

# Forschungsdatenmanagement im Überblick

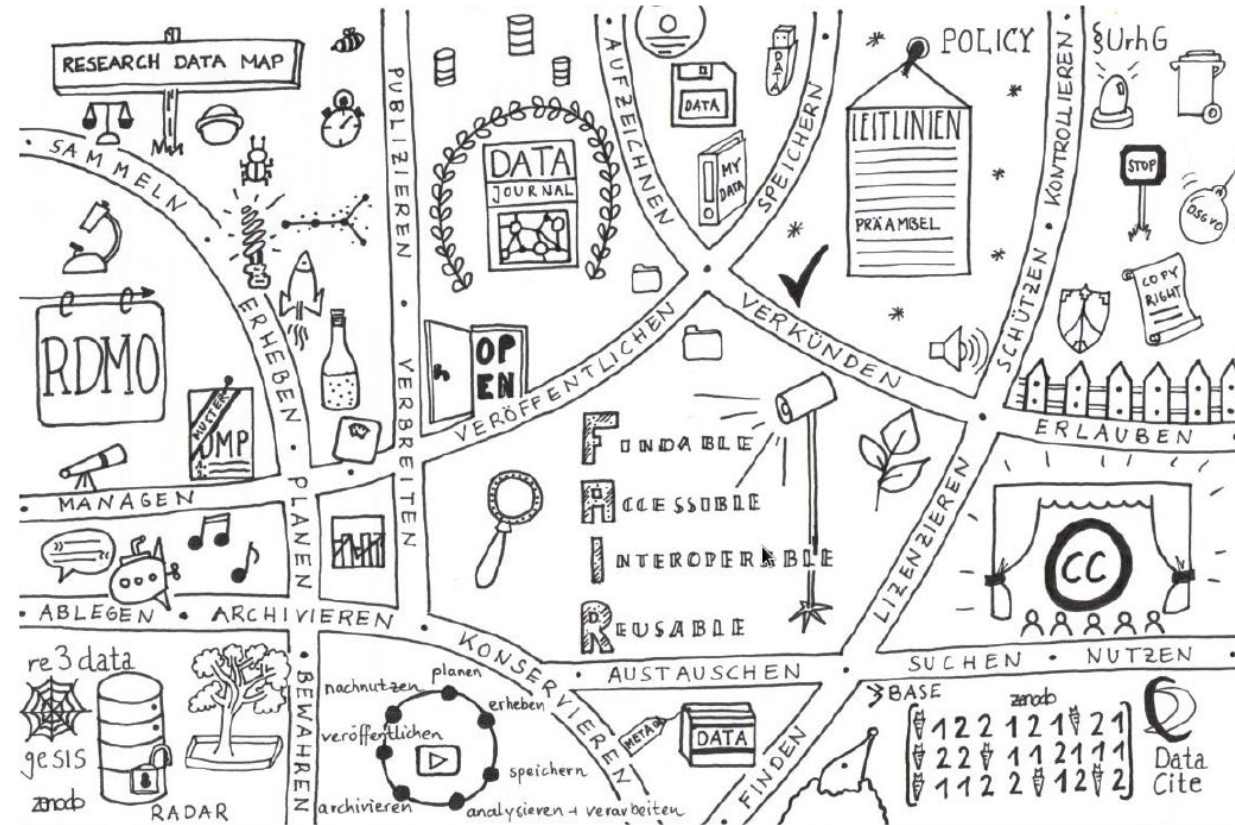


Bild von Martina Jackenroll aus „Wie organisiere ich meine Dateien und Ordner? Einführung in das Forschungsdatenmanagement“ lizenziert unter [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)

**Forschungsdaten sind alle Daten, die im wissenschaftlichen Forschungsprozess entstehen, verarbeitet werden oder dessen Ergebnis sind.**

In der [Checkliste zum Umgang mit Forschungsdaten](#) zählt die DFG einige Beispiele für Forschungsdaten auf:

„Zu Forschungsdaten zählen u. a. Messdaten, Laborwerte, audiovisuelle Informationen, Texte, Surveydaten oder Beobachtungsdaten, methodische Testverfahren sowie Fragebögen. Korpora, Software und Simulationen können ebenfalls zentrale Ergebnisse wissenschaftlicher Forschung darstellen und werden daher ebenfalls unter den Begriff Forschungsdaten gefasst.“

Forschungsdatenmanagement (FDM) umfasst alle Maßnahmen im Lebenszyklus von Forschungsdaten: von der Planung, Erhebung, Speicherung, Dokumentation, Aufbereitung und Analyse der Daten bis zu ihrer Veröffentlichung, Archivierung und Nachnutzung.

