



Forschungsdatenmanagement: Von der Planung und Organisation bis hin zur Veröffentlichung

Vortragender: Kevin Lang
Montag, 28.09.2023

Agenda

- 1) Vorstellung**
- 2) Forschungsdatenmanagement**
- 3) Leitlinien zum Umgang mit Forschungsdaten**
- 4) Datenmanagementplan**
- 5) Modelle zum Umgang mit Forschungsdaten**
 - Five S Data Model
 - 3-2-1 Backup-Regel
 - FAIR Data Prinzipien

Was dieser Workshop nicht beinhaltet...

- Rechtliche Aspekte (Urheberrecht, Datenschutz, ...)
- Speichermöglichkeiten und Softwareanwendungen
- Open Science

Vorstellung

Zur eigenen Person

Kevin Lang, Master of Science

- **2011 – 2016:** Bachelor-Studium, Medieninformatik
- **2016 – 2018:** Master-Studium, Computer Science and Media
- **Studentische Arbeit:**
 - Schwerpunkte: Natural Language Processing, Machine Learning, Artificial Intelligence und Big Data
- **Seit 2018:** Stabsstelle für Forschungsdaten und Mitglied des Thüringer Kompetenznetzwerk Forschungsdatenmanagement (TKFDM)

Bauhaus-Universität Weimar

Stabsstelle Forschungsdatenmanagement

Strategie zur Digitalisierung (12/2017)

Themen zu Forschungsdatenmanagement:

- Kompetenznetzwerk
- Forschungsdatenleitlinie (FD-Policy)
- Open Data (Open Science)
- Digitale Bibliothek Thüringen
- Infrastrukturen
- Nationalen Forschungsdateninfrastrukturen (NFDI)
- Langzeitarchivierung



(aus [Thüringer Strategie zur Digitalisierung im Hochschulbereich](#), 2017, zuletzt aufgerufen am 17.06.2019)

Strategie zur Digitalisierung 2021-2025

Neue Themen zu Forschungsdatenmanagement:

- Kompetenznetzwerk(TKFDM) und Thüringer Zentrum für Lernende Systeme und Robotik (TZLR)
- Themenfelder: Vernetzung, Schulung, Open Science, Langzeitarchivierung, High-Performance Computing, Rechtliche Hilfe und Einbindung in die Curricula
- Gründung eines Strategiebeirats
 - Forschungsdatenrepositorium REFODAT
 - Data Steward Programm



(aus [Thüringer Strategie zur Digitalisierung im Hochschulbereich](#), 2021, zuletzt aufgerufen am 17.05.2021)

TKFDM: Portal



Suche

Deutsch | English

AKTUELLES VERANSTALTUNGEN AKTIVITÄTEN MATERIALIEN ÜBER UNS

FDM
THÜRINGEN

**WILLKOMMEN AUF DEN SEITEN DES THÜRINGER KOMPETENZNETZWERKS
FORSCHUNGSDATENMANAGEMENT (TKFDM)**

www.forschungsdaten-thueringen.de

TKFDM: Services

- **Beratung**
 - Datenmanagementplan, rechtliche Grundlagen, Software, Portale, Formate, ...
- **Schulung**
 - thüringenweite Schulungen, Workshops, Train-the-Trainer und Informationsveranstaltungen
 - auch auf Anfrage möglich
- **Vernetzung**
 - Zwischen den Hochschulen, Forschungseinrichtungen, NFDI-Konsortien und anderen Interessierten



TKFDM: Materialien und Veranstaltungen

- Verschiedene Informationsflyer
- Handreichungen & Best-Practices
 - FDM Förderung, Forschungsdatenrepositorien, Open-Source, Datenschutz, ...
 - Qualitätskontrolle, eLabFTW, GitLab, LaTeX, BEXIS, FDM in Kursen, ...
 - FAIR Assessment Tools
- 23 Dinge zu Forschungsdatenmanagement
- ScaryTales: 60 Geschichten zu schlechtem Datenmanagement



Organisationen



Rat für
Informations
Infrastrukturen



Bausteine Forschungsdatenmanagement
Empfehlungen und Erfahrungsberichte für die Praxis von
Forschungsdatenmanagerinnen und -managern

deutsches forschungsnetz



Deutsche Initiative für
Netzwerkinformation e. V.



RESEARCH DATA ALLIANCE



| D | C | C



Warum Forschungsdatenmanagement?

Probleme bei der Wiederverwendbarkeit von Forschungsdaten



[Data Sharing and Management Snafu in 3 Short Acts](#) by NYU Health Sciences Library

Probleme bei der Wiederverwendbarkeit von Forschungsdaten

- Daten nicht öffentlich, obwohl öffentlich finanziertes Projekt
- fehlende Planung von Projektordnerstrukturen
- Wahl des Speichermedium (z.B. USB-Stick)
- fehlende Sicherung der Daten
- Verwendung proprietärer Formate
- keine Metadaten bzw. Dokumentation der Daten
- fehlende ID für Autoren

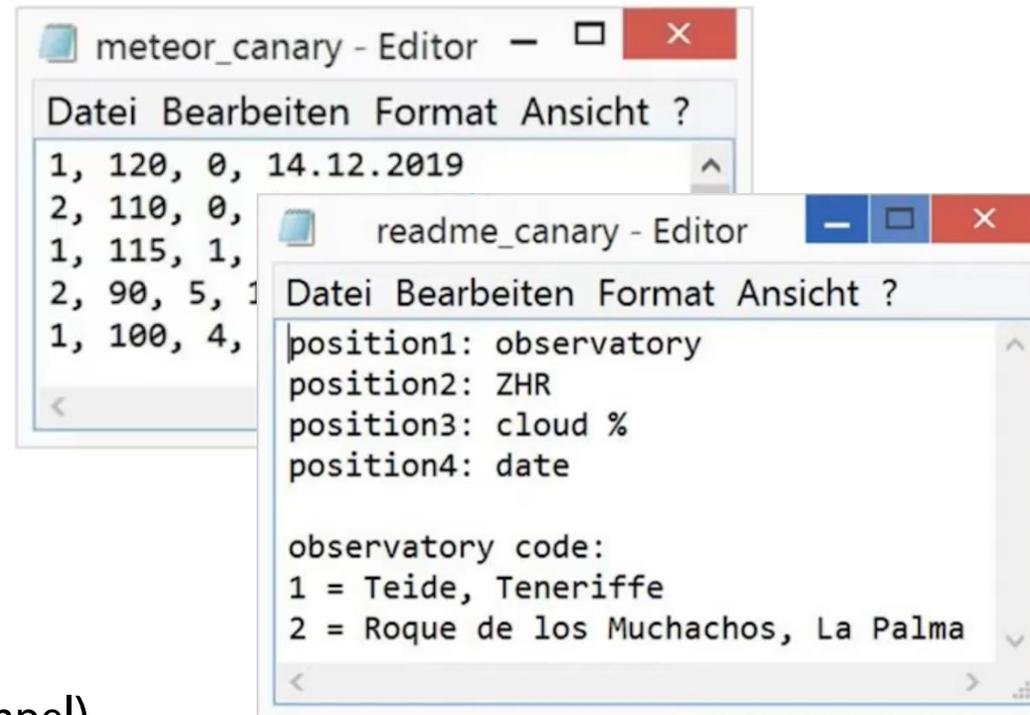


Weitere Beispiele

- Forschung: Sternschnuppen von verschiedenen Observatorien
- Datensätze: Datum, Bewölkungsgrad, Zenithal Hourly Rate, Bundesstaat

	A	B	C	D
1	date	cloud %	ZHR	observatory
2	19-12-14	2	105	NSW
3		5	80	TAS
4		0	130	WA
5	19-12-15	40	25	NSW
6		15	40	TAS
7		0	135	WA
8	19-12-16	10	70	NSW
9		5	90	TAS
10		0	140	WA

- unvollständige Datensätze
- verschieden definierte Abkürzungen
- unterschiedliche Formate (wie Zeitstempel)



Leitlinien zum Umgang mit Forschungsdaten

Was ist eine FD-Policy?

Beschreibt, wie Mitglieder einer Institution mit Forschungsdaten umgehen sollten.

Gründe

- Schaffung von Bewusstsein zum Thema Forschungsdatenmanagement (top down Ansatz)
- Unterstützung der frühzeitigen Planung
- Daten auffindbar, verständlich und wiederverwendbar machen
- Zuarbeit an Drittmittelgeber (DFG, BMBF, EU Horizon 2020 Programm, ...)

