



# Research Data Management

Carolin Odebrecht

Humboldt-Universität zu Berlin, Computer and Media Service

<https://orcid.org/0000-0003-4887-7701>

Anja Herwig

Humboldt-Universität zu Berlin, University Library

<https://orcid.org/0000-0003-1703-3979>





# General information

Presentation slides and materials:

Link to HU-Box will be shared in the chat and afterwards again via e-mail

Questions:

Either in the chat or via the blue hand below the list of participants

Website Research Data Management:

<https://hu.berlin/dataman>



# Services of the research data management initiative

- Advice on handling of research data
- Communicating with and transfer to internal and external services
- Support for data management planning
- Workshops on research data management
- Information events within teaching for PhDs and students
- Cooperation with legal department and data protection officers

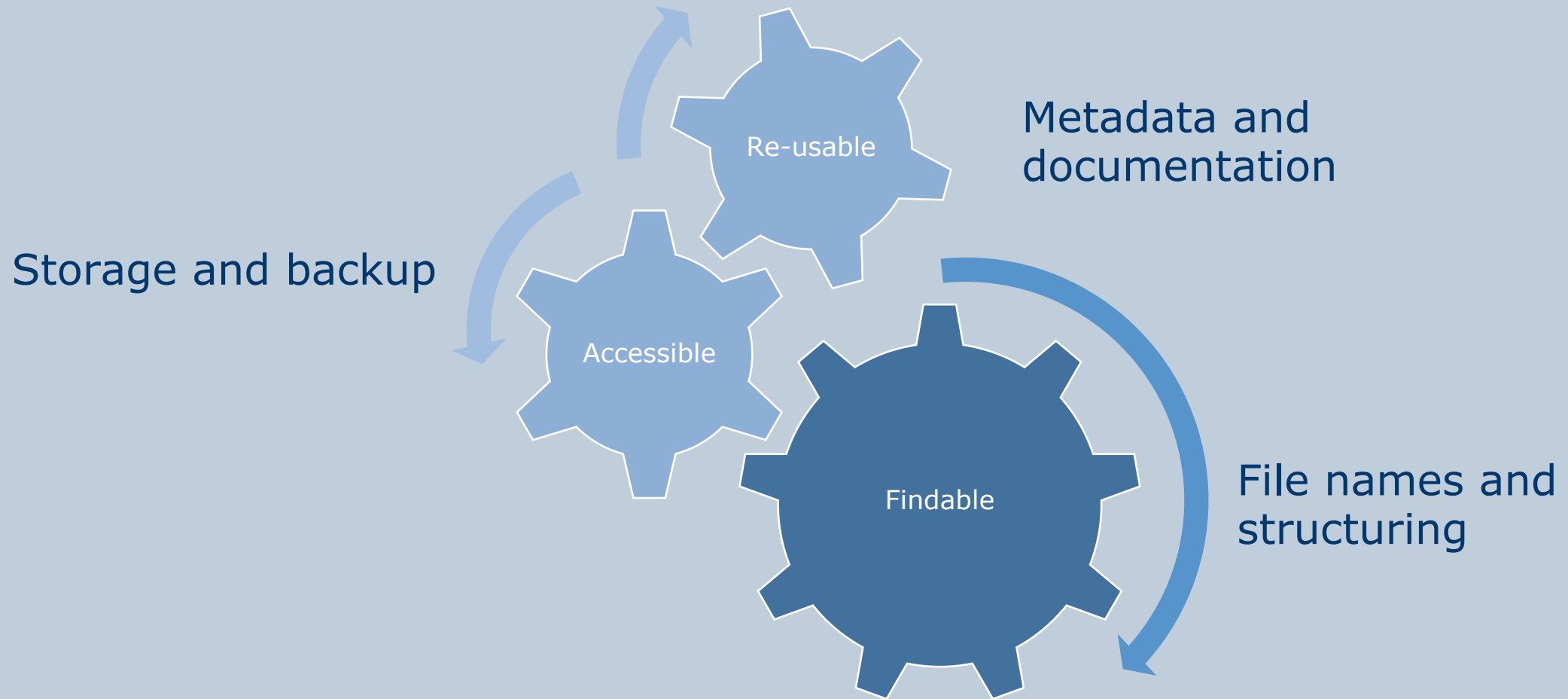
# Welcome!



What's your name?

What are your expectations for this workshop?

# Agenda for today





# Why research data management?

## Own interests

- Good scientific practice
- Knowledge management/knowledge transfer
- Protection against loss of data

## External interests

- Research funders
- Publishers
- Guidelines of research associations and university management



Video: Data Sharing and Management Snafu in 3 Short Acts

[https://youtu.be/66oNv\\_DJuPc](https://youtu.be/66oNv_DJuPc)





# Activity: What is your research data?

Go to [www.menti.com](http://www.menti.com) and use the code **1808 6812**

Tell us in **a few keywords** about the types of research data that you produce within your dissertation.

You can add as many types as you like/need.