## Warum FDM? (Teil 1)

- Teil der guten wissenschaftlichen Praxis
- Erfüllen der Vorgaben von Forschungsförderern und Verlagen
- Reproduzierbarkeit von Forschungsergebnissen
- Glaubwürdigkeit, Nachvollziehbarkeit der eigenen Forschung durch Dritte
- Vereinfachung zukünftiger Nachnutzung der eigenen Daten und Interpretierbarkeit der Daten langfristig sicherstellen

## Warum FDM? (Teil 2)

- Minimieren des Datenverlustrisikos
- Leichtere Zusammenarbeit mit Projektpartnern
- Referenzierbarkeit durch die Nutzung von persistenten Identifikatoren
- Erhöhung der eigenen Sichtbarkeit
- Klärung von rechtlichen und ethischen Aspekten des eigenen Forschungsvorhabens
- Unterstützung von Open Science

## Daten nachnutzen

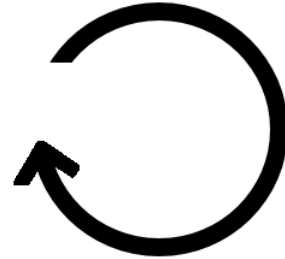
- Daten bzw. Schlussfolgerungen prüfen
- Daten für Lehre oder Forschungsvorhaben nutzen

## Daten archivieren

- endgültige Datenversion auswählen
- Ort für Langzeitarchivierung auswählen (Datenträger oder Repositoryum)

## Vorhaben planen

- Datenmanagement planen
- DMP erstellen
- Recherche nach bereits vorhanden Daten



- rechtliche und ethische Aspekte klären
- Lizenzen festlegen
- Veröffentlichungsort (Repositoryum, Journal) auswählen

## Daten veröffentlichen

## Daten erheben

- Experimente, Befragungen, Simulationen, usw. durchführen
- Erlaubnis zur Datennutzung einholen

## Daten verarbeiten

- Daten erfassen, digitalisieren, transkribieren, übersetzen
- Daten prüfen, validieren, bereinigen, anonymisieren
- Daten dokumentieren und mit Metadaten versehen
- Daten verwalten und sichern
- Daten für die Langzeitarchivierung vorbereiten (Dateiformate)
- Daten interpretieren und Schlussfolgerungen ableiten

Ein DMP ist ein „lebendes“ Dokument, welches vor Projektbeginn erstellt und während der Projektlaufzeit fortlaufend erweitert und ergänzt werden sollte.

## DMP der UB Chemnitz

Die meisten DMPs sind als Fragenkatalog aufgebaut.

Zum Beispiel:

Angaben zum Projekt; Art der Daten; Verfahren der Speicherung, Archivierung und Veröffentlichung der Daten; Rechtliche Aspekte; Datenverantwortlicher

## Anforderungen von Drittmittelgebern

- DFG, BMBF: Bei der Antragstellung sind Überlegungen zum FDM notwendig
- Horizon Europe: DMP innerhalb von 6 Monaten nach dem Beginn der Förderung verpflichtend.

## Tools zur DMP Erstellung

- [RDMO](#)
- [ARGOS](#)
- [DataWiz](#)
- [DMPOnline](#)
- [DMPTool](#)

## Findable (Auffindbar)



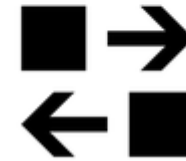
Die Daten und ihre Metadaten sind für Menschen und Maschinen einfach und eindeutig auffindbar, wobei vor allem maschinenlesbare Metadaten wichtig sind.

## Accessible (Zugänglich)



Der Zugang zu den Daten ist über ein standardisiertes Kommunikationsprotokoll eindeutig geregelt. Ferner muss der Zugang zu den Metadaten dauerhaften Bestand haben.

## Interoperable (Interoperabel)



Die Daten und ihre Metadaten sind aufgrund der Verwendung formaler Standards mit anderen Daten kombinierbar, wodurch der automatisierte Austausch zwischen Computeranwendungen möglich ist.

## Reusable (Wiederverwendbar)



Für eine gute Nachvollziehbarkeit und Reproduzierbarkeit der Daten sind diese sowie die Metadaten reichhaltig zu dokumentieren. Dazu gehört auch die Vergabe einer verständlichen Nutzungslizenz.

