-01-10
RightwingLinks	2018-12-20

<https://datorium.gesis.org>

# Das Forschungsdatenrepositorium der UHH

...dient allen Angehörigen der Universität Hamburg zur Archivierung und Veröffentlichung von (abgeschlossenen) Datensätzen und Dokumenten aus allen Bereichen der Forschung.

- Anmeldung mit B-Kennung (Shibboleth)
- S3 Speicher Anbindung (5 Petabyte)
- Self-Service Portal
- CoreTrustSeal Zertifizierung in Beantragung
- Sie entscheiden, ob Ihr Eintrag *open*, *restricted* oder *closed access* ist
- Alle Dateitypen möglich
- Upload bis 50GB über die Weboberfläche
- Mindestens 10 Jahre Aufbewahrungszeit
- Jeder Eintrag erhält eine DOI (inkl. Versionen)

<https://www.fdr.uni-hamburg.de/>

# Repositorien mit Dini-Zertifikat

- standardisierte Beurteilung von (Dokumenten-) Servern und erbrachter Dienstleistungen
- Qualitätskontrolle für Dokumenten- und Publikationsservices
- Kriterien u.a.
  - Sichtbarkeit des Gesamtangebotes
  - Betreuung der Autorinnen und Autoren
  - Sicherheit, Authentizität und Integrität des technischen Systems
  - Langzeitverfügbarkeit und Auffindbarkeit der dort archivierten Dokumente
- In Deutschland derzeit ~50 DINI-zertifizierte Repositorien (ausschließlich Dokumenten-Repositorien mit Open Access-Anteil)

# Vertrauenswürdige Repositorien: Zertifizierung (Data Seal of Approval)

- Kriterien für DSA (self assessment)
- Daten können im Internet gefunden werden
- Daten sind verfügbar (eindeutige Rechte und Lizenzen)
- Daten sind nachnutzbar (Format)
- Daten sind zuverlässig/vertrauensvoll
- Daten haben PIDs, so dass auf sie verwiesen werden kann



# Daten in den Repositorien sind nicht selbstredend

- Der Katalog erläutert nur unzureichend
- Deshalb sind Metadaten unabdingbar
- Detailinformationen auf den Landing Pages und in Zusatzdokumenten
- Landing Pages können langfristig nicht mehr zugänglich sind
- Datenproduzenten stehen für Nachfragen nicht mehr zur Verfügung

*Daten sollten nachnutzbar sein!*

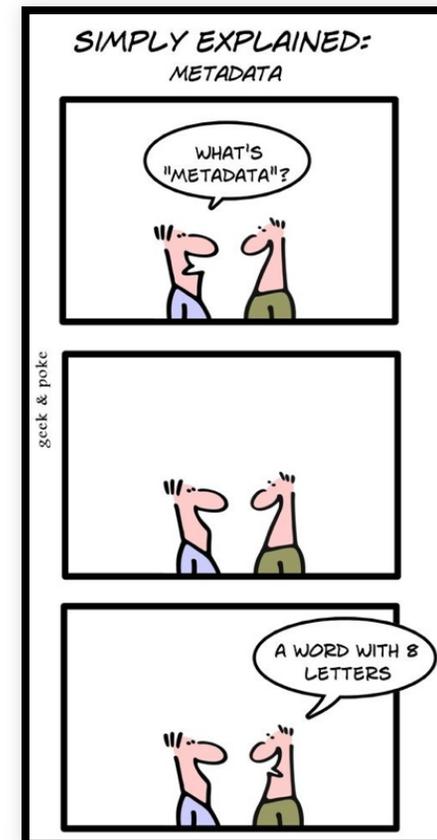
## Damit Daten auch zukünftig nutzbar sind, müssen:

- Daten in einem Katalog gefunden werden
- Daten in einem Format sein, das noch geöffnet werden kann
- Formate in einer format-registry beschrieben sein
- alle Daten und Dokumentation der Archivierungsaktionen in einem Informationspaket zusammengefasst sein
- Daten nachvollziehbar sein (Lücken zwischen dem erwarteten Wissen künftiger Nutzer und dem notwendigen Wissen schließen)

# Was sind Metadaten?

Metadaten sind „Daten über Daten“ und beinhalten beschreibende Informationen über den Kontext.

- bei der Datengenerierung verwendete Technologie (z.B. Hardware und Software)
- administrative Angaben (Projekt, Beteiligte, Institution usw.)
- Beziehungen zu anderen Daten (Publikationen, Personen, Projekten usw.)



# Metadaten-Typen

<b>Metadaten-Typ</b>	<b>Ziel</b>	<b>Beispiel</b>
Beschreibende Metadaten (“descriptive metadata”)	Identifizierung und Auffindbarkeit	Titel, Verfasser, Abstract, Schlagwörter...
Verwaltungsmetadaten (“administrative metadata”)	Erstellung und Speicherung	Rechtmanagement und Zugriffsrechte, Aufbewahrung...
Strukturelle Metadaten (“structural metadata”)	Aufbau der Ressource	Kapitel, Abschnitte, Struktur, Dateien...
Technische Metadaten (“technical metadata”)	technische Merkmale	Umfang, Codierungen, technische Struktur, Formate

# Übung „Metadaten“

- Finden Sie sich in zwei Gruppen zusammen
- Suchen Sie sich ein Objekt aus
- Beschreiben Sie die Metadaten (10 min)
- Anschließend stellt jede Gruppe die Ergebnisse kurz vor.

# Warum sind Metadaten wichtig?

- Metadaten sichern eine langfristige Datennutzung und Interpretation
- Gut dokumentierte Daten sind auch von für Personen, die nicht an ihrer Erhebung beteiligt waren, verständlich
- Strukturierte Metadaten können maschinell gelesen und verarbeitet werden, was eine schnelle, einfache und verlässliche Suche ermöglicht

# Übersicht zu disziplinspezifischen Metadatenstandards

<http://www.dcc.ac.uk/resources/metadata-standards>  
<http://rd-alliance.github.io/metadata-directory/subjects/>

- Stichworte
  - Data Documentation Initiative (Sozialwissenschaften)
  - GESIS; Datorium oder SSOAR (Social Science Open Access Repository)
  - Dublin Core (bibliographisches Datenformat)
  - Marc21 (bibliographisches Datenformat)

## Hilfreiche Links

- Zentrum für FDM (<https://www.fdm.uni-hamburg.de/>)
- IR für GW <https://www.dokserv.gwiss.uni-hamburg.de>
- Forschungsdatenrepositorium <https://fdr.uni-hamburg.de> (coming soon)
- DMP Online <http://dmp.fdm.uni-hamburg.de/> (coming soon)
- [www.forschungsdaten.org](http://www.forschungsdaten.org)
- [www.forschungsdaten.info](http://www.forschungsdaten.info)



## Referenzen - Leseempfehlungen

- Ludwig/Enke, 2013, „Leitfaden zum Forschungsdatenmanagement“
- [Gesis.org](https://www.gesis.org): Anwendung der Datenschutzgesetzgebung auf die empirische Sozialforschung (z.B. Watteler & Kinder-Kurlanda 2015),
- Sicherheitspraktiken in Organisationen (z.B. Poller et al. 2017), und
- Datenethik (z.B. Kinder-Kurlanda & Zimmer 2017, Kinder-Kurlanda & Ehrwein 2015).
- Katsanidou, Alexia, Laurence Horton, and Uwe Jensen. 2016. "Data Policies, Data Management and the Quality of Academic Writing." *International Studies Perspectives* online first.



VIELEN DANK!