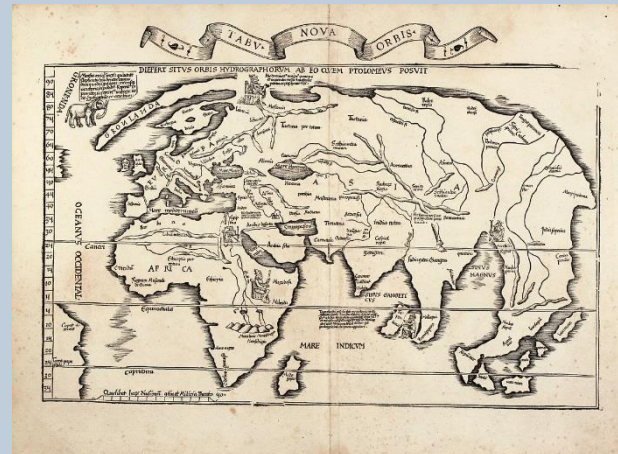


# What is your research data?

expression data  
reaction time-movement  
matlab files numbers in excel sheets  
dna sequencing data  
field notes  
eeg  
fmri data  
interview transcripts  
elisa reads  
fmri dicom behavioral data  
microscopy images  
policies  
eye tracking data  
synaptic currents flow-cytometry data

# What is research data?

- No fixed definition of research data
- In general: „all data, that originate during scientific activity or serve as basis for research results“ ([Kindling/Schirmbacher 2013](#))



243 I: Und wie ging es dann weiter? #00:26:10#

244 A: Ohje,...das ist schon so lange her! Lassen Sie mich überlegen... [Name des Ehemanns] und ich sind  
 245 dann zurück nach [Stadt im Ruhrgebiet]. Er hat dort eine Anstellung bei [Firma] bekommen und wir  
 246 konnten uns endlich ein Auto leisten. Das war aber auch bitter nötig bei der ganzen Fahrerei und den  
 247 drei Kindern! Ich habe ihm in der Zeit den Rücken freigehalten. Bin mit den Kindern viel unterwegs  
 248 gewesen, damit er am Wochenende Ruhe hatte. Vor allem wenn er Schicht hatte, war es nicht immer  
 249 einfach. Da musste der Staubsauger auch mal ruhen (lacht). Meine Schwester fand das ja nie richtig.  
 250 Daheim bleiben und Kinder hüten – niemals, hat sie gesagt. Aber für mich war das damals in  
 251 Ordnung. Ich hatte meine Aufgaben und er seine. Habe ich nie bereut. Jeder muss das für sich  
 252 entscheiden, so seh' ich das. #00:34:03#

253 I: War das später problematisch? #00:35:42#


254 A: Nachdem ich mit den dreien allein war, war es nicht einfach eine Stelle zu finden. Zehn Jahre raus  
 255 aus dem Beruf und kaum Praxiserfahrung nach der Lehre. Ich musste viele Briefe [Bewerbungen]  
 256 schreiben bis Herr [Arbeitgeber] mich genommen hat. Ich bin ihm dafür bis heute dankbar!  
 257 Irgendwie musste ich ja meine Kinder ernähren. Und er hatte Verständnis, wenn eins mal krank war...  
 258 Natürlich war da der Umzug. Meine Schwester ist dafür extra aus [Stadt in Thüringen] gekommen.  
 259 Für die Kinder hat es mir am meisten Leid getan. Aus der gewohnten Umgebung herausgerissen zu  
 260 werden, die Freunde nicht mehr zu sehen. Aber es ging nun mal nicht anders. Ich hatte keine Wahl.  
 261 Eine Alleinerziehende stellt halt nicht jeder ein. War damals so, ist heute so. #00:44:08

Photos: Creative Commons CC0 Public Domain

# Principles of Humboldt-Universität zu Berlin



## Research Data Policy

HUMBOLDT-UNIVERSITÄT ZU BERLIN 

**Grundsätze zum Umgang mit Forschungsdaten an der Humboldt-Universität zu Berlin**

**Präambel**

Der verantwortungsvolle Umgang mit Forschungsdaten ist für die Nachvollziehbarkeit der Forschung, den wissenschaftlichen Fortschritt und die Verbreitung wissenschaftlicher Erkenntnis unerlässlich. Die vorliegenden Grundsätze richten sich an alle forschenden HU-Angehörigen, die sowohl als eigenständige Forschende angesprochen sind als auch in ihrer Funktion als Lehrende und Verantwortliche für die Betreuung des wissenschaftlichen Nachwuchses. Ihre Aufgabe besteht auch darin, Studierende und Promovierende über den adäquaten Umgang mit Forschungsdaten zu informieren und fachspezifische Kompetenzen und Standards zu vermitteln.

**Was sind Forschungsdaten?**


Als Forschungsdaten werden alle Daten bezeichnet, die während des Forschungsprozesses entstehen oder sein Ergebnis sind. Sie werden abhängig von der Forschungsfrage und unter Anwendung verschiedener Methoden erzeugt bzw. gesammelt, bearbeitet, analysiert und schließlich publiziert und/oder archiviert. Demzufolge treten die Forschungsdaten in jeder Wissenschaftsdisziplin in unterschiedlichen Medientypen, Aggregationsstufen und Formaten auf. Für die Bereitstellung und Nachnutzung von Forschungsdaten ist es notwendig, den Entstehungskontext und die benutzten Werkzeuge zu dokumentieren.

**Grundsätze**

Unter Berücksichtigung der [Vorschläge zur Sicherung guter wissenschaftlicher Praxis der Deutschen Forschungsgemeinschaft](#) und der [Satzung der Humboldt-Universität zu Berlin zur Sicherung guter wissenschaftlicher Praxis und zum Umgang mit Vorwürfen wissenschaftlichen Fehlverhaltens](#) werden die nachfolgenden Grundsätze formuliert:

1. Forschende HU-Angehörige sind verpflichtet, die Forschungsdaten sicher zu speichern, angemessen aufzubereiten und zu dokumentieren sowie langfristig aufzubewahren. Die Verantwortung für die Gewährleistung dieser Prozesse liegt bei den HU-Angehörigen, die das Forschungsvorhaben leiten.
2. Alle forschenden HU-Angehörigen sind aufgefordert, die in ihrer wissenschaftlichen Tätigkeit entstehenden Forschungsdaten gemäß den im jeweiligen Fachgebiet etablierten Regelungen bzw. Standards aufzubereiten. Sie dokumentieren den gesamten Forschungszyklus sowie die verwendeten Werkzeuge und Verfahren.