**Definition** von Offenen Daten nach der „[Open Knowledge Foundation Deutschland](#)“:

„Offene Daten sind Daten, die von jedermann frei verwendet, nachgenutzt und verbreitet werden können – maximal eingeschränkt durch Pflichten zur Quellennennung und dem Prinzip des „share-alike“, d.h. Daten nur unter unveränderten urheberrechtlichen Nutzungsmöglichkeiten zu teilen. Davon ausgenommen sind persönliche Daten.“

Wichtige **Eigenschaften** von Offenen Daten:

- Verfügbarkeit und Zugang
- Wiederverwendung und Nachnutzung
- Universelle Beteiligung



## **FAIRe Daten sind keine offenen Daten!**

Während offene Daten immer frei verfügbar und zugänglich sein müssen, kann der Zugang zu FAIRen Daten beschränkt sein. Die FAIR-Prinzipien fordern lediglich, dass der Zugang klar geregelt sein muss. Gibt es zum Beispiel nur einen frei zugänglichen Kontakt, bei dem die Datennutzung erfragt werden kann, so sind die Daten FAIR aber nicht offen.

## **Es gilt der Grundsatz:**

**„so offen wie möglich, so eingeschränkt wie nötig“**

- Auswahl eines Datenablageortes hinsichtlich Speicherkapazität, Sicherheitsstandards und Zugriffsverwaltung
- hierarchische Ordnerstruktur für mehr Übersichtlichkeit (Einzelordner nicht überfüllen)
- klare und strukturierte Dateibenennung (Gleichgewicht zwischen einer inhaltsspezifischen Bezeichnung und der Länge des Dateinamens)
- über Versionierung die Veränderungen an einer Datei über die Projektzeitlauf hinweg protokolliert (über ReadMe-Datei oder direkt bei der Dateibenennung)
  - Die TU Chemnitz unterstützt den Dienst [GitLab](https://www.gitlab.com/), wodurch die Versionskontrolle mittels Git möglich ist.
- Datenorganisation unter den Projektpartnern abstimmen, damit man jederzeit nachvollziehen kann, was, wie und weshalb getan wurde.

Damit Projektpartner, Dritte oder auch man selbst zu einem späteren Zeitpunkt die Daten eines Forschungsvorhabens nachvollziehen kann, sollten alle Schritte im Zusammenhang mit der Erstellung und Bearbeitung der Daten nachvollziehbar dokumentiert werden.

## Metadaten

- Maschinenlesbare strukturierte Daten, die bestimmte Merkmal der Daten beschreiben, z.B. Titel der Datei, Name des Primärforschers/Urhebers, Projektnummer, Zeitraum und Ort der Datenerhebung usw.
- Es gibt bereits einheitliche vordefinierte Standards, z.B. [Dublin Core](#).



← **rich metadata** →  
 41 Informationen

← **poor metadata** →  
 2 Informationen



Bild aus [Open und FAIR Data: Der Beitrag der SAGW](#) von Beat Immenhauser lizenziert unter [CC BY 4.0](#)

Durch eine geeignete Datensicherung sollen die Daten vor Verlust sowie unberechtigten Manipulationen und Zugriffen geschützt werden.

- Geeignete Speichermedien (PC, Laptop, USB-Stick, Cloud) und Nutzungsorte (Speicherung nur lokal am Arbeitsplatz oder auch Zugriff von zu Hause oder unterwegs) auswählen
  - [Speicherdienste](#) der TU Chemnitz, z.B. AFS-Projektverzeichnis oder TUCcloud
- Backup-Strategie einhalten (Fragen: Welche Daten sollen auf jeden Fall gesichert werden? Wie oft werden Sicherheitskopien angelegt?)
  - 3-2-1 Backupregel (mindestens 3 Kopien, an mindestens 2 verschiedenen Orten und einer davon dezentral)
  - Backup-Dienst der TU Chemnitz: [RSYNC](#)
- Verschlüsselung von Dateien / Verzeichnissen durchführen und geeignete Passwörter nutzen

Nach dem Projektende sollte eine endgültige Versionen der Daten ausgewählt und zur langfristigen Nachnutzung bereitgestellt werden. Die gute wissenschaftliche Praxis empfiehlt eine Archivierung von mindestens 10 Jahren.

