

FORSCHUNGSDATENMANAGEMENT – WIE OFFEN KANN ES SEIN?

Gerald Jagusch, Universitäts- und Landesbibliothek Darmstadt

Open Up!-Webinar für die Hochschule der Medien Stuttgart | 5. Dezember 2023

AGENDA

- 1** (sehr!) kurze Einführung in das FDM
- 2** Unterschiede zwischen FAIR und Open Data
- 3** FDM und das Open-Science-Paradigma
- 4** Erfahrungsbeispiele aus der TU Darmstadt
- 5** ... und dem Konsortium NFDI4Ing
- 6** Fazit & Ausblick

(SEHR!) KURZE EINFÜHRUNG IN DAS FDM

▪ Forschungsdaten

▪ z.B. [laut DFG](#):

- „Zu Forschungsdaten zählen u. a. **Messdaten, Laborwerte, audiovisuelle Informationen, Texte, Surveydaten** oder **Beobachtungsdaten, methodische Testverfahren** sowie **Fragebögen**.
- **Korpora** und **Simulationen** können ebenfalls zentrale Ergebnisse wissenschaftlicher Forschung darstellen und werden daher ebenfalls unter den Begriff Forschungsdaten gefasst.
- Da Forschungsdaten in einigen Fachbereichen auf der Analyse von **Objekten** basieren (z. B. Gewebe-, Material-, Gesteins-, Wasser- und Bodenproben, Prüfkörper, Installationen, Artefakte und Kunstgegenstände), muss der Umgang mit diesen ebenso sorgfältig sein und eine fachlich adäquate Nachnutzungsmöglichkeit, wann immer sinnvoll und möglich, mitgedacht werden.
- Ähnliches gilt, wenn **Software** für die Entstehung oder Verarbeitung von Forschungsdaten erforderlich ist.“

▪ Datenmanagement

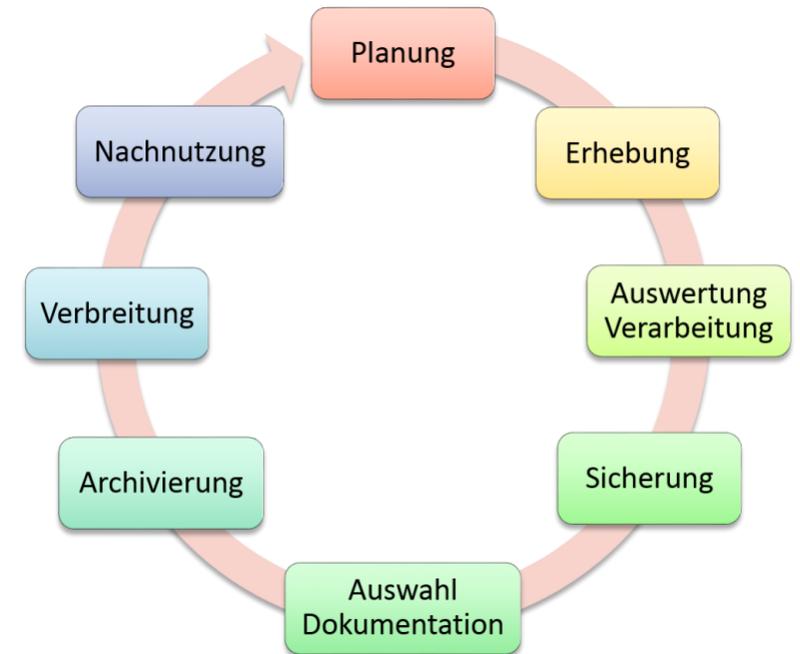
- „... ist die Menge aller methodischen, konzeptionellen, organisatorischen und technischen Maßnahmen und Verfahren zur Behandlung der Ressource „Daten“ mit dem Ziel, sie mit ihrem maximalen Nutzungspotenzial in die Geschäftsprozesse einzubringen und im laufenden Betrieb deren optimale Nutzung zu gewährleisten.
- Darüber hinaus muss ein professionelles Datenmanagement auch die Aspekte der Daten-/Informationsqualität und des Datenschutzes berücksichtigen.“ ([Wikipedia](#))

→ Forschungsdatenmanagement

(SEHR!) KURZE EINFÜHRUNG IN DAS FDM

Forschungsdatenmanagement behandelt die Themen

- Datendokumentation (Metadaten) (*Wo ist hier die Grenze?*)
- Datenorganisation (Dateiformate, Datenaustausch)
- Datenspeicherung und -archivierung
- Rechtliche Aspekte (Urheberrecht, Datenschutz, Lizenzierung)
- Datenpublikation (Repositorien, Datenzentren, Persistent Identifiers (PIDs))



(SEHR!) KURZE EINFÜHRUNG IN DAS FDM

- Forschungsdatenmanagement gibt es seit es Forschung gibt.
- Den Begriff FDM gibt es etwa seit 20 Jahren.
- In den letzten 10 Jahren hat sich das FDM in der Breite außerordentlich qualitativ und quantitativ weiter entwickelt und professionalisiert.
- Bewusstsein für das „Problem“ FDM ist inzwischen überall vorhanden.
- Lösungen sind idR individuell oder sehr fachspezifisch.
- Es gibt (noch) nahezu keine Anreize im Wissenschaftssystem gutes FDM zu machen.