## Guidelines

HUMBOLDT-UNIVERSITÄT ZU BERLIN 

**Handlungsempfehlungen**

*in Ergänzung zu den Grundsätzen zum Umgang mit Forschungsdaten an der Humboldt-Universität zu Berlin*

**Vorbemerkung**

Die Humboldt-Universität zu Berlin misst den Forschungsdaten als wertvolle Grundlage wissenschaftlicher Erkenntnis eine hohe Bedeutung zu. Unter Berücksichtigung der Satzung über die [Grundsätze zur Sicherung guter wissenschaftlicher Praxis](#) hat der Akademische Senat ["Grundsätze zum Umgang mit Forschungsdaten an der Humboldt-Universität zu Berlin"](#) beschlossen. Die vorliegenden Handlungsempfehlungen ergänzen das o.g. Dokument und geben praktische Hinweise zur Umsetzung dieser Grundsätze.

Zur Erfüllung der Grundsätze wird eine zentrale Unterstützung der Universität benötigt. Unter Federführung des Vizepräsidenten für Forschung und durch die Zusammenarbeit der Serviceeinrichtungen Computer- und Medienservice, Universitätsbibliothek und Servicezentrum Forschung werden zu diesem Zweck ein Konzept erarbeitet und Serviceangebote aufgebaut. Diese werden in die vorhandene Serviceinfrastruktur integriert. Die Abstimmung erfolgt mit den Fakultäten und Instituten. Bei der Gestaltung der Serviceangebote orientiert sich die Humboldt-Universität an internationalen Entwicklungen und kooperiert mit anderen Hochschulen. Die Koordination erfolgt durch die im Sommer 2012 gestartete Forschungsdatenmanagement-Initiative.<sup>1</sup>

**Allgemeiner Hinweis**

Bei der Veröffentlichung von Forschungsdaten empfiehlt die Humboldt-Universität, sich an dem Grundprinzip „Zugänglich, wenn möglich, eingeschränkt, wenn notwendig“ zu orientieren. Die Nutzung von geeigneten Datenarchiven oder Repositorien fördert dabei die Sichtbarkeit und Nachnutzungsmöglichkeiten der Forschungsergebnisse. Darüber hinaus unterstützt die Humboldt-Universität die Prinzipien von „Open Data“.<sup>2</sup>

Im Folgenden werden Empfehlungen zum Umgang mit Forschungsdaten für verschiedene Phasen eines Forschungsvorhabens gegeben.

1. **Planungsphase: vor dem Forschungsvorhaben**

**Anforderungen der Förderorganisationen**

Wenn Ihr Forschungsvorhaben im Rahmen eines Drittmittelprojektes durchgeführt wird, erkundigen Sie sich im Vorhinein nach eventuell geltenden Richtlinien zur Nachnutzung von im Projekt entstehenden Forschungsdaten. Einige Förderorganisationen verlangen explizit, Forschungsdaten nach dem Projektende öffentlich zugänglich zu machen. Insbesondere werden

<sup>1</sup>Die Aktivitäten der Initiative werden auf der Website <http://www.oms.hu-berlin.de/stataman> dokumentiert. Für nähere Erläuterungen siehe Royal Society (2012): Science as an open enterprise S. 12. Online zugänglich unter [http://royalsociety.org/uploads/Files/Royal\\_Society\\_Contentpolicypolicy/saee/2012-08-20-SAOE.pdf](http://royalsociety.org/uploads/Files/Royal_Society_Contentpolicypolicy/saee/2012-08-20-SAOE.pdf)

Version 1.3 vom 21.07.2014 - 1 -

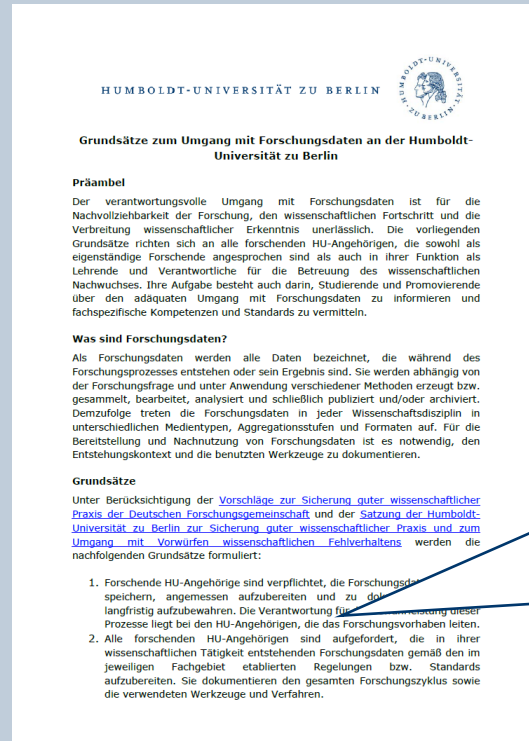




# Research Data Policy

## Policy

„Humboldt-Universität zu Berlin has committed itself to establishing the means to enable this policy to be fulfilled.“



„HU researchers are committed to the secure storage, processing, documentation and preservation of their research data for the long term. Responsibility for ensuring these processes lies with HU staff in charge of research projects or programmes.“



# Discipline-specific policies

Some examples:

- [Guidelines on the Handling of Research Data in Biodiversity Research](#)
- [Data Management in Psychological Science: Specification of the DFG Guidelines](#)
- [Provision and Use of Quantitative Research Data in Educational Research \(available in German only\)](#)
- [Research data management in the social sciences, behavioral sciences and economics \(German Data Forum; available in German only\)](#)



# Good Scientific Practice

- Principles that ensure a culture of research integrity
- Duty for every researcher and every institution
- Regulation on different levels





# Good scientific practice

## Institutions

- HU [Satzung](#)
- TU [Grundsätze](#)
- FU [Satzung](#)
- Charité [Geschäftsstelle](#)

## Funding Agencies

- Deutsche Forschungsgemeinschaft. (2019). Guidelines for Safeguarding Good Research Practice. Code of Conduct. <https://doi.org/10.5281/zenodo.3923602>
- European Commission H2020 [Grant Guide](#)

# Guidelines for Safeguarding Good Research Practice



## Example DFG

- G 7: Cross-phase quality assurance
- G 8: Stakeholders, responsibilities and roles
- G 9: Research design
- **G 10: Legal and ethical frameworks, usage rights**
- **G 11: Methods and standards**
- **G 12: Documentation**
- **G 13: Providing public access to research results**
- **G 14: Authorship**
- G 15: Publication medium
- G 16: Confidentiality and neutrality of review processes and discussions
- **G 17: Archiving**



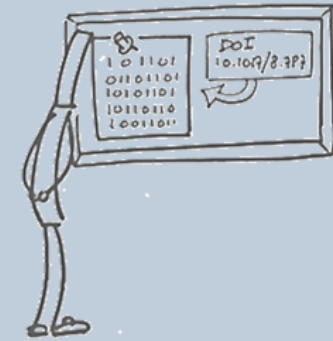
# FAIR Principles

- **F**indable  
Research data should be easy to find.
- **A**ccessible  
Research data should be easy to access.
- **I**nteroperable  
Research data should be interoperable with other systems.
- **R**e-usable  
Research data should be reusable.

## FAIR DATA PRINCIPLES



FINDABLE



ACCESSIBLE

HOW DO YOU OPEN A .XZQ FILE?



INTEROPERABLE



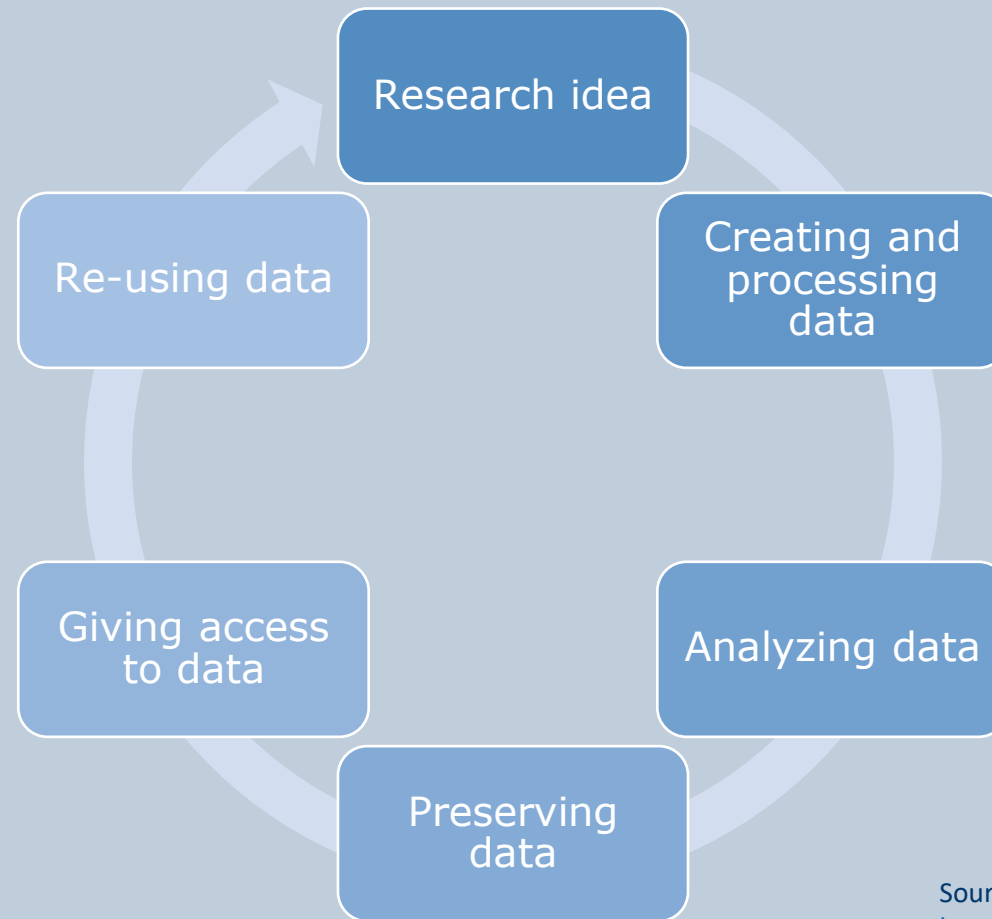
REUSABLE

FAIR Principles:

<https://www.go-fair.org/fair-principles/>



# Research data lifecycle



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
















# File names and structuring



So many boxes. I forgot to label my boxes.

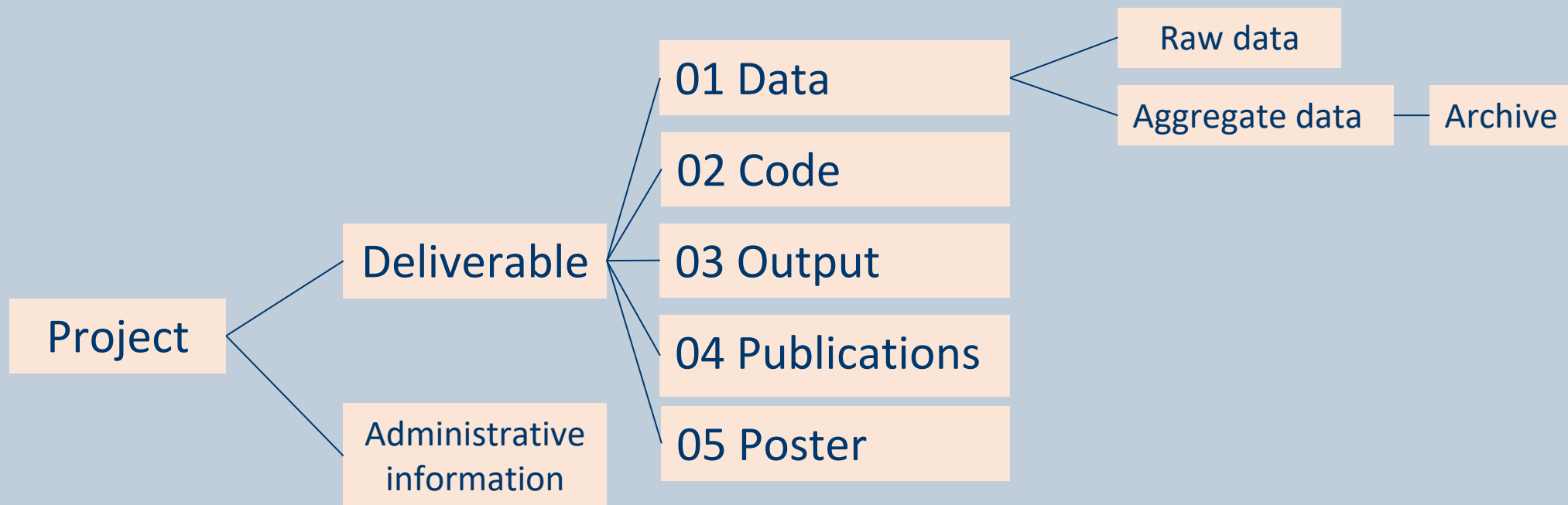
__Lehm	1_Ren	Abgaben	ACorona	ACorona-Scann
AG_Bib	AGEP BFI	AGEP Göhring	AGEP O	AG-Recher
Aktions	Aktuelle	Alma.Leh	Auslage	Ausleihe G
Ausleih	Aussonderungen	Aussonderungen	AZUBI_F	Benutzung
Beschar	Betriebs	Bibliomet	BoysDay	Bücher-Ver
Cano	CMS-Fe	Coffee-L	Coyo_H	DB_Schulung
DBS_wi	Digi-AG	Digitalis	DIN_nur	Diss_Scan
Disserte	Dubl_Lin	EDV_Was	Einband	Elektronis
Elsevier	Endnote	Erfassung	Etiketten	Evifa_Mier
Fachref	Fachrefe	FAK.Leh	Fernleih	FID_Harbei
Flaeche	Forschung	Forschung	Forschung	Fortbildung
Fortbilo	Fotos_Fil	Fremdph	Gleitzeit	Gleitzeitb
Gleitzei	Golda	Goobi-Vie	GZ_Dien	GZ_Foyer
GZ_Sca	GZ_Schli	GZ_Schul	GZ_Stud	GZ_Weihn
GZ_Wei	Handapp	Herzberg	Heuser	Historische
hofer_b	IDS	Infokomp	Inforun	Jahresber
jour fixe	Kalender	Kasse	Klausur	KOBV_Open
Kuehls	Lehman	Lehrbuch	Lenz	Lieferanten
Lilienth	Literatur	Magazin	Miersch	Mini-Schul
Mitarbe	Mittelab	MK-Auss	Nexis-S	OeA_Aust
Öffnun	OpenAcc	OpenAcc	Praesent	Praesentat
Praktiku	Primo.L	Provenie	Referen	Referenda
Reggen	Restauri	ReWi-Sta	RFID_Ba	Rieger
Roeder	Röhrich	Sammlun	SAP_Bra	Schimmel
Schnup	Schulung	Schulung	Schulung	Schulz_D
SG_Aus	social	SP_Umste	Speicher	SSG_Inhalt
Stahnke	Statistik	Statistik	steinbo	Sunderma
Szczepa	TA_für_d	Tauschlis	TeamEO	Telefonie
Theken	Tutoren	Tutorial	UB I	UB II
UB III	UB IV	UB2B_Ru	UBimPu	Umsatzber
Umzug	Umzug	Unguelte	UrhWiss	Urlaubspla
Verluste	Vorlagen	Vw_Erw	Wahlvor	Webredak
Webrec	Webreda	Webreda	Webreda	Webredak
Worksh	Zoll_Bre	ZwB_E3	ZwB-Wi	Zweitverö
.DS_Sto	Derror.doc	9231299C	Arbeits	ARK Staats
Auftr.Sc	Auskunf	Bedarfsa	Benutzun	Berechnung
Bestellu	beyerb	BR_AD_A	CD-Lauf	Datenaus
Datenar	dB52005	Dienstpla	DNSName	DNSName
DNSNa	GBZProj	Gleitzeit	GZ_Büro	HUHS0604
infodier	INIFILE	jedit4.3	kleine_S	Kontakt
Liste_Ar	Magazin	Maske_Z	Mehrfach	OA_Schulung
oktav 3	PIN.doc	Profiles	Profiles	Rettung F
RVK_Sig	RVK_Sig	RVK-Sig	Sign.sch	Standard
Theken	UBTemp	url_liste	Vanden	Verabschie
viereck	Vorakzessi	Zeitschri	Zeitschri	