Leitlinien und Empfehlungen von Organisationen

- 2010 [Grundsätze zum Umgang mit Forschungsdaten](#)
- 2014 [Management von Forschungsdaten - eine zentrale strategische Herausforderung für Hochschulleitungen](#)
- 2015 [Leitlinien zum Umgang mit Forschungsdaten](#)
- 2019 [Leitlinien zur Sicherung guter wissenschaftlicher Praxis](#)

Leitlinien an der Bauhaus-Universität Weimar

- 2012 [Richtlinie zur Sicherung guter wissenschaftlicher und künstlerischer Praxis](#)
- 2014 [Strategie zur Sicherung und Verwertung von geistigem Eigentum](#)
- 2015 [Leitlinienpapier zur Forschungsförderung](#)
- 2016 [Open-Access-Policy](#)
- 2019 [Gemeinsame IT-Nutzungsordnung](#)
- 2020 [Leitlinie zum Umgang mit Forschungsdaten](#)
- 2021 [Handlungsempfehlung zum Umgang mit Forschungsdaten](#)

**Bauhaus-
Universität
Weimar**

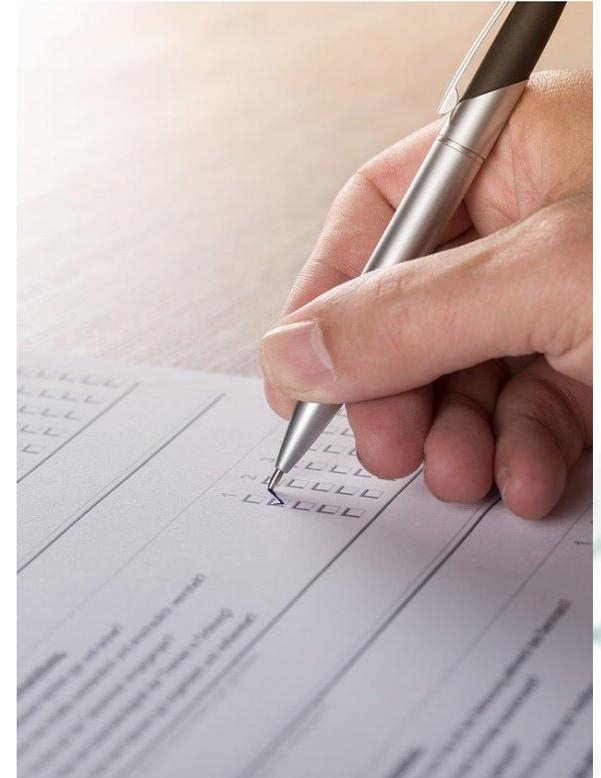
Definitionen

- **Forschungsdaten:**

- Daten der wissenschaftlichen und künstlerischen Forschung, die während des Forschungsprozesses generiert, gesammelt, verarbeitet oder analysiert werden.
- Dazu gehören auch Forschungsergebnisse und gegebenenfalls benötigte Werkzeuge und Verfahren.

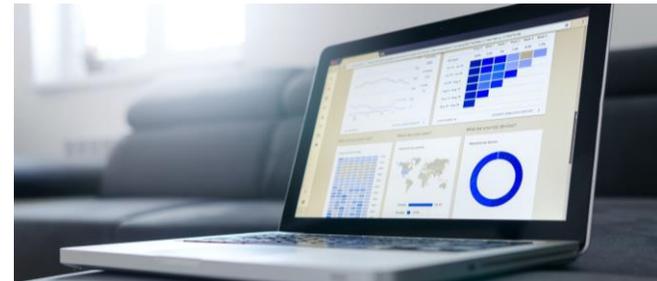
- **Forschungsdatenmanagement:**

- Die gesamte Planung, Erfassung, Verarbeitung, Dokumentation, Archivierung und Publikation der Forschungsdaten.



Punkte der Leitlinie

- Verantwortlichkeiten
- Dokumentation und Aufbewahrung
- Veröffentlichung und Nachnutzbarkeit



Verantwortlichkeiten

Die Leiterinnen und Leiter eines Forschungsvorhabens sind verpflichtet, ein Forschungsdatenmanagement gemäß den Prinzipien guter wissenschaftlicher und künstlerischer Praxis zu etablieren. **Alle Mitglieder des Projekts sind in den Umgang mit Forschungsdaten einzuweisen.**



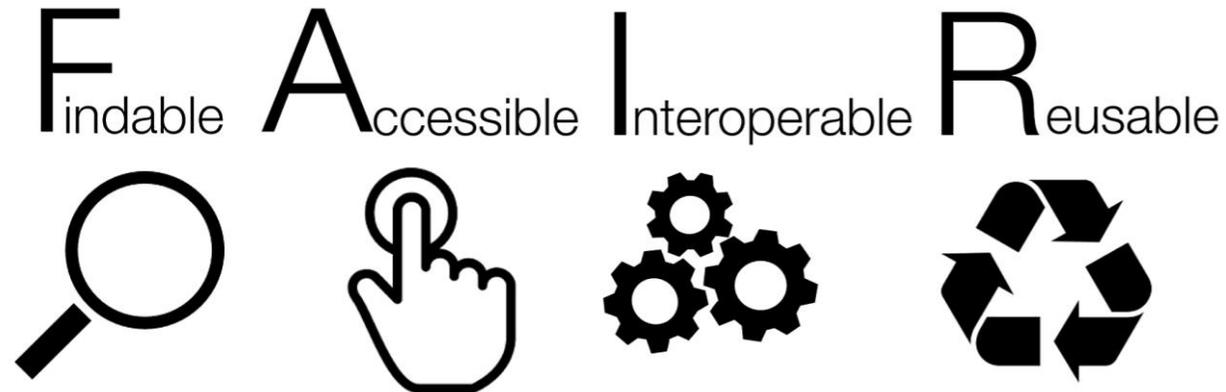
Dokumentation und Aufbewahrung

Während der Durchführung des Forschungsvorhabens sind die **Herkunft, Art und Verarbeitung** der Forschungsdaten zu dokumentieren. Etablierte **Methoden und Standards** sollen dabei angewendet und Nutzungsrechte festgelegt werden. Die wissenschaftlich relevanten Forschungsdaten sollen für einen angemessenen Zeitraum von **mindestens 10 Jahren** bei der entstehenden Einrichtung oder auf einem Repository aufbewahrt werden.



Veröffentlichung und Nachnutzbarkeit

Die Veröffentlichung der Forschungsdaten sollte zeitnah nach **den FAIR-Prinzipien** („**F**indable, **A**ccessible, **I**nteroperable, **R**eusable“) auf anerkannten Archiven und Fachrepositorien erfolgen. Sowohl **Materialien, Informationen, angewandte Methoden und eingesetzte Software** sollen, soweit es möglich und zumutbar ist, verfügbar gemacht werden. Bei der Veröffentlichung von Forschungsdaten müssen u. a. der Schutz personenbezogener Daten, das Urheberrecht sowie berechnigte Interessen Dritter gewahrt bleiben.



Handlungsempfehlung zum Umgang mit Forschungsdaten

- Erweiterung zur Leitlinie
- 10 Seiten mit Inhaltsangabe
- Leitfaden im Projekt:
 - Planung
 - Durchführung
 - Abschluss

1	Präambel	3
2	Projektplanung	4
3	Projektdurchführung	5
3.1	Organisation der Dateien	5
3.2	Rechte und Pflichten	5
3.3	Dokumentation.....	5
3.4	Versionierung und Sicherung	6
3.5	Virtuelle Forschungsumgebungen.....	6
3.6	Nutzung externer IT-Dienste	7
4	Projektabschluss.....	8
4.1	Veröffentlichung nach den FAIR-Prinzipien	8
4.2	Repositorien für Forschungsdaten	8
4.3	Auswahl und Löschung	9
4.4	Lizenzierung.....	9
4.5	Langzeitverfügbarkeit.....	9
5	Zentrale Serviceeinrichtungen	10
5.1	Die Universitätsbibliothek	10
5.2	Das Dezernat Forschung.....	10
5.3	Das Rechenzentrum (SCC)	10
5.4	Rechtliche und sonstige Beratung	10
5.5	Die Bauhaus Research School.....	10

Datenmanagementplan

Was ist eine Datenmanagementplan?

Ein DMP ist ein Dokument für die Planung und Dokumentation der Daten innerhalb eines Projekts.

Gründe

- Koordination zwischen Projektpartnern
- Verständnis und Wiederverwendung der Daten
- Frühzeitiges Erkennen von Problemen
- Grundlage für Drittmittelgeber...