(SEHR!) KURZE EINFÜHRUNG IN DAS FDM

In Deutschland vielfache Ansätze zum FDM (→ Bildungsföderalismus! → Wissenschaftsfreiheit → digitales als „Neuland“):

- Lokale FDM-Servicestellen der Hochschulen und Forschungseinrichtungen, meist als joint venture von Bibliothek, Rechenzentrum und Forschungsverwaltung
- ab ca. 2013: „FDM-Landesinitiativen“ der Bundesländer (ganz unterschiedlich aufgestellt Projekte und Strukturen)
- Seit 2020: Nationale Forschungsdateninfrastruktur (NFDI), gefördert von der GWK, v.a. vom Bund (27 fachspezifische Konsortien, zahlreiche Sektionen, Arbeitsgruppen, Gremien)

auf europäischer Ebene: European Open Science Cloud (EOSC) → EU-weite Infrastruktur

Weltweit (seit 2013): Research Data Alliance (RDA) → Austausch und Empfehlungen, Standards

Und natürlich auch kommerzielle Anbieter, z.B. Elsevier

UNTERSCHIEDE ZWISCHEN FAIR UND OPEN DATA

FAIRe Daten?

2016 wurden die [FAIR Guiding Principles for scientific data management and stewardship](#) in der Zeitschrift *Scientific Data* veröffentlicht.

- **Findability** = Auffindbarkeit (PIDs, öfftl. Indexierung der Metadaten)
- **Accessibility** = Zugänglichkeit (technische Standards)
- **Interoperability** = Kompatibilität (reiche & standardisierte Metadaten, Linked data)
- **Reusability** = Wiederverwendbarkeit (rechtlich: Lizenzierung, Provenienzangaben)

Diese Grundsätze legen Schwerpunkt auf maschinelle Verarbeitbarkeit (d. h. die Fähigkeit von Computersystemen, Daten zu finden, darauf zuzugreifen, zu interagieren und sie wiederzuverwenden, ohne dass der Mensch eingreifen muss).

Haben sich seitdem rasant durchgesetzt als „buzz word“ und als Goldstandard für gutes FDM.

Hinsichtlich „FAIRness“ können Datensätze automatisiert geprüft werden z.B. mit [F-UJI](#).

UNTERSCHIEDE ZWISCHEN FAIR UND OPEN DATA

Offene Daten?

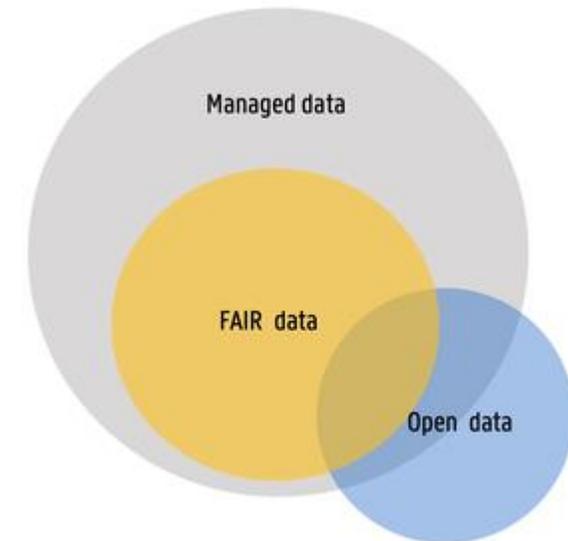
„Als offene Daten werden Daten verstanden, die ohne jedwede Einschränkung zur freien Nutzung, Weiterverbreitung und freien Weiterverwendung zugänglich gemacht worden sind. Es handelt sich bei ihnen unter anderem um Lehrmaterial, Geodaten, Statistiken, Verkehrsinformationen, **wissenschaftliche Publikationen, medizinische Forschungsergebnisse** oder Hörfunk- und Fernsehsendungen.“

- Open data ist viel breiter als „Forschungsdaten“

UNTERSCHIEDE ZWISCHEN FAIR UND OPEN DATA

FAIR ≠ Open

- FAIR schließt Open nicht ein, denn ggf. müssen nur die (high level) Metadaten öfftl. zugänglich sein, d.h. nur die Existenz und ggf. rudimentäre Informationen sind offen zugänglich
- Open schließt FAIR nicht ein, denn eine (offene) Lizenzierung erfüllt nur 1 der 15 FAIR-Kriterien.
- „Perfekte“ Forschungsdaten sind „FAIR open data“
- FAIR ist weit wichtiger als open