- Auswahl von Daten (begrenzte Speicherkapazitäten)
  - Sind die Daten notwendig um Forschungsergebnisse zu verifizieren?
  - Sind die Daten auch langfristig von Interesse?
  - Wäre eine erneute Datenerhebung unverhältnismäßig teuer?
- Bereits zu Beginn der Projektarbeit sollten offene und langzeitstabile Dateiformate zur Abspeicherung verwendet werden.
  - Texte: PDF/A, ODT oder TXT
  - Tabellen: CSV, ODS oder XLSX
  - Bilder/Grafiken: TIFF, PNG oder JPEG2000

Unerlässlich für das dauerhafte Wiederauffinden und die Zitierbarkeit von Daten ist ein persistenter Identifikator, der einem Datensatz zugeordnet wird und stets eindeutig und direkt auf diesen verweist.

## DOI (Digital Object Identifier)

- für Publikationen und Forschungsdaten
- einzigartige Folge von alphanumerischen Zeichen:
  - Präfix (Organisationskennung)
  - Suffix (Objektkennung)
- Beispiel:
  - <http://dx.doi.org/10.25532/OPARA-28>

## ORCID (Open Researcher and Contributor ID)

- eindeutige Zuordnung wissenschaftlicher Leistungen zu einer Person
- vom Forschenden selbst angelegt und gepflegt
- [Informationen der UB Chemnitz zu ORCID](#)
- Beispiel:
  - <https://orcid.org/0000-0001-9503-8367>

## Warum Daten veröffentlichen?

- Vorgabe von Forschungsförderern erfüllen
- eigene Arbeit sichern / Datenverluste minimieren
- Ergebnisse überprüfbar und nachvollziehbar machen
- Grundlage für neue Forschungsvorhaben und Kooperationen bereitstellen
- effizientere Nutzung der Daten ermöglichen
- eigene Sichtbarkeit erhöhen

Repositorien sind eine spezielle Form von Archiven zur Speicherung von digitalen Objekten.

## Sie unterscheiden sich:

- nach der Art der zu speichernden Objekte (Publikationen oder Forschungsdaten)
- nach der Domäne der enthaltenen Daten
  - Institutionell, z.B. RADAR oder OpARA
  - Fachlich, z.B. GESIS oder DRYAD
  - Generisch/Fächerübergreifend, z.B. Zenodo oder Figshare
  - Medienspezifisch, z.B. TIB AV-Portal oder GitHub
- nach der Speicherfrist der Daten
- nach den Nachnutzungsmöglichkeiten der Daten

**Wie findet man ein passendes  
Repositorium?**  
Nutzen Sie das Register  
[Re3Data](#)!



## Fachzeitschrift

- Fokus liegt auf dem Artikel
- Interpretation der Daten
- zugrundeliegende Daten eventuell nur auf Anfrage

vs.

## Datenjournal

- Fokus liegt auf der Veröffentlichung von Forschungsdaten
- Beschreibung der Daten
- zum Teil journaleigene Repositorien

## Beispiele von Datenjournals

- [Scientific Data](#)
- [Journal of Open Psychology Data \(JOPD\)](#)
- [The Journal of Chemical & Engineering Data](#)

## Achtung !

Nachfolgend möchten wir Sie auf einige Informationen zu rechtlichen und ethischen Aspekten im Zusammenhang mit FDM hinweisen. Diese Hinweise stellen jedoch keine rechtsverbindlichen Aussagen dar.

Die Universitätsbibliothek Chemnitz möchte Sie mit diesen Informationen beim Management Ihrer Daten unterstützen. Es handelt sich aber lediglich um unverbindliche Informationen und Hilfestellungen zum Thema.

**Wir sind nicht befugt rechtsverbindliche Aussagen bzw. Beratungen anzubieten. Alle angebotenen Informationen sind ohne Gewähr auf Richtigkeit und Vollständigkeit!**

Routinemäßig erhobene Daten und Fakten (z.B. Messdaten, Metadaten) unterliegen an sich nicht dem deutschen Urheberrecht.

**ACHTUNG!**

Es können andere Schutzrechte bestehen (Leistungsschutzrechte des UrhG, Datenschutz, Arbeitsrecht, Kooperationsverträge)!

”In Zweifelsfällen sollte daher von einer grundsätzlichen Schutzfähigkeit der Forschungsdaten ausgegangen werden.” [Quelle 1, Seite 91]

Urheberrechtlich geschützte Forschungsdaten dürfen nur mit Zustimmung des Rechteinhabers (z.B. durch eine Lizenz - [CC-BY](#) , [ODC](#) , [DL-DE](#)) oder aufgrund gesetzlicher Vorgaben (z.B. [§60a](#) / [§60c](#) UrhG in Lehre / wissenschaftlicher Forschung) verwertet werden.