Drittmittelgeber

Förderer	Anforderung	Einreichung mit Antrag	Inhalte nach...	Aktualisierung
 <p>Horizon Europe THE NEXT EU RESEARCH & INNOVATION PROGRAMME (2021 – 2027)</p>	DMP	ja	Horizon Europe Programme Guidelines Data Management Plan Template	nur bei großen Änderungen
 <p>DFG Deutsche Forschungsgemeinschaft</p>	Angaben zum Umgang mit Daten	ja	Checkliste zur Beschreibung des Umgangs mit Forschungsdaten in Forschungsvorhaben	nein
 <p>Bundesministerium für Bildung und Forschung</p>	abhängig vom Projekt	nur falls verlangt	abhängig vom Projekt (z.B. empirischen Bildungsforschung Checklist)	abhängig vom Projekt
 <p>VolkswagenStiftung</p>	DMP	ja	Richtlinien nach Practical Guide von Science Europe	nein

Ursprungsquelle: Kerstin Helbig, HU Berlin

Weiterführung: [Forschungsdatenmanagement in der Forschungsförderung](#)

Anmerkungen zur Erstellung

- Living Document
- Kurze und einfache Beschreibungen
- Alle Beteiligten involvieren
- Stakeholder einbeziehen (Geldgeber, Institut, Forschungscommunity, Mitarbeiter etc.)



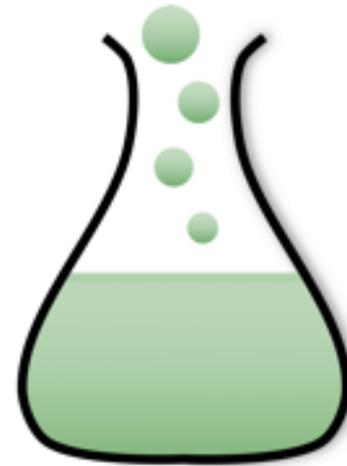
Inhalt: Generelle Informationen

- Permanente ID
- Förderer
- Projekttitle
- Projektbeschreibung (Abstrakt)
- Namen der Beteiligten (und IDs)
- Kontakt Projektleitung oder Abteilungen
- Erstellungs- bzw. Änderungsdatum (ggf. Version)
- Vorgaben/Referenzen von Dritten



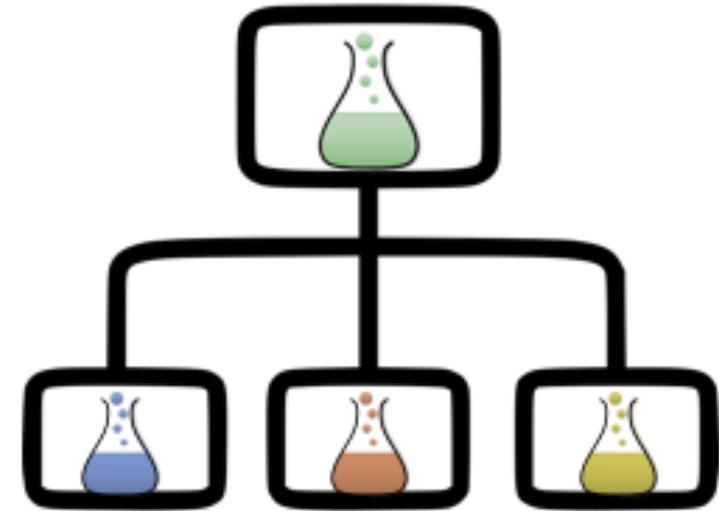
Inhalt: Herkunft bzw. Erhebung der Daten

- Grund der Erhebung?
- Vorhandene Datensätze?
- Verbindung von Datensätzen?
- Einschränkungen?
- Software? Methoden?
- Dokumentation der Erhebung?
- Art, Format und Größe?
- Gründe für Formate?



Inhalt: Dokumentation und Qualität

- Welche Metadatenstandards?
(wie DDI, TEI, EML, MARC, CMDI)
- Organisation der Daten?
(Ordnerstruktur, Versionskontrolle, ...)
- Art der Dokumentation?
(Datenbank, Wiki, Readme, Datei-Header, ...)
- Werkzeuge für (Weiter-)Verwendung der Daten?
- Kontrolle der Datenqualität?



Inhalt: Speicherung und Sicherung

- Medien für Datenspeicherung?
- (Automatisch) Sicherung?
- Einrichtung für die Services?
- Wiederherstellungsplan?
- Zugang zu den Daten und Organisation?
- Datenschutz und Leitlinien?



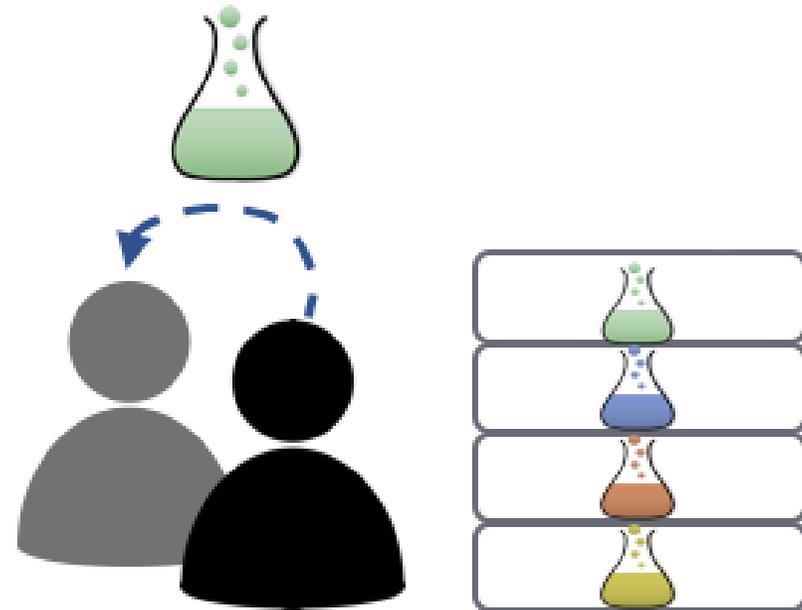
Inhalt: Rechtliche und ethische Aspekte

- DSGVO konform?
- Einverständnis zur Erhebung/Sammlung?
- Pseudonymisierung/Anonymisierung?
- Aufbewahrung und Verschlüsselung?
- Autorisierte Zugänge?
- Rechte an den Daten?
- Verstöße oder Interessen Dritter?
- Institutionelle oder internationale Richtlinien?



Inhalt: Weitergabe, Veröffentlichung und Erhaltung

- Wann Veröffentlichung bzw. Archivierung?
- Auswahl der Plattform?
- Anfragen bzw. Zugang zu den Daten?
- Einschränkungen oder nur Teilveröffentlichungen? (z.B. Embargo)
- Löschung von Daten?
- Gründe zur Weiterverwendung?
- Werkzeuge zum Anschauen/Bearbeiten?



Inhalt: Ressourcen und Verantwortlichkeiten

- Rollen und Pflichten?
- Projekte in Kollaboration?
- Erstellung und Aktualisierung des DMP?
- Kosten für Zeit und Personal?
- Kosten für Hardwareanforderungen?
- Kosten für Einhaltung der FAIR-Prinzipien?



Externe Hilfe und Angebote

- [Musterpläne](#)
- [Research Data Management Organiser](#)
 - forschungsdaten.info
- [DMPonline](#)
- [Data Stewardship Wizard](#)
- [ARGOS](#)



Kleine Pause bis ...
10:55



Modelle zum Umgang mit Forschungsdaten

Übersicht

- **Five S Data Model**
 - Organisation von Forschungsdaten
- **3-2-1 Backup Regeln**
 - Sicherung von Forschungsdaten
- **FAIR-Prinzipien**
 - Veröffentlichung von Forschungsdaten

Five S Data: Ursprung

- Five S stammt aus dem Konzept von „Kaizen“ aus dem Toyota-Produktionssystem, welches soviel bedeutet wie: „*Die Handlung, Schlechtes besser zu machen*“

1	Seiri (整理)	Sort	Aussortieren
2	Seiton (整頓)	Set in Order	Aufräumen
3	Seisō (清掃)	Shine	Arbeitsplatzsauberkeit
4	Seiketsu (清潔)	Standardize	Anordnung zur Regel machen
5	Shitsuke (躰)	Sustain	Alle Punkte einhalten und verbessern

- bekannt bei herstellungsorientierten Unternehmen wie Toyota, Boeing oder Hewlett-Packard
- Bezug zu Forschungsdaten 2019 von University of Helsinki geschaffen (über RDA veröffentlicht)