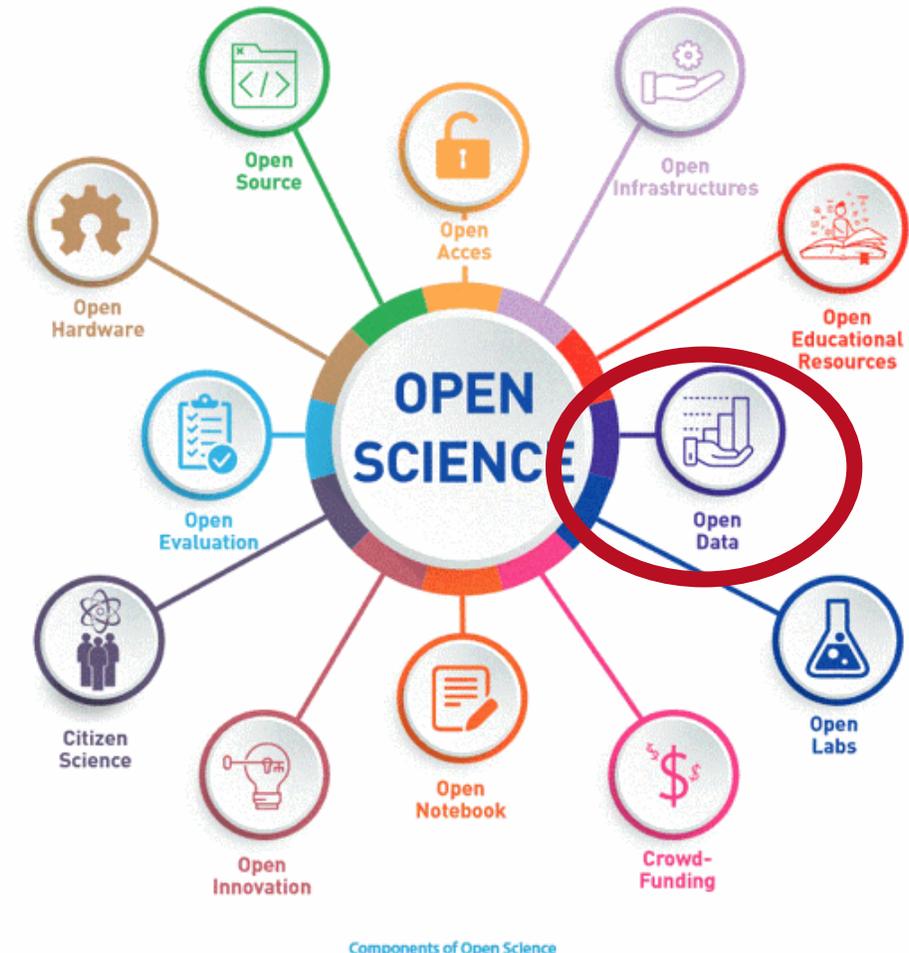


<https://www.ugent.be/en/research/datamanagement/after-research/fair-data.htm>

FDM UND DAS OPEN-SCIENCE-PARADIGMA

Open Science in der UNESCO-Definition umfasst FDM an vielen Stellen:

- Open data
- aber auch
 - Open Source (→ Forschungssoftware)
 - Open Infrastructures
 - Open Labs
 - Open Notebook (→ Metadaten)
 - Open Innovation (→ Datentransfer)
 - Open Evaluation (→ FDM-Anreize)
 - Open Hardware (→ Lesbarkeit von Daten)



Components of Open Science

ERFAHRUNGSBEISPIELE AUS DER TU DARMSTADT

- Lokale Servicestelle „[TUdata](#)“ der TU Darmstadt
 - Teil der FDM-Landesprojekts [Hessische Forschungsdateninfrastrukturen](#) (HeFDI)
 - Ingenieurwissenschaften als Hauptkunden der Services, z.B. Repository [TUdatalib](#)
 - große Heterogenität im FDM an der TU
 - vom open data Pionier in der Biologie oder Informatik
 - bis zu Fachgebieten ohne Awareness
 - (noch) keine Open-Science-Strategie, Anreize und Verpflichtungen fehlen
 - Game changer sind gute Services, die Lücken schließen oder Forscherleben erleichtern, z.B.
 - Antragsberatung
 - Elektronische Laborbücher
 - Jupyter Notebooks
 - Datenspeicherung
 - ...
 - Zukunft: Semantische Vernetzung aller Informationen (?)

TU data

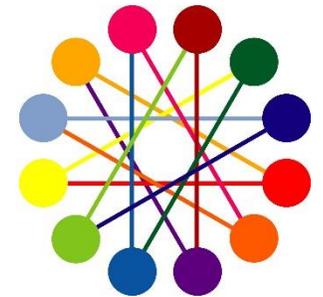


ERFAHRUNGSBEISPIELE AUS DEM KONSORTIUM NFDI4ING

Nationale Forschungsdateninfrastruktur für die Ingenieurwissenschaften (NFDI4Ing)

- Seit 2020
- Ingenieurwissenschaften haben bislang
 - sehr wenig gemeinsame Infrastruktur im FDM
 - wenig FDM-Standards
 - kaum Strukturbildung
 - auch international kaum gute FDM-Vorbilder
 - FAIR data innerhalb der eigenen Instituts als Herausforderung
- aktuelle Themen:
 - [Datentreuhandmodelle](#) & Verbindung zum Datenmanagement in der Industrie (z.B. [Gaia-X](#))
 - Gesetