

Ich hab es hier irgendwo

Grundlagen der Arbeit mit Daten

Felix Rau & Patrick Helling
(Universität zu Köln)



Data Center for the Humanities
Kölner Datenzentrum
für die Geisteswissenschaften

Worum geht's?

Kontrolle beim Umgang mit Daten.

Fahrplan

Zeit		Inhalt
14:00 Uhr	5 min	Intro
14:05 Uhr	30 min	Daten & Co.
14:35 Uhr	10 min	FDM Prinzipien
14:50 Uhr	45 min	Workflows
15:20 Uhr	20 min	Diskussion 
15:50 Uhr	10 min	Abschluss 

Wer sind wir


Felix Rau



Patrick Helling

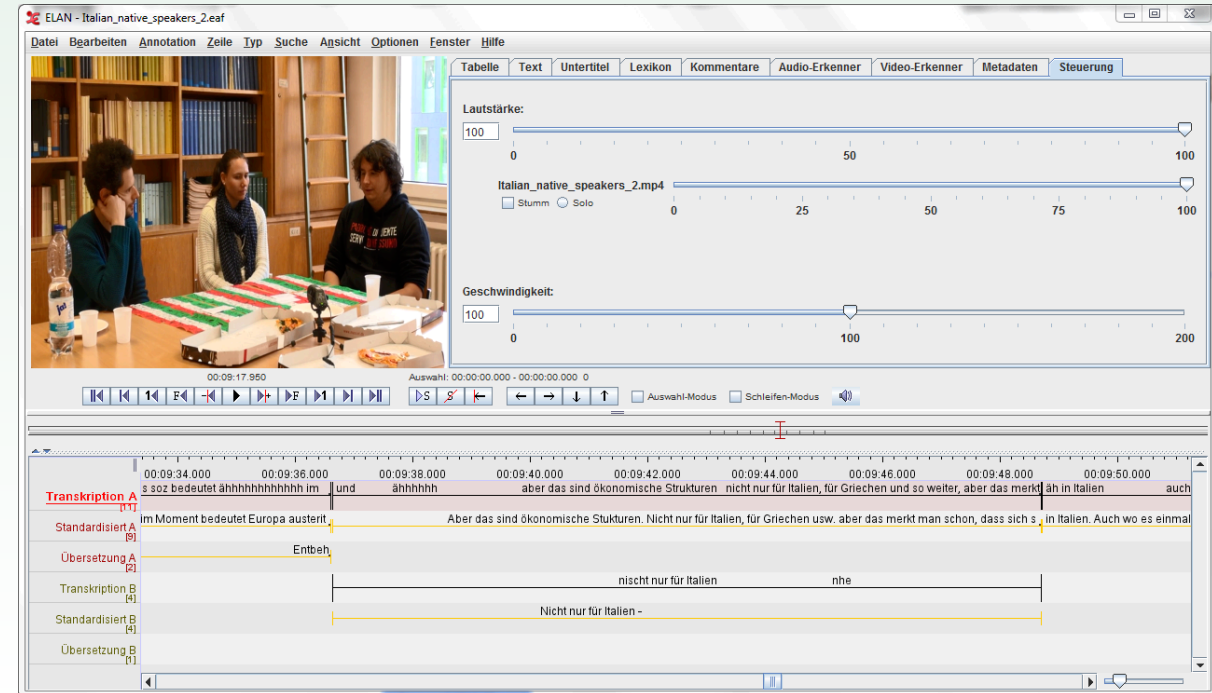


Daten & Co.

Dauer	Inhalt
5 min	Daten
5 min	Metadaten
5 min	Paradaten
5 min	Anderer Output
10 min	Diskussion 

Daten

In der Linguistik:
Repräsentation eines
gesprochenen,
gebärdeten oder
geschriebenen
Sprachevents.



Datentypen

Literatur 📖, Bilder 🖼️, Audio 🎵, Fotos 📷,
Datenbanken 💿, Videos 🎥, Objekte 🏺, Eye tracking
👁️, Tabellen 📊, Manuskripte 📝, Software 💾, fmri-
Daten 🧠, ...

Daten

Leonelli 2016 defines data as

“ any product of research activities, ranging from artifacts such as photographs to symbols such as letters or numbers, that is collected, stored, and disseminated in order to be used as evidence for knowledge claims. ”

Leonelli, Sabina. 2016. *Data-Centric Biology: A Philosophical Study*. Chicago ; London: The University of Chicago Press.

Datenmodelle

Sprachliche Interaktionssituation → Videoaufnahme
→ Transkript → Übersetzung → Morphem-
Glossierung → Intonationsannotation
30 Aufnahmen 🖐️ ⇒ numerischer Datensatz
(Morphem-Intonations-Korrelation)

Daten und Datenmodelle

- Welche Daten habe ich?
- Welche Derivate (Transformationen/Annotationen) habe ich?
- Wie sind diese verschiedenen Datenmodelle verknüpft?