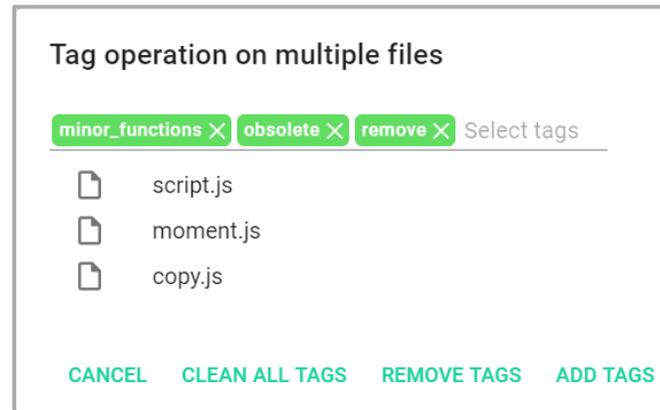
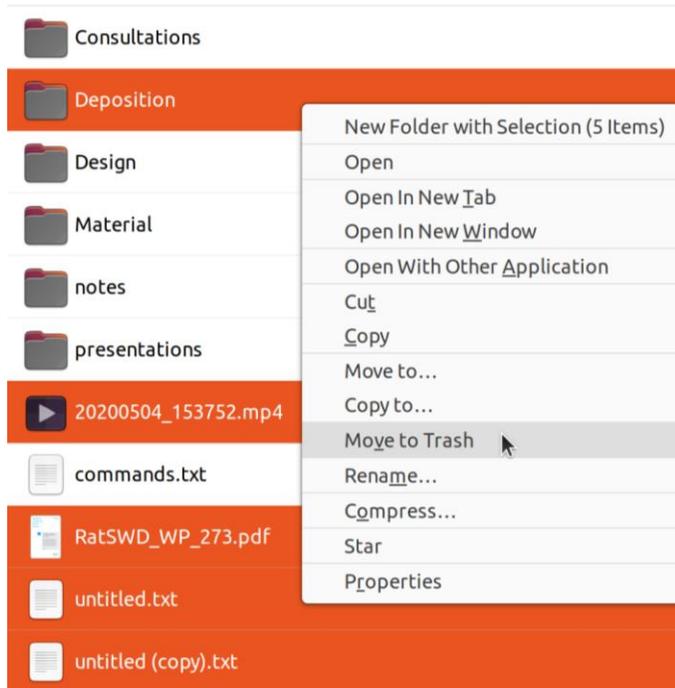
改 善
"kai" "zen"

Five S: 1) Sort

- Ziel:
 - Zeitgewinnung bei der Suche
 - Gewinnung von mehr Datenspeicher
- Umsetzung:
 - unnötige Dateien/Ordner löschen, oder für geplante Löschung markieren
 - vor temporären Dateien sauber halten



Löschung, Markierung und Unterbindung



Quelle: [TagSpaces](#)

.gitignore

```
10
11 # Generated files
12 bin/
13 gen/
14 out/
15
```

Quelle: [GitHub](#)

Five S: 2) Set in Order

- Ziel:
 - Eine Organisation aufbauen um Arbeitsvorgänge zu vereinfachen
- Umsetzung:
 - Sinnvolle Ordnerstrukturen
 - Namenskonventionen
 - Readme für Erklärungen von Ausnahmen

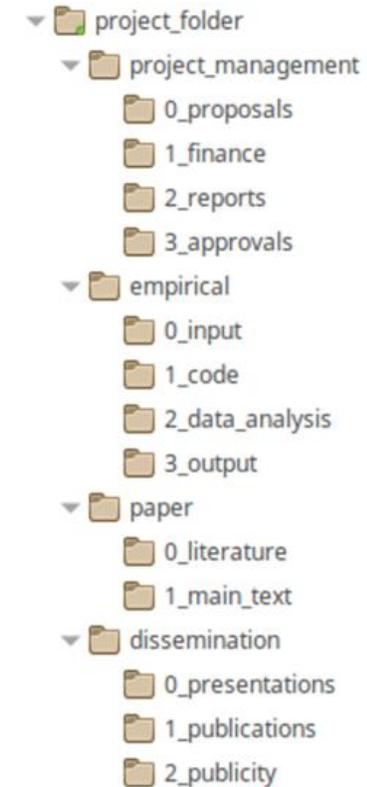


Ordnerstrukturen

Schlechtes Beispiel:

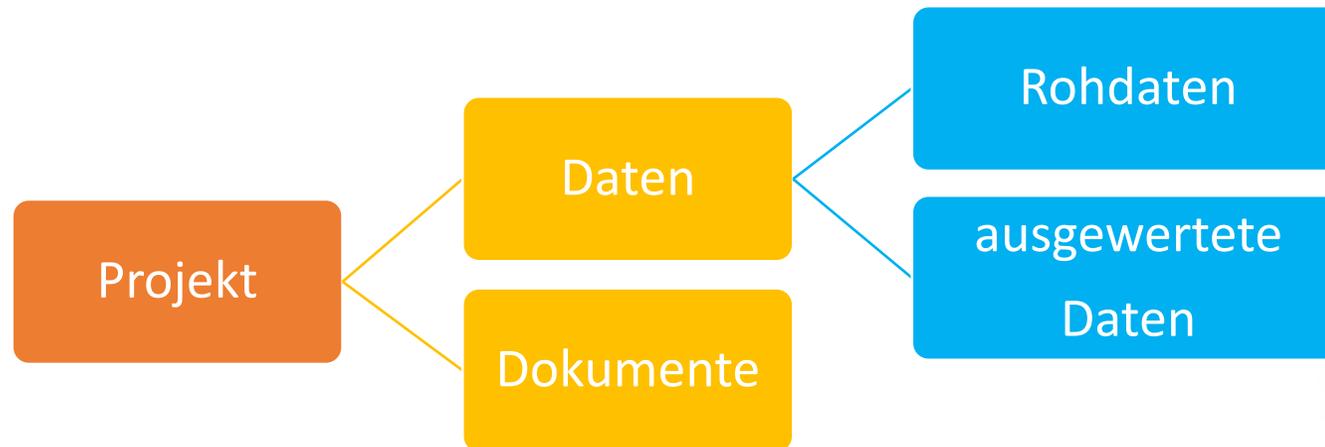
Name	Date Modified	Size	Kind
▶ 1_PROJEKTIT	9 May 2019 at 11.39	--	Folder
▶ 2_PEOPLE	25 Mar 2019 at 9.10	--	Folder
▶ 3_MANUSCRIPTS	25 Apr 2019 at 13.05	--	Folder
▶ 4_APPLICATIONS AND GRANTS	Yesterday at 23.46	--	Folder
▶ 5_SEMINARS and MEETINGS	12 Mar 2019 at 20.54	--	Folder
▶ 6_PERSONAL	21 Dec 2018 at 12.44	--	Folder
▶ 7_ADMIN	28 Mar 2019 at 9.48	--	Folder
▶ 8REFEREE TASKS	7 May 2019 at 13.27	--	Folder
▶ 9ABSTRACTS AND PRESENTATIONS	28 Feb 2019 at 10.52	--	Folder
▶ 10LUENNOT JA OPETUS	8 May 2019 at 21.38	--	Folder
▶ 11POPULAR SCIENCE	17 Jan 2019 at 21.34	--	Folder
▶ Articles	21 Dec 2018 at 12.27	--	Folder
▶ Bernasconi_thesis	6 Jul 2016 at 14.48	--	Folder
▶ elokuvia	18 Dec 2017 at 4.44	--	Folder
▶ Jonna's documents_old	Today at 13.08	--	Folder
▶ Photos	14 May 2019 at 0.15	--	Folder
▶ photos_from_anytrans	15 Nov 2017 at 23.12	--	Folder
▶ R	31 Jan 2019 at 9.03	--	Folder
▶ Sample lists	20 Jul 2015 at 12.03	--	Folder
▶ Team speciant	12 Mar 2018 at 13.53	--	Folder

Gutes Beispiel:



Regeln für Ordnerstrukturen

- Klare, hierarchische Strukturen orientiert an Inhalten und Vorgängen
- Kurze nachvollziehbare Namen
- 🗂️✖️👍 : 7 Ordner in die Breite, 3 Ordner in die Tiefe



Probleme bei Namen

news_release_2015-07-14.doc
news_release_2015-07-15.doc
news_release_2015-07-16.doc
news_release_7172015.doc
Untitled-1.txt
Untitled-2.txt
Untitled-3.txt