Quelle 1: M. Putnings, H. Neuroth und J. Neumann: [Praxishandbuch Forschungsdatenmanagement](#).

Beim Datenschutz geht es um den Schutz von personenbezogenen Daten.

Daten sind personenbezogen, wenn sie die Identifizierbarkeit der Person hinter den Daten zulassen.

Die Identifizierbarkeit kann durch

- ... **Anonymisierung** (kein Rückschluss auf die Person dahinter) verhindert werden. Dann findet das Datenschutzrecht keine Anwendung. Aber solch ein Grad an Anonymisierung ist in der Praxis selten.
- ... **Pseudonymisierung** (Verwendung einer Codierung) verschleiert werden. Da durch Entschlüsselung ein Rückschluss auf die Person möglich ist, gilt hierbei das Datenschutzrecht.

Quelle 1: M. Putnings, H. Neuroth und J. Neumann: [Praxishandbuch Forschungsdatenmanagement](#).

Quelle 2: [Datenschutz-Grundverordnung \(DSGVO\)](#)

## Einwilligung:

„Personenbezogene Daten dürfen nur dann erhoben oder verarbeitet werden, wenn eine Einwilligung der betroffenen Personen vorliegt oder ein anderer gesetzlicher Erlaubnistatbestand die Datenverarbeitung gestattet (Art. 6 Abs. 1 DSGVO).“ [Quelle 1, Seite 103]

Bei Fragen zum Datenschutz können Sie sich an den [Datenschutzbeauftragten der TU Chemnitz](#) wenden.

Quelle 1: M. Putnings, H. Neuroth und J. Neumann: [Praxishandbuch Forschungsdatenmanagement](#).

Quelle 2: [Datenschutz-Grundverordnung \(DSGVO\)](#)

Das BMBF geförderte Projekt DataJus setze sich in seiner Projektlaufzeit von 2017 bis 2019 mit den rechtlichen Rahmenbedingungen des FDM auseinander. Das Ergebnis sind unter anderem zwei Entscheidungsbäume.

## Entscheidungsbäume

- Urheberrechtliche Fragestellungen für die Veröffentlichung von Forschungsdaten
- Datenschutzrechtliche Fragestellungen für die Veröffentlichung von Forschungsdaten

Weitere Informationen finden Sie auf  
der [Projektwebseite von DataJus](https://www.tu-chemnitz.de/ub/publizieren/forschungsdaten).

Eine Lizenz ist eine vertragliche Vereinbarung, die festlegt wer die Daten wie nutzen darf. Das offene Format der **Creative-Commons-Lizenzen** wird ab Version 4.0 für die Lizenzierung von Forschungsdaten empfohlen.

Am bekanntesten sind CC-Lizenzen für die Lizenzierung von freien Texten und Materialien. Hierzu noch einige Informationen.

Informationen zu CC-Lizenzen auf den [Webseiten der UB Chemnitz](#).

Informationen zur [Auswahl einer passenden Lizenz](#).

Informationen zur [TULLU-Regel](#) für die korrekte Verwendung von offen lizenzierten Werken.

## Verantwortung gegenüber

- ... der Wissenschaft
  - kein Wissenschaftsbetrug: Fabrikation, Falsifikation und Plagiat [Quelle 1, Seite 120]
  - [Europäischer Verhaltenskodex für Integrität in der Forschung](#)
- ... dem Forschungsgegenstand
  - Menschen (informierte Einwilligung), Tiere, Institutionen, Objekte
- ... der Gesellschaft

Bei Fragen zur Forschungsethik können Sie sich an die [Ethikkommission der TU Chemnitz](#) wenden.



Die Nationale Forschungsdateninfrastruktur (NFDI) ist ein bundesweit verteiltes Netzwerk zum strategischen Ausbau des FDM in Deutschland. Es gibt unterschiedliche fachspezifische Konsortien, deren Aktivitäten vom [NFDI-Verein](#) koordiniert werden.

Es gibt [27 Konsortien](#), zum Beispiel:

- [MaRDI](#) (Mathematische Forschungsdateninitiative)
- [KonsortSWD](#) (Konsortium für die Sozial-, Bildungs-, Verhaltens- und Wirtschaftswissenschaften)
- [NFDI4Chem](#) (Fachkonsortium Chemie für die NFDI)
- [NFDI4Ing](#) (NFDI für die Ingenieurwissenschaften)
- [NFDI-MatWerk](#) (NFDI für Materialwissenschaft und Werkstofftechnik)
- [NFDI4DataScience](#) (NFDI für Datenwissenschaften und Künstliche Intelligenz)

Fast jedes Bundesland hat seine eigene FDM-Initiative bzw. FMD-Projekte. In Sachsen gibt es die Initiative SaxFDM. Hierbei handelt es sich um einen Zusammenschluss von mehreren Hochschulen und Forschungseinrichtungen.