Metadaten

“ Metadaten sind strukturierte Informationen über Forschungsdaten, die das Sammeln, Speichern, die Analyse und Verbreitung der Daten ermöglichen und unterstützen. ”

Metadaten (Beispiele)

The screenshot shows a spreadsheet with a table titled "Bundle Information". The table has columns for various metadata fields. The first row is a header row with a purple background. The second row contains instructions for each column. The third row contains the actual column headers. The fourth and fifth rows contain example data.

	Bundle Information							Location
1	Instructions for the metadata file should be unique. Can be identical to BundleCharacteristics but using only ASCII characters and no spaces, no "-", no apostrophe, no "", no special characters, e.g. &S&A...	Title of the bundle (human readable) that will be displayed online. For better recognition through the files this should be unique.	A short description for this bundle.	Keywords: e.g. songs, conversation, ... List them in one field with commas.	Language name that will be displayed. If there is more than one language, list them in one field with commas. Choose a value you have already filed in the matching column in General Info to connect the information to the bundle.	Please specify date as follows: YYYY-MM-DD	Title of the bundle will be published, usually the year in which you fill out the sheet.	The most relevant location for the bundle (just one), e.g. the location of recording. Choose a value you have already filed in the matching column in General Info to connect the information to the bundle.
2	Fields that must be filled in are marked in orange	MetadataFileName	BundleDisplayTitle	BundleDescription	BundleKeyword	LanguageDisplay Name	BundleRecordingDate	BundlePublicationYear
3	Example	WidgetSongs	Widget Songs	Recordings made by Gumbert in Nuremberg in 1880, we received a copy of the recording on tape from Almut Degen.	Interviews, legacy data, digitized tapes, Rubricated.	Zugheim	1999-12-29	2024 Köln
4	Example	AdditionalWidgetSongs	Additional Widget Songs	An additional bundle.	Interviews, legacy data, digitized tapes, Rubricated.	Zugheim	2009-12-29	2024 Köln
5								
6								
7								
8								
9								
10								
11								
12								
13								
14								
15								
16								
17								
18								
19								
20								
21								
22								
23								
24								
25								
26								
27								
28								
29								

The screenshot shows the Zenodo record page for "RDM Demerit Badges". The page includes the Zenodo logo, the title "RDM Demerit Badges", the authors "Rau, Felix" and "Mollenhauer, GJ Elisabeth", and the publication date "Published March 6, 2025 | Version 1.0.0". The page also features a "Show affiliations" button, a "Cite all versions?" section, and a "Versions" section showing the current version 1.0.0. The page is part of the "Data Center for the Humanities der Universität zu Köln".

RDM Demerit Badges

Rau, Felix · Mollenhauer, GJ Elisabeth^{1,2}

Published March 6, 2025 | Version 1.0.0

RDM Demerit Badges

The RDM Demerit Badges were developed by Felix Rau and Elisabeth Mollenhauer at the Data Center for the Humanities at the University of Cologne. The Demerit Badges are intended to raise awareness for common mistakes in research data management (RDM) and to encourage good RDM practices as well as a positive culture around mistakes. The badges were designed to be used in workshops, lectures, and other teaching or outreach settings. The home of the RDM Demerit Badges is <https://dch.phil-fak.uni-koeln.de/rdm-demerit>.

All materials are licensed under the [CC-BY 4.0](#) license.

See also:

Rau, F. & Mollenhauer, E. (2025). Fehlerkultur und die Vermittlung von Forschungsdatenmanagement. D44 2025 Under Construction (D442025). Bielefeld, Deutschland. <https://dx.doi.org/10.5281/zenodo.14943156>

File types and function

In this resource, we provide files that can be used to reproduce the stickers (image files) and files to adapt and modify the stickers (Adobe Illustrator files). The InfoSheets are intended to contextualize and explain the badges in an offline teaching setting. They are provided as PDF files for reproduction and as InDesign files for modification.

Badges

- Adobe Illustrator files for exporting JPEGs for printing ([badges_illustrator_print](#))
- Adobe Illustrator files for exporting web images ([badges_illustrator_web](#))
- JPEG files for printing stickers ([badges_png_print](#))
- PNG files for web pages ([badges_png_web](#))

InfoSheets

- InDesign files for exporting PDFs ([infosheets_indesign](#))
- PDF files for printing ([infosheets_pdf](#))

19 VIEWS **17 DOWNLOADS**

Versions

Version 1.0.0 Mar 6, 2025

Cite all versions? You can cite all versions by using the DOI 10.5281/zenodo.14983643. This DOI represents all versions, and will always resolve to the latest one. [Read more.](#)

External resources

Indexed in

[OpenAIRE](#)

Paradaten

“ The paradata of a data set or survey are data about the process by which the data were collected. [Wikipedia](#) ”

Paradaten (Beispiele)

- Ursprünglich aus der Umfrageforschung
- Heute auch in anderen Disziplinen
- Beispiele: Zeitstempel, keystroke-Daten, Sprechtempo, Befragungszeitpunkt, ...
- Können als eine Art Metadaten betrachtet werden

Anderer Output

- Code
- Paper
- Was habe ich und wie ist es mit den Daten verknüpft?