COMPANY_LOGO_SMALL.JPG

Picture#2_v01!.jpg

photo.jpg

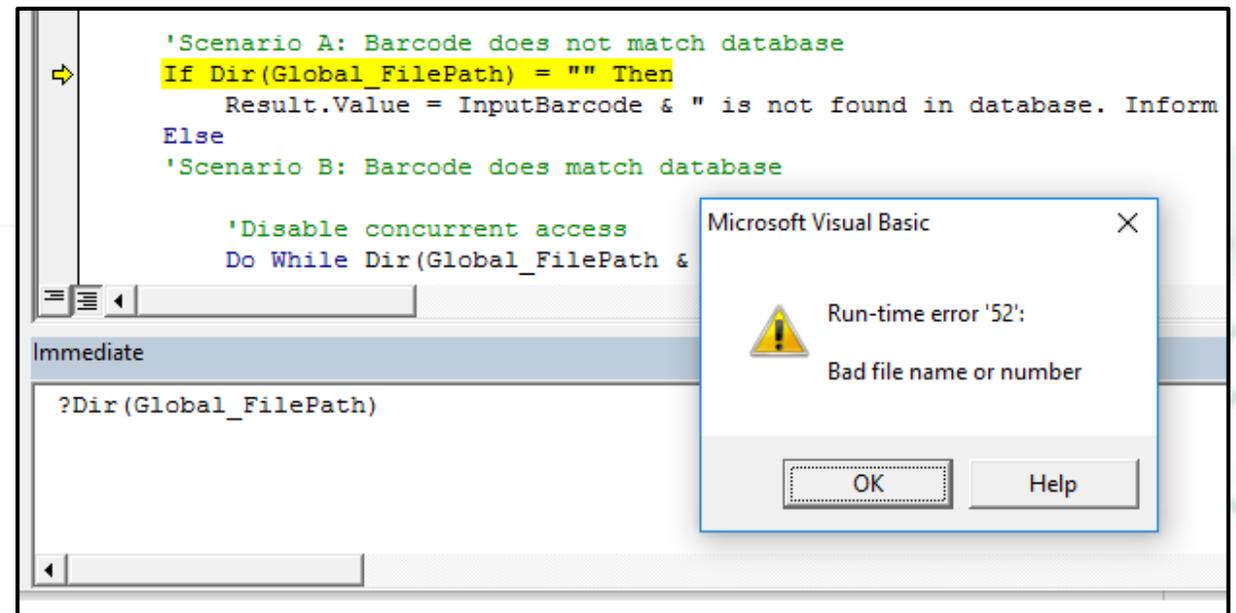
393923022_30252433004.png

mockup_final.psd
mockup_v6_final.psd
mockup_v6_final2.psd
mockup2_v6_final_revised3.psd

index-copy.html
index-copy-2.html
copy-of-index-copy.html

projects.php
projects-david.php

Quelle: 20px.com



Quelle: stackoverflow.com

Namenskonventionen

Technisch:

- Keine **Sonderzeichen** (wie !?:#,*,~.)
- Keine **Umlaute** oder ähnliches (ü, ö, ä, ß, é, ç, ...)
- "_" oder "-" anstatt **Leerzeichen**
- Keine **Großbuchstaben**
- Max. **256 Zeichen** (inkl. Pfad zur Datei)

➤ Beispiele:

- vfu_uebersicht_v04.csv
- 20201104_tkfdm_fact_sheet_open_source_software.pdf

Inhaltlich:

- Veröffentlichungsdatum
- Autor*in, Organisation
- Titel
- Kategorie
- Version (v01..v99, submission, alpha, final, ...)

Five S: 3) Shine

- Ziel:
 - Qualität erhalten und ggf. anpassen
 - Dokumentieren und verständlich halten
- Umsetzung:
 - Vorgänge kontrollieren und selber durchsetzen
 - Regelmäßige Routinen



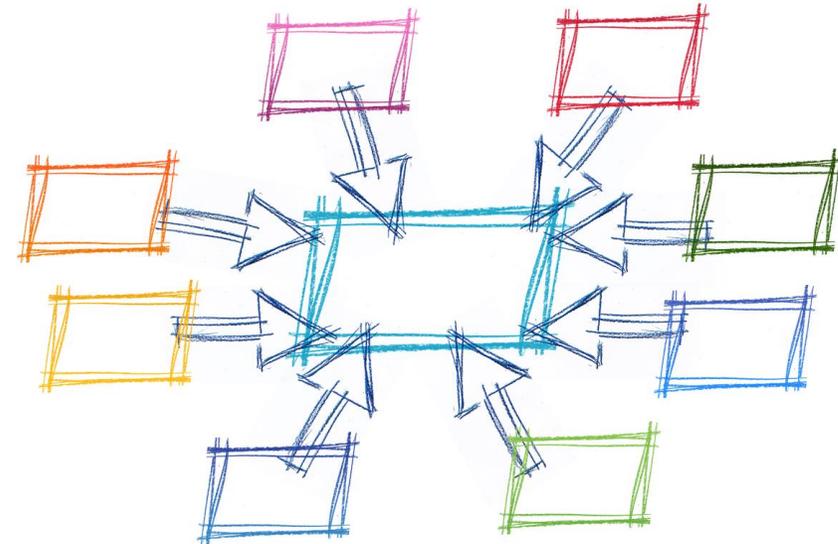
Routinen einführen



- ✓ Kein Datenmüll auf Desktop?
- ✓ Alles zugeordnet?
- ✓ Richtige Benennung?
- ✓ ToDos markiert?

Five S: 4) Standardize

- Ziel:
 - Prozesse und Termine etablieren, um erste 3 "S" zu ermöglichen
- Umsetzung:
 - Best-Practices, Leitlinien und Regeln dokumentieren
 - Diskussion mit Kolleg*innen, Verantwortlichkeiten klären



Standard Operating Procedure (SOP)

- Struktur der Ordner
- Erlaubte Formate
- Bezeichnung der Ordner und Dateien
- Ablegungsorte und Beschaffungsmaßnahmen
- Rechte und Pflichten zur Nutzung und Erweiterung
- Anwendung internationaler und/oder disziplinspezifischen Standards



Organisationen für Standardfestlegungen



schema.org

NIST
National Institute of
Standards and Technology



W3C[®]

Standards für Metadaten finden...



RESEARCH DATA ALLIANCE

<http://rd-alliance.github.io/metadata-directory/>
<https://rdamsc.bath.ac.uk/>



<https://www.dcc.ac.uk/guidance/standards/metadata>



<https://terminology.tib.eu/ts>

Five S: 5) Sustain

- Ziel:
 - Mit Selbstdisziplin und Gewohnheit entwickelte Organisation erhalten
- Umsetzung:
 - Workshops, Übungseinheiten
 - Einweisung neuer Mitarbeiter*innen
 - Neue Methoden/Technologien



Workshops und Feedback

- Regelmäßige Workshops für Vorstellung der Organisation
- Umgang mit Werkzeugen erklären (z.B. Konsolenbefehle, GitLab, LaTeX, eLabFTW)
- Neue Methoden und Technologien diskutieren



Handreichung zum 5S Data Modell



The 5S Methodology in Research Data Management

Structuring and maintaining file organization

Developed by the Thuringian Competence Network for Research Data Management

Contributors: Kevin Lang (Bauhaus-Universität Weimar)
Annett Schrötter (Friedrich-Schiller Universität Jena)
Roman Gerlach (Friedrich-Schiller Universität Jena)
Jessica Rix (Technische Universität Ilmenau)
Nadine Neute (Universität Erfurt)
Anne Lehmann (Universität Erfurt)

Version: 1.0, 08.02.2021
DOI: 10.5281/zenodo.4494258

All texts are licensed as:
[CC BY 4.0 International](https://creativecommons.org/licenses/by/4.0/)

All comics are licensed as:
[CC BY-NC 4.0 Attribution-NonCommercial 4.0 International](https://creativecommons.org/licenses/by-nc/4.0/)



doi.org/10.5281/zenodo.4494257

3-2-1 Backup Regel: Ursprung

- Historische goldene Regel in Unternehmen zur Organisation von Backups
- 2009 von Fotograf Peter Krogh formuliert
 - 3 Kopien
 - 2 verschiedene Technologien
 - 1 externer Standort
- Erweitert als 3-2-1-0 oder 3-2-1=0 Backup Regel
 - 0 Fehler bei Wiederherstellung

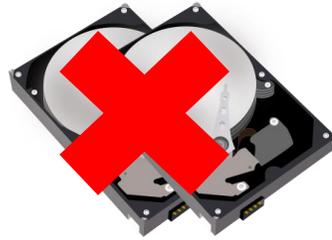


3-2-1 Backup Regeln: 3 Kopien

- 1 Arbeitskopie auf dem Arbeitsplatzrechner, 2 Backup-Kopien
- Warum 3? **Wahrscheinlichkeitsrechnung!**



**1 zu
100**



**1 zu
10 000!**



**1 zu
1000 000!!!**

- im privaten Gebrauch meist 2 Kopien ausreichend
- bei Institutionen und Gewerbe 3 oder mehr Kopien

3-2-1 Backup Regeln: 2 verschiedene Technologien

- unterschiedliche Fehleranfälligkeiten bei verschiedenen Speichertechnologien und Medientypen
- Unterschiede in:
 - Anzahl von Schreib-/Lesezyklen
 - Haltbarkeit des Materials
 - Schnittstellen
 - Sicherheitsaspekte
- durch Medienbruch lässt sich Datenverlust verringern