Die Mitglieder von [SaxFDM](#) stehen im regelmäßigen Austausch miteinander und organisieren unter anderem:

- eine jährliche FDM-Tagung
- sowie „Digital Kitchen“.

- M. Jackenkroll: Wie organisiere ich meine Dateien und Ordner? Einführung in das Forschungsdatenmanagement. Chemnitz 2021. URL <https://www.tu-chemnitz.de/ub/kontakt-und-wir/texte/presse.html#weitere>
- B. Weiner, S. Wünsche, S. Kühne, P. Voigt, S. Frericks, C. Hoffmann, R. Elze, & R. Gey: Crashkurs Forschungsdatenmanagement. 30. April 2020. Zenodo. URL <https://doi.org/10.5281/zenodo.3778431>
- Universitätsbibliothek Chemnitz: Forschungsdaten. Zugriff 12.12.2022. URL <https://www.tu-chemnitz.de/ub/publizieren/forschungsdaten>
- Universitätsbibliothek Chemnitz: Informationen zu Creative-Commons-Lizenzen. Zugriff 20.12.2022. URL <https://www.tu-chemnitz.de/ub/publizieren/creativecommons.html>
- Forschungsdaten.info . Zugriff 12.12.2022. URL <https://forschungsdaten.info>
- DFG: Checkliste für einen angemessenen Umgang mit Forschungsdaten in DFG-Projekten. Version 21.12.2021. URL [https://www.dfg.de/foerderung/grundlagen\\_rahmenbedingungen/forschungsdaten/index.html](https://www.dfg.de/foerderung/grundlagen_rahmenbedingungen/forschungsdaten/index.html)
- Open Knowledge Foundation Deutschland. Zugriff 14.12.2022. URL <https://okfn.de/>
- B. Immenhauser: Open und FAIR Data: Der Beitrag der SAGW. 2. November 2018. Zugriff 15.12.2022. URL [https://sagw.ch/fileadmin/redaktion\\_sagw/dokumente/Veranstaltungen/2018/Open\\_Data/Praesentationen/3\\_Beat\\_Immenhauser\\_1\\_.pdf](https://sagw.ch/fileadmin/redaktion_sagw/dokumente/Veranstaltungen/2018/Open_Data/Praesentationen/3_Beat_Immenhauser_1_.pdf)
- M. Putnings, H. Neuroth und J. Neumann: Praxishandbuch Forschungsdatenmanagement. Berlin 2021. URL <https://doi.org/10.1515/9783110657807>
- Datenschutz-Grundverordnung (DSGVO) der EU. Einsehbar unter <https://dsgvo-gesetz.de>

- P. Brettschneider: Daten lizenzieren. 29. Mai 2020. Zenodo. URL <https://doi.org/10.5281/zenodo.3865204>
- S. Borski und J. Muuß-Merholz: OER leicht gemacht mit der TULLU-Regel. 5. Dezember 2016. Zugriff 20.12.2022. URL <https://open-educational-resources.de/oer-tullu-regel/>
- ALLEA (All European Academies): Europäischer Verhaltenskodex für Integrität in der Forschung. Überarbeitete Fassung. 2018. URL [https://allea.org/wp-content/uploads/2018/06/ALLEA-European-Code-of-Conduct-for-Research-Integrity-2017-Digital\\_DE\\_FINAL.pdf](https://allea.org/wp-content/uploads/2018/06/ALLEA-European-Code-of-Conduct-for-Research-Integrity-2017-Digital_DE_FINAL.pdf)
- Nationale Forschungsdateninfrastruktur (NFDI) e.V. Zugriff 20.12.2022. URL <https://www.nfdi.de/>
- SaxFDM (Forschungsdatenmanagement in Sachsen). Zugriff 21.12.2022. URL <https://saxfdm.de/>



Dieses Werk - ausgenommen das Logo der TU Chemnitz - ist lizenziert unter einer [Creative Commons Namensnennung 4.0 International Lizenz](https://creativecommons.org/licenses/by/4.0/). Bitte beachten Sie die Rechte an den zitierten Materialien.