Wie sieht es bei euch aus? ✨

Wie sieht es bei euch aus?

- Daten
 - Rohdaten
 - Derivate/Datenmodelle
- Metadaten
- Paradata
- Anderer Output

FDM Prinzipien und Workflows

Dauer Inhalt

10 min FDM Prinzipien

10 min Workflows

10 min Datensuche 

20 min Diskussion 

FDM Prinzipien

- Dokumentiere Strukturen
- Dokumentiere Entscheidungen
- Dokumentiere die Änderungen (z.B. via Code)
- Versioniere Daten
- Automatisiere Prozesse
- Wiederholbarkeit (Repeatability)
- Kontrolle über Daten und Prozess

FAIR

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

- Findable
- Accessible
- Interoperable
- Reusable

Workflows

- Acquiring (erheben/erwerben)
- Keeping (Daten halten)
- Handling (handhaben)
- Releasing (veröffentlichen/löschen/archivieren)

Acquiring (erheben/erwerben)

- Dokumentieren, wie Daten erhoben wurden
 - Paradata
- Datenqualität
- Datenformate
- Rechtliche Aspekte

Datensuche

1. VLO (<https://vlo.clarin.eu/>)
2. BASE (<https://www.base-search.net/>)
3. FDB (<https://www.forschungsdaten-bildung.de/>)
4. Google Dataset Search
(<https://datasetsearch.research.google.com/>)
5. SSH Open Marketplace
<https://marketplace.sshopencloud.eu/>
6. R3Data (Repositories)
<https://www.re3data.org/search>
7. DataCite Commons <https://commons.datacite.org/>

Keeping (Daten halten)

- Speicherung
- Backup
- Dateitypen
- Ordnerstruktur
- Dateibenennung

Speicherung und Backup

- Speicherung
 - Lokal / Cloud / Institutionell
- Backup
 - 3-2-1 Regel
 - Automatisch
 - Regelmäßige Tests

Rau, F. (2023). *DCH Project Data Storage Guidelines* (1.0.0). Zenodo. [doi:10.5281/zenodo.7760967](https://doi.org/10.5281/zenodo.7760967)

Dateitypen

- Offene, standardisierte Formate
- Arbeits- und Archivierungsformate

Ordnerstruktur

```
project/  
├── 01_admin/  
├── 02_data/  
│   ├── 01_raw/  
│   └── 02_processed/  
├── 03_code/  
├── 04_docs/  
├── 05_output/  
├── README.md  
└── LICENSE
```

Rau, F. (2023). *DCH Folder Structure Guidelines* (1.0.0). Zenodo. [doi:10.5281/zenodo.7452113](https://doi.org/10.5281/zenodo.7452113)

Dateibenennung

1. Vergib eindeutige, aussagekräftige, lesbare, aber prägnante Namen
2. Wähle Namen, die in allen Datei- und Betriebssystemen sicher sind
3. Strukturiere den Dateinamen und nutze Dateinamenerweiterungen
4. Erleichtere die alphabetische Sortierung
5. Dokumentiere das Benennungsmuster

Beispiel Dateibenennung

[ISO 639-3]_[speaker]_[date]_[number]_[data type].[file extension]

ger_LK_2023-05-31_001_raw-audio.wav

ger_LK_2023-05-31_001_raw-video.m4v

Rau, F. (2023). DCH File Naming Guidelines (1.0.0). Zenodo. [doi:10.5281/zenodo.7447485](https://doi.org/10.5281/zenodo.7447485)

Handling (handhaben)

- Versionierung
- Dokumentation der Transformationen und Analyse
- Alles in Code
- Automatisierung
- (Analyse)

Releasing (veröffentlichen/löschen/archivieren)

- Datenarchivierung
- Datenpublikation
- Datenlöschung

Datenarchivierung

- Metadaten
- Datenstandards
 - Nachhaltigkeit der Formate
- 10 Jahre Archivierung (Mindestanforderung)
- Langzeitarchivierung

Datenpublikation

- Repositoriumssuche
- Metadatenstandards

Repositoryumssuche

- OSF (<https://osf.io/>)
- Zenodo (<https://zenodo.org/>)
- Fachspezifische Repositorien
 - R3Data (<https://www.re3data.org/search>)
 - <https://www.konsortswd.de/angebote/forschende/alle-datenzentren/>
- Institutionelle Repositorien (Lokale Einrichtungen)

Eure Perspektive

- Wie organisiert ihr eure Daten?
- Welche Herausforderungen seht ihr?
- Welche Best Practices könnt ihr teilen?
- Gibt es offene Fragen oder Unsicherheiten?
- Wie könnte ein idealer Workflow für euch aussehen?
- Welche Tools oder Ressourcen nutzt ihr bereits?

Danke!

info-dch@uni-koeln.de



Data Center for the Humanities
Kölner Datenzentrum
für die Geisteswissenschaften