3-2-1 Backup Regeln: 1 externer Standort

- Was passiert, wenn alle Kopien an einem Ort sind?
- Mögliche Probleme:
 - Brand
 - Überschwemmung
 - Stromausfall
 - Hacker-Angriffe
 - Netzwerkbefehle/
Softwareupdates
- Getrennte Räume und Technik



3-2-1 Backup Regeln: 0 Fehler

- 0 Fehlertoleranz bei der Wiederherstellung von Dateien
- Welche Probleme können passieren?
 - Vorgang/Kontakt
 - Backup/Versionierung deaktiviert (Default Settings)
 - zu große Intervalle
 - Art und Stand der gesicherten Dateien
- Immer System testen und richtig konfigurieren



FAIR-Prinzipien: Ursprung

- Erste Veröffentlichung im März 2016
 - **F**indable (Auffindbarkeit)
 - **A**ccessible (Zugänglichkeit)
 - **I**nteroperable (Anwendbarkeit)
 - **R**eusable (Wiederverwendbarkeit)
- Hohe Akzeptanz in vielen Organisationen (G20, GO FAIR, CODATA, RDA, DFG, ...)
- Versuche zur Erweiterung:
 - Nach Fachbereichen/
Anwendungen
 - CARE-Prinzipien
 - Open Data



FAIR-Prinzipien: Findable

- Ziel:
 - Die Daten sollten mit viel Kontext auffindbar sein.
- Umsetzung:
 - Persistenter Identifikator
 - Viele Metadaten
 - Repository (→ re3data.org)
 - Auffindbar in Suchmaschinen oder Register



Das richtige Repository finden...

- Suche über:  **re3data.org**
REGISTRY OF RESEARCH DATA REPOSITORIES



PID Datensatz



PID Autor*innen



Metadaten



Download/Export



Beschreibung



Zugriff



Lizenzen



Überblick/Vorschau



Versionierung



Anmeldung/Bearbeitung



Suchmaschinen

Generelle Forschungsdaten-Repositoryen



DRYAD



figshare



Persistente Identifikatoren

- Autor*innen: ORCID, ResearcherID
- Publikation: DOI, URN, HDL
- ***neu*** Einrichtungen: ROR

- Vorteile:

- Plattform übergreifend
- immer abrufbar (eindeutige Resource)
- eindeutig auch nach Umzug, Institutionswechsel oder Namensänderung
- Auflistung von Metadaten



RESEARCHERID



Metadaten

The screenshot shows a metadata page for a 'Fact Sheet: Research Data Repositories'. The page is dated March 16, 2020, and is in English. It includes a definition of research data repositories, a list of quality features, a table of files, and a list of versions. The page is part of the Zenodo digital library.

March 16, 2020

Fact Sheet: Research Data Repositories

Gerlach, Roman, Rex, Jessica, Lang, Kevin, Neude, Nadine, Schwartz, Volker

English:

Repositories for research data are used to publish or archive digital data from your own research and thereby make it accessible to interested parties. The motivation for this is that research results from publications can be understood and reproduced with the help of the published research data, but can also be used for new research questions if necessary. They also serve as good support to meet the guidelines of the FAIR principles.

With this fact sheet, we would like to provide all interested parties with important information about research data repositories. First, the quality characteristics of a good repository are discussed and free which services and offers are available to the researchers free of charge. This includes the register for subject-related repositories re3data.org, as well as interdisciplinary repositories such as [Zenodo](http://Zenodo.org), [Dryad](http://Dryad.org) or the Digitale Bibliothek Thüringen.

German:

Repositorien für Forschungsdaten dienen dazu, digitale Daten aus der eigenen Forschung zu veröffentlichen oder zu archivieren und sie damit für Interessierte zugänglich zu machen. Die Motivation hierfür ist, dass Forschungsergebnisse aus Publikationen mit Hilfe der veröffentlichten Forschungsdaten nachvollziehbar und reproduzierbar werden, ggf. aber auch für neue Forschungsfragen verwendet werden können. Weiterhin dienen sie auch als gute Stütze um die Leitlinien der FAIR-Prinzipien zu erfüllen.

Mit dieser Handreichung möchten wir allen Interessierten wichtige Informationen rund um das Thema Forschungsdatenrepositorien zur Verfügung stellen. Dabei wird zuerst auf die Qualitätsmerkmale eines guten Repositoriums eingegangen und dann welche Services und Angebote für die Forscher kostenfrei zur Verfügung stehen. Darüber stellt das Register für fachbezogene Repositorien re3data.org, als auch fachübergreifende Repositorien wie z.B. [Zenodo](http://Zenodo.org), [Dryad](http://Dryad.org) oder die Digitale Bibliothek Thüringen.

fix minor layout issues, justified text, graphic positions

Preview

Repositories for research data

Definition

Repositories for research data are used to publish or archive digital data from our research and to make it accessible to interested parties. The motivation for this is that research results from publications become traceable and reproducible with the help of the published research data. It is also possible to publish a research dataset without a corresponding publication so that as many other researchers as possible can use it for their research questions.

The publication of the research data of a project is already required or recommended by many third-party funders and is part of good scientific practice. When choosing a data repository, attention should be paid to various quality features that can contribute to the understanding and visibility of the data set.

Quality features for repositories

Files (2669 KB)

Name	Size	Preview	Download
fact_sheet_research_data_repositories.pdf	431.0 KB	Preview	Download
fact_sheet_research_data_repositories.pdf	431.0 KB	Preview	Download

Citations (0)

Show only: Literature (0) Dataset (0) Software (0) Unknown (0)

Citations to this version

No citations.

Communities

Thüringer Kompetenzzentrum Forschungsdatenmanagement

240 views **204** downloads

OpenAIRE

Publication date: March 16, 2020

DOI: [10.5281/zenodo.3900922](https://doi.org/10.5281/zenodo.3900922)

Keyword(s): research data repositories, Forschungsdatenrepositorien, FAIR, research data, data, Datenmanagement, Open Access, Open Access, Open Access, Open Access

Communities: Thüringer Kompetenzzentrum Forschungsdatenmanagement

License (for files): Creative Commons Zero v1.0 Universal

Versions

Version	Created	Mar 16, 2020
Version 2	10.5281/zenodo.3900922	Mar 16, 2020
Version 1	10.5281/zenodo.3760348	Mar 16, 2020

Cite all versions? You can cite all versions by using the DOI 10.5281/zenodo.3760327. This DOI represents all versions, and will always resolve to the latest one. [Read more](#).

Share

Cite as

Gerlach, Roman, Rex, Jessica, Lang, Kevin, Neude, Nadine, & Schwartz, Volker. (2020, March 16). Fact Sheet: Research Data Repositories. Zenodo. <http://doi.org/10.5281/zenodo.3900922>

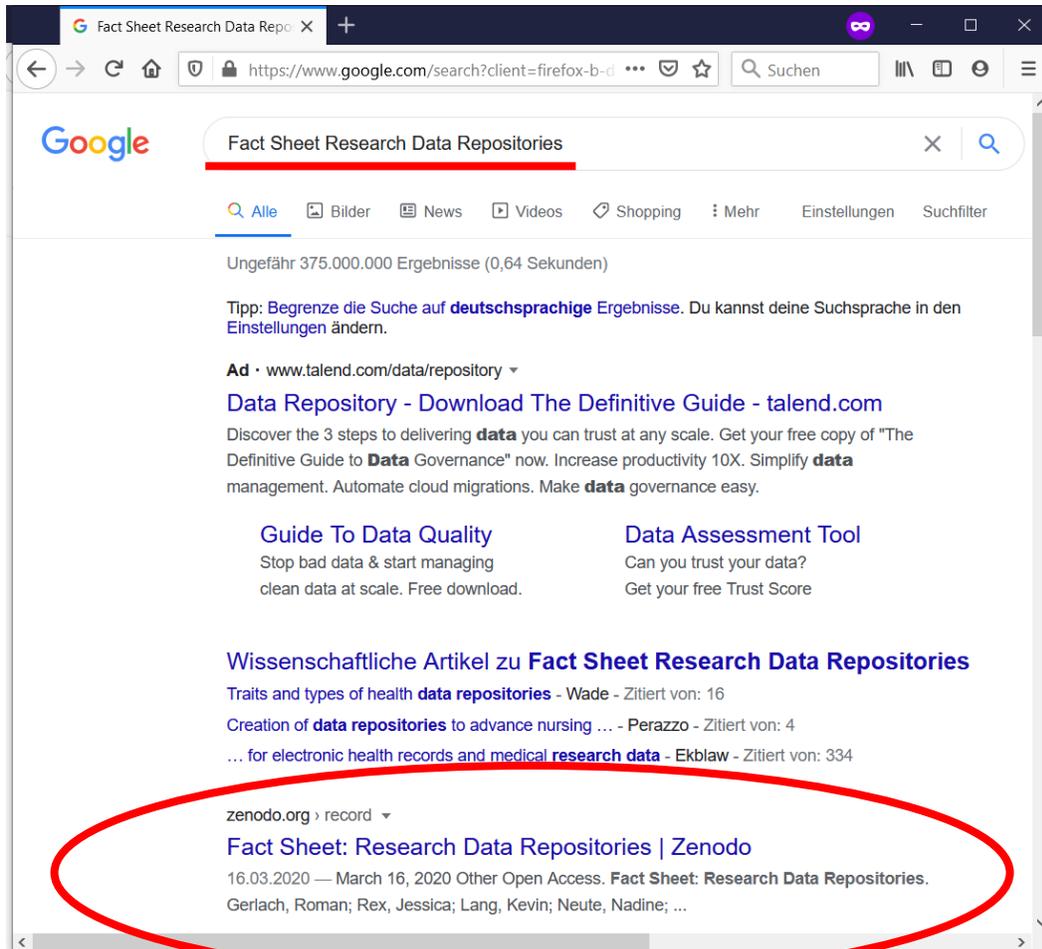
Start typing a citation style.