# File structure

 01Posteingang	01.12.2020 11:51	Dateiordner
 02Allgemeines	03.02.2016 13:16	Dateiordner
 03Verwaltung	24.02.2020 10:06	Dateiordner
 04Monographien-Erwerbung	19.11.2020 16:43	Dateiordner
 05Benutzung	23.11.2020 14:03	Dateiordner
 06Servicekräfte	30.10.2020 15:31	Dateiordner
 07Zeitschriften	30.11.2020 14:11	Dateiordner
 08Elektronische Medien	09.12.2011 10:12	Dateiordner
 09Oeffentlichkeitsarbeit und Infokompet...	03.09.2020 11:20	Dateiordner
 10Statistik	17.09.2020 14:35	Dateiordner
 11Archiv	02.06.2008 13:55	Dateiordner
 12SHK intern	24.02.2020 13:44	Dateiordner
 Nawi (huzub37c.user.hu-berlin.deUBProj...	21.07.2020 11:33	Verknüpfung

# File structure

- Clear directory structure  
Meaning should be clearly visible; also for other researchers
- Hierarchical structures make finding easier:

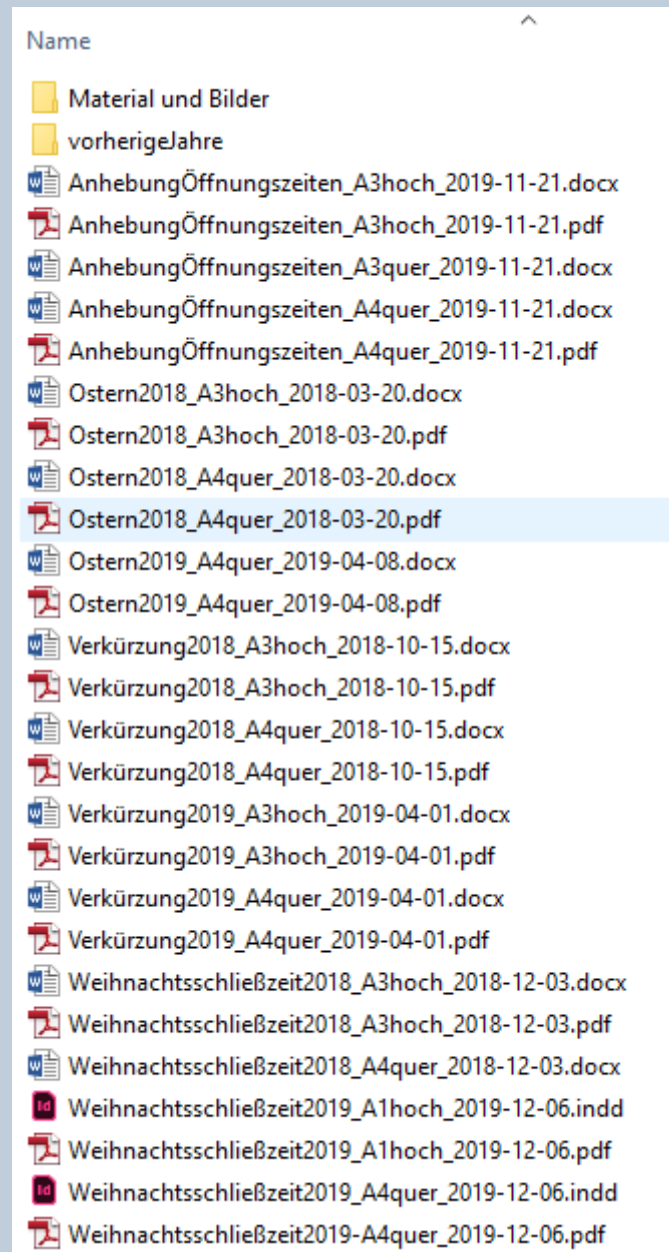


# File naming



Source: MANTRA – Research Data Management Training: Organising data. <http://datalib.edina.ac.uk/mantra/organisingdata>, [last access: 08.12.2020], the image is licensed under Creative Commons Attribution 4.0 International (CC BY 4.0), <https://creativecommons.org/licenses/by/4.0/>

# File naming







# File naming

- Use conventions for file naming
  - Document any naming conventions or used abbreviations
    - e.g. [Project]\_[Interview]\_[Place]\_[Person-ID]\_[YYYYMMDD].mp4
    - Use date/time stamps or a separate ID (e.g. v1.0.0) for each version
    - The date should be at the beginning or end of the filename to facilitate sorting
  - Avoid special characters { } [ ] < > ( ) \* % # ` ; " , : ? ! & @ \$ ~
  - Use the underscore (\_) or capitalization of the first letter to separate names
- Check the file versions
  - Put obsolete file versions separately after a backup
  - A version control software can be very helpful (→ CMS support)



# GitLab

- For versioning of files and content
- Cooperative development of software code
- Registration via HU-Account
- Cooperation with external users possible (invitation and registration)

Website: <https://scm.cms.hu-berlin.de>

# Storage and backup



I will need that back when you are finished,  
that is my only copy.