Export

BibTeX, CSL, DataCite, Dublin Core, DCAT, JSON, JSON-LD, GeoJSON, MARCXML, Merdley

- Persistenter Identifikator
- Titel
- Autor*innen (mit ID wie ORCID)
- Veröffentlichungsdatum
- Publikationen
- Stichwörter
- Kategorien/Gruppen
- Version
- Zugriffszahlen
- ...

In Suchmaschinen auffindbar



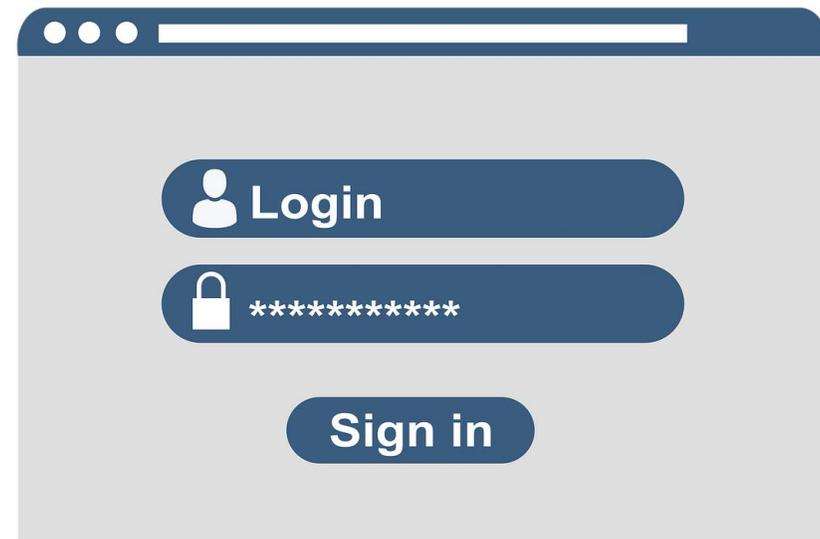
➤ mit Suche (wie Titel oder anderen Metadaten) im Repositorium finden

➤ Suchmaschinen:

- Google, Bing, ...
- [Google Dataset Search](#)
- [OpenAIRE Search](#)

FAIR-Prinzipien: Accessible

- Ziel:
 - Klar definierter Zugang für die Daten
- Umsetzung:
 - Downloadmöglichkeiten oder API
 - Authentifizierungs- bzw. Autorisierungsschritte
 - Bedingungen
 - Bestehende Metadaten



Downloadmöglichkeiten

- HTML:

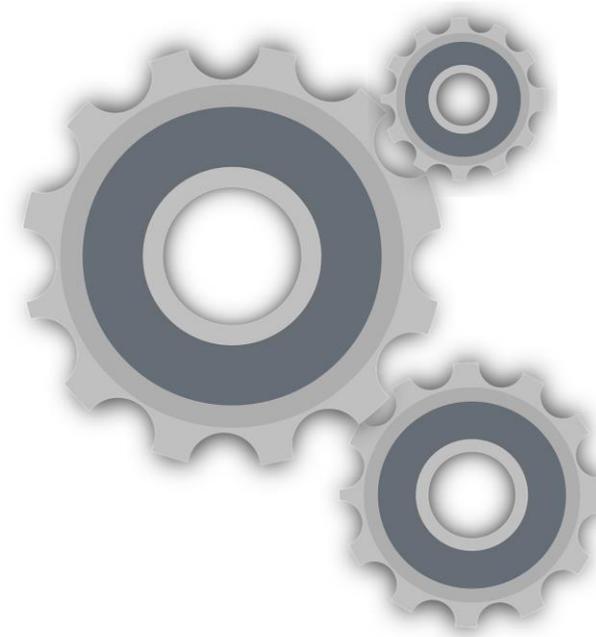
Files (868.9 kB)		
Name	Size	
fact_sheet_research_data_repositories.pdf	431.0 kB	Preview Download
md5:786a3c1dedcc4db99d44b2fd6c1464da ⓘ		
handreichung_forschungsdatenrepositorien.pdf	437.9 kB	Preview Download
md5:1b7b371c218dc7cd84483078e934833b ⓘ		

- API:

```
{
  "key": "my-file.zip",
  "mimetype": "application/zip",
  "checksum": "md5:2942bfabb3d05332b66eb128e0842cff",
  "version_id": "38a724d3-40f1-4b27-b236-ed2e43200f85",
  "size": 13264,
  "created": "2020-02-26T14:20:53.805734+00:00",
  "updated": "2020-02-26T14:20:53.811817+00:00",
  "links": {
    "self": "https://zenodo.org/api/files/44cc40bc-50fd-4107-b347-00838c79f4c1/dummy_example",
    "version": "https://zenodo.org/api/files/44cc40bc-50fd-4107-b347-00838c79f4c1/dummy_example",
    "uploads": "https://zenodo.org/api/files/44cc40bc-50fd-4107-b347-00838c79f4c1/dummy_example"
  },
  "is_head": true,
  "delete_marker": false
}
```

FAIR-Prinzipien: Interoperable

- Ziel:
 - Die Datenformate und Metadaten entsprechen üblichen Standards
- Umsetzung:
 - Verwendung offener Formate
 - Vokabular nach Standards
 - Referenzen zu verwandten (Meta-)Daten



Offene Formate

Dateityp	Offene Formate	Proprietäre Formate
Texte	TXT, TEX, ODT, HTML, RTF	DOC, DOCX
Tabellen	CSV, TSV	XLS, SAV
Bild	PNG, SVG	PSD, AI, PUB
Audio	VORBIS, FLAC, OPUS	WMA, MP3
Video	FFmpeg, OpenH264, Xvid	MPEG-2, MPEG-4, MVC

- Text-basierend frei lesbar (keine binären Formate)
- Kann mit jeder Software verarbeitet werden
- Geeignet für Archivierung

FAIR-Prinzipien: Reusable

- Ziel:
 - Dokumentation und Verwendungsmöglichkeiten
- Umsetzung:
 - Datennutzungslizenz vergeben
 - Beschreibung der Erstellung/Herkunft und Veränderung
 - Verweis auf benötigte Anwendungen



Lizenzen

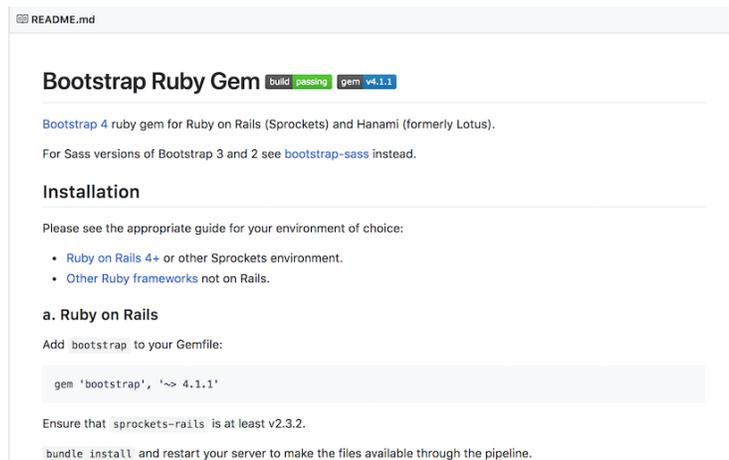
- Def.: Festlegung der Nutzungsbedingungen von Daten
- Vorlagen: CC (Creative Commons), ODC (Open Data Commons), GPL (GNU General Public Licences)

Icon	Kürzel	Beschreibung
	BY	Attribution / Namensnennung
	NC	Non commercial / Nicht kommerziell
	ND	No derivatives / Keine Bearbeitung
	SA	Share alike / Weitergabe unter gleichen Bedingungen

Dokumentation

- Zuhilfenahme durch Datenmanagementplan
- Ausführung der Dokumentation z.B. in:

➤ README.md



Bootstrap Ruby Gem build passing gem v4.1.1

Bootstrap 4 ruby gem for Ruby on Rails (Sprockets) and Hanami (formerly Lotus).