# Backup and security of data





# Backup and security of data

CMS-  
Support

- Backup
  - At least three copies
  - Use at least two different systems
  - One of the systems should be decentralized (cloud)
- Test the data recovery at the beginning and at regular intervals
- Protect your (sensitive) data
  - Hardware (for example in a separate lockable room)
  - Encrypt your files
  - Use save passwords (→ KeePass)
  - At least two people should have access to your data



# HU-Box

- Secure and mobile storage for files
- Automated backup
- Data is stored on HU servers
- Flexible access options for group work and sharing
- Integration as virtual drive possible (Drive Client)
- Ability to work with personal data within encrypted and password protected folders
- Collaborative writing with OnlyOffice in the browser version

Website: <https://box.hu-berlin.de>

Installation instructions: <https://blogs.hu-berlin.de/hu-box/hilfe-faq/> (in German)

# HU-Box



**HU-Box** | Hochladen | Neu | Freigegeben | Dateien in dieser Bibliothek | [Benutzer] | [Einstellungen]

Gruppen / Datawomen / Forschungsdatenmanagement / Workshops und Veranstaltungen / Schulungen

Name	Größe	Letzte Änderung
<input type="checkbox"/> Agrar-Bio		2017-09-11
<input type="checkbox"/> BWB Rechtsworkshop		2019-10-10
<input type="checkbox"/> BWB Train-the-Trainer FDM		2019-11-29
<input type="checkbox"/> Chemie		
<input type="checkbox"/> Datenmodellierung		
<input type="checkbox"/> DMP-Schulung		
<input type="checkbox"/> Ethnologie		
<input type="checkbox"/> Evaluationsbogen		
<input type="checkbox"/> FDM SLF		
<input type="checkbox"/> Geowissenschaften		
<input type="checkbox"/> Humboldt Graduate School		
<input type="checkbox"/> IZBF		
<input type="checkbox"/> Schulung Erwachsenenbildung PSE		
<input type="checkbox"/> Schulungsmaterial		
<input type="checkbox"/> Workshop Psychologie		
<input type="checkbox"/> Workshop Sozialwissenschaften		
<input type="checkbox"/> Teilnehmerzahlen-Schulungen.xlsx		

Name	Änderungsdatum	Typ	Größe
<input checked="" type="checkbox"/> Agrar-Bio	03.06.2020 12:05	Dateiordner	
<input checked="" type="checkbox"/> BWB Rechtsworkshop	03.06.2020 12:01	Dateiordner	
<input checked="" type="checkbox"/> BWB Train-the-Trainer FDM	03.06.2020 12:01	Dateiordner	
<input checked="" type="checkbox"/> Chemie	03.06.2020 12:05	Dateiordner	
<input checked="" type="checkbox"/> Datenmodellierung	03.06.2020 12:01	Dateiordner	
<input checked="" type="checkbox"/> DMP-Schulung	14.10.2020 11:46	Dateiordner	
<input checked="" type="checkbox"/> Ethnologie	03.06.2020 12:05	Dateiordner	
<input checked="" type="checkbox"/> Evaluationsbogen	03.06.2020 12:05	Dateiordner	
<input checked="" type="checkbox"/> FDM SLF	03.06.2020 12:05	Dateiordner	
<input checked="" type="checkbox"/> Geowissenschaften	03.06.2020 12:04	Dateiordner	
<input checked="" type="checkbox"/> Humboldt Graduate School	03.06.2020 12:04	Dateiordner	
<input checked="" type="checkbox"/> IZBF	03.06.2020 12:03	Dateiordner	
<input checked="" type="checkbox"/> Schulung Erwachsenenbildung PSE	03.06.2020 12:03	Dateiordner	
<input checked="" type="checkbox"/> Schulungsmaterial	22.09.2020 08:06	Dateiordner	
<input checked="" type="checkbox"/> Workshop Psychologie	03.06.2020 12:03	Dateiordner	
<input checked="" type="checkbox"/> Workshop Sozialwissenschaften	03.06.2020 12:03	Dateiordner	
<input checked="" type="checkbox"/> Teilnehmerzahlen-Schulungen.xlsx	28.10.2020 15:36	Microsoft Excel-Ar...	25 KB



# File formats



But is there a way to read the hexadecimal?





# File formats

For long-time archiving files should be:

- unencrypted,
  - uncompressed,
  - non-proprietary/patent-encumbered,
  - and in an open, documented standard
- 
- More information on open file formats:  
[https://en.wikipedia.org/wiki/List\\_of\\_open\\_formats](https://en.wikipedia.org/wiki/List_of_open_formats)

# Activity: Which file formats are recommended for long-term archiving?



Please answer the Zoom survey!



# File formats

File type	Recommendation	Avoid
Tables	CSV, TSV, SPSS portable	XLS, SPSS
Text	TXT, ODT, HTML, PDF/A only if layout is important	DOC, PPT, RTF
Multimedia	Container: MP4, MKV Codec: Dirac, FLAC, FFV1	Flash, MOV, WMV
Images	TIFF, PNG, DNG, SVG	GIF, JPG
Data exchange	XML, RDF, JSON	ODB, MDB, DBF

Save files in a recommended format in addition to the original software format!

Good and detailed overview:

<https://www.ianus-fdz.de/it-empfehlungen/dateiformate> (in German)

# Documentation



What is the content of the field called „Sam2“?



# Activity: Writing a documentation

You will now be separated into two groups. In each group one object will be presented. The object and its use should be described as well as possible.

Each group will receive a link to an editable document in the HU-Box. The documentation is to be noted there.

After 10 minutes the breakout session ends and the other group checks the documentation for content and completeness.