For Sass versions of Bootstrap 3 and 2 see [bootstrap-sass](#) instead.

Installation

Please see the appropriate guide for your environment of choice:

- Ruby on Rails 4+ or other Sprockets environment.
- Other Ruby frameworks not on Rails.

a. Ruby on Rails

Add bootstrap to your Gemfile:

```
gem 'bootstrap', '~> 4.1.1'
```

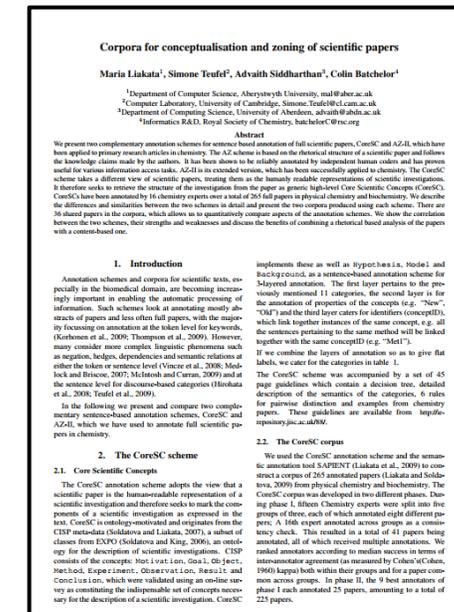
Ensure that sprockets-rails is at least v2.3.2.

`bundle install` and restart your server to make the files available through the pipeline.

➤ Wiki



➤ Publikation



Corpora for conceptualisation and zoning of scientific papers

Maria Liakata¹, Simone Teufel², Advait Siddharthan³, Colin Batchelor⁴

¹Department of Computer Science, Aberystwyth University, mal@aber.ac.uk
²Computer Laboratory, University of Cambridge, Simone.Teufel@cl.cam.ac.uk
³Department of Computing Science, University of Aberdeen, advait@abdn.ac.uk
⁴Informatics R&D, Royal Society of Chemistry, batchelc@rsc.org

Abstract

We present two complementary annotation schemes for sentence based annotation of full scientific papers, CoreSC and AZ-IL, which have been applied to primary research articles in chemistry. The AZ scheme is based on the rhetorical structure of a scientific paper and follows the knowledge claims made by the authors. It has been shown to be reliably annotated by independent human coders and has proven useful for various information access tasks. AZ-IL is its extended version, which has been successfully applied to chemistry. The CoreSC scheme takes a different view of scientific papers, treating them as the humanly readable representation of scientific investigations. It therefore seeks to retrieve the structure of the investigation from the paper as generic high-level Core Scientific Concepts (CoreSC). CoreSC has been annotated by 16 chemistry experts over a total of 265 full papers in physical chemistry and biochemistry. We describe the differences and similarities between the two schemes in detail and present the two corpora produced using each scheme. There are 26 abstract papers in the corpora, which allows us to quantitatively compare aspects of the annotation schemes. We show the correlation between the two schemes, their strengths and weaknesses and discuss the benefits of combining a rhetorical based analysis of the papers with a content-based one.

1. Introduction

Annotation schemes and corpora for scientific texts, especially in the biomedical domain, are becoming increasingly important in enabling the automatic processing of information. Such schemes look at annotating mostly abstracts of papers and less often full papers, with the majority focusing on annotation at the token level for keywords (Greenhouse et al., 2009; Thompson et al., 2009). However, many consider more complex linguistic phenomena such as negation, hedges, dependencies and semantic relations at either the token or sentence level (Vince et al., 2008; Medlock and Brinco, 2007; McInish and Carran, 2009) and at the sentence level for discourse-based categories (Hoshida et al., 2008; Teufel et al., 2009).

In the following we present and compare two complementary sentence-based annotation schemes, CoreSC and AZ-IL, which we have used to annotate full scientific papers in chemistry.

implements these as well as Hypothesis (i.e. *Model*), and *Background*), as a sentence-based annotation scheme for 3-layered annotation. The first layer pertains to the previously mentioned 11 categories, the second layer is for the annotation of properties of the concepts (e.g. "New", "Old") and the third layer caters for identifiers (conceptID), which link together instances of the same concept, e.g. all the sentences pertaining to the same method will be linked together with the same conceptID (e.g. "Method").

If we combine the layers of annotation so as to give flat labels, we cater for the categories in table 1.

The CoreSC scheme was accompanied by a set of 45 page guidelines which contain a decision tree, detailed description of the semantics of the categories, 9 rules for pairwise distinction and examples from chemistry papers. These guidelines are available from <http://mposney.jku.ac.uk/RW>.

2. The CoreSC scheme

2.1. Core Scientific Concepts

The CoreSC annotation scheme adapts the view that a scientific paper is the human-readable representation of a scientific investigation and therefore seeks to mark the components of a scientific investigation as expressed in the text. CoreSC is ontology-motivated and originates from the CISP meta-data (Sokolova and Liakata, 2007), a subset of classes from EXPO (Sokolova and King, 2006), an ontology for the description of scientific investigations. CISP consists of the concepts: *Topic*, *Level*, *Goal*, *On-ject*, *Method*, *Experiment*, *Observation*, *Result* and *Conclusion*, which were validated using an online survey as constituting the indispensable set of concepts necessary for the description of a scientific investigation. CoreSC

2.2. The CoreSC corpus

We used the CoreSC annotation scheme and the semantic annotation tool S-PRESENT (Liakata et al., 2009) to construct a corpus of 265 annotated papers (Liakata and Sokolova, 2009) from physical chemistry and biochemistry. The CoreSC corpus is developed in two different phases. During phase 1, fifteen Chemistry experts were split into five groups of three, each of which annotated eight different papers. A 16th expert annotated across groups as a consistency check. This resulted in a total of 41 papers being annotated, all of which received multiple annotations. We ranked annotators according to median success in terms of inter-annotator agreement (as measured by Cohen's Kappa, 1960) kappa) both within their groups and for a paper common across groups. In phase II, the 9 best annotations of phase I each annotated 25 papers, amounting to a total of 225 papers.

Vielen Dank für Ihre Aufmerksamkeit.

Weiterführende Veranstaltungen (aufgezeichnete Coffee Lectures)

- Rechte an Forschungsdaten
- Versionierung mit Git - Die Einführung
- Kollaborative Plattformen - Was dem Einzelnen nicht möglich ist, das schaffen viele
- Digitale Daten - Der richtige Typ zu ihrem Format!
- Forschungsdatenbeschreibung - Das Anlegen von Codebooks und Readme-Dateien

Werbung

- Postkarten, Kulis, Tassen, Webcam-Cover, ...
- ScaryTales Kartenspiel



Quellen

- Forschungsdatenmanagement und das TKFDM:
 - [Portal von TKFDM](#)
 - [TKFDM Community auf Zenodo](#)
 - [Research Data Scarytales](#)
 - [Data Sharing and Management Snafu in 3 Short Acts](#)
 - [FDM kurz erklärt \(TIB\)](#)
 - [#2MinutenFDM](#)
- Datenmanagementplan:
 - Tabelle Drittmittelgeber: [20200701_fdmtage_weimar_kerstin_helbig.pdf](#)
 - Inhalte: [se_rdm_practical_guide_final.pdf](#)
 - Grafiken: [essentials-for-a-data-management-plan-for-spectroscopists](#)
- Stockimages mit CC0 Lizenz von:
 - [pixabay.com](#)
 - [unsplash.com](#)

Quellen

- Five S DATA:
 - <https://www.helsinki.fi/en/research/organizing-data-folders-with-5sdata-method>
 - <https://kanbanize.com/de/lean-management-de/wert-verschwendung/was-sind-die-5s-in-lean>
- 3-2-1 Backup:
 - <https://www.storage-insider.de/was-ist-die-3-2-1-backup-regel-a-782641/>
 - <https://spanning.com/blog/simplifying-the-3-2-1-rule-for-data-protection-in-the-age-of-the-cloud/>
- FAIR Principles:
 - [The FAIR Guiding Principles for scientific data management and stewardship](#)
 - [FAIR Principles auf go-fair.org](#)
 - [How FAIR are your data?](#)
 - <https://www.gida-global.org/care>
 - [Making FAIReR assessments possible](#)