## Documentation exercise – Group 1

Name: Can opener

category: kitchen utensil

description: ca. 20 cm long metal object, with two handles and a black plastic knob

purpose: opening cans

things needed: opener, can, can should be placed on an even and stable surface

Use instructions: place the opener on top of can horizontally. along the rim of the can in between the cutting part and the turning part. Squeeze handles. For best results, turn the black plastic knob.



## Documentation exercise – Group 2

**Filename:** can\_opener\_20220215

Components : A handler, a cutting wheel and a wheel rotative component.

**Use:** The purpose of the object is to open metal cans. It works by hinging to the side of the can and a rotating wheel with spikes pierces the metal surface. You then rotate the wheel across the circumference of of the can to open it completely. You can then empty the contents and recycle the can.

User description : The following object is employed to manually cut through metallic cans. In order to use this object, press the wheel on the can lip and rotate it in order to cut through the metal. Once finished, simply remove the metallic lid.

**Appearance:** It is made of metal and it has really durable structure.

**Storage:** The object is typically stored in one's kitchen.



# Readme

- Documents the research process and the data
- Includes:
  - Abstract
  - Hypotheses
  - Information on the collection of data (methods, units, time, place, devices)
  - Measures for data cleansing (deletion of outliers, weighting)
  - Structure of the data and its relationships to each other
  - Explanation of variables, labels and codes
  - Differences between different versions
  - Information on access and terms of use



# Data Dictionary

A data dictionary can help you to keep track of your data:

File name	Data Type	Creator	Date	Description	Rights	Long-term availability
analysis1.xlsx	Excel table	John Doe	2015	First results from microscopy experiment	CC0	yes
worksite1.tif	TIF image	Jane Doe	2015/04/24	Image of worksite prior to excavation	CC BY	no
worksite2.tif	TIF image	Jane Doe	2015/08/15	Image of worksite after discovery of object 1	CC BY	yes
Object1.png	PNG image	John Doe	2015/08/16	Object 1, discovered at worksite	CC BY	yes



# Metadata



Everything you need to know is in the article!

# Example metadata

Determinants of Unemployment in the European Union. An empirical Study of the Federal Republic of Germany (FRG), France, Great Britain and Italy		
DOI	<a href="https://doi.org/10.4232/1.8198">10.4232/1.8198</a>	
Version	1.0.0	
Resource Type	Dataset	
Creator	• Hubert, Frank	
Publication Date	2006	
Language	German	
Classification	<ul style="list-style-type: none"> <li>• ZA:               <ul style="list-style-type: none"> <li>◦ Historical Studies Data</li> </ul> </li> </ul>	
Description	<ul style="list-style-type: none"> <li>• <b>Abstract</b>                Since the oil price shock in 1974 unemployment increased significantly and also did not really decline in periods of economic ups ... <a href="#">see more</a></li> </ul>	

Temporal Coverage	• 1961 - 1993
Geographic Coverage	<ul style="list-style-type: none"> <li>• Germany (DE)</li> <li>• France (FR)</li> <li>• United Kingdom (GB)</li> <li>• Italy (IT)</li> </ul>
Collection Mode	Types of sources: Scientific publications, Statistics published by the OECD Statistics published by the European commission Annual expert ... <a href="#">see more</a>
Data and File Information	<ul style="list-style-type: none"> <li>• Type of Units: Text Unit</li> <li>Number of Units: 37</li> <li>Number of Variables: 111</li> </ul>
Availability	<ul style="list-style-type: none"> <li>● Download</li> </ul> A - Data and documents are released for academic research and teaching.
Rights	All metadata from GESIS DBK are available free of restriction under the Creative Commons CC0 1.0 Universal Public Domain Dedication. Howeve ... <a href="#">see more</a>

Source: Hubert, Frank (2006): Determinanten der Arbeitslosigkeit in der EU. Eine empirische Analyse für die Bundesrepublik Deutschland, Frankreich, Großbritannien und Italien. Version: 1.0.0. GESIS Datenarchiv. Datensatz. <http://doi.org/10.4232/1.8198>



# Metadata

An overview of (discipline-specific) metadata standards can be found here:

<http://rd-alliance.github.io/metadata-directory/subjects>

- Examples:
  - Dublin Core (multidisciplinary)
  - ISO 19115 (geosciences)
  - Darwin Core (biodiversity)
  - Data Documentation Initiative (social sciences)
- Support: [Subject librarians of the University Library](#)



# Questions?



I think I cannot use your data.



# Helpful tools

- Versioning: [Git](#), [Subversion](#), [Bitbucket](#)
- File formats: [PRONOM](#), Wikipedia's [list of file formats](#)
- Password manager: [KeePass](#)
- Data management planning tools: [DMPonline](#), [RDMO](#)
- Electronic lab/field notebooks: [Labfolder](#), [sciNote](#), [Rspace ELN](#), etc.
- E-learning modules: <https://campus.hesge.ch/researchdatamanagement/>



# What to keep in mind

**Lifecycle** describes the process from data creation to publication

Find the **balance** between perfection and feasibility

Take advantage of the various **support services** offered by HU

**Research data management** offers

- ✓ Trustworthy data
- ✓ Reproducible data
- ✓ Easier handling of large amounts of data
- ✓ Generation of durable data

# Feedback



Rate this workshop and its content!  
Please answer the Zoom survey.



# Thank you very much for your interest!

Carolin Odebrecht, research data management coordinator

Anja Herwig, data librarian

[forschungsdaten@cms.hu-berlin.de](mailto:forschungsdaten@cms.hu-berlin.de)

Further information about research data management  
as well as our workshops and training:

<https://hu.berlin/dataman>



Follow us on Twitter [@DatawomenHUB](https://twitter.com/DatawomenHUB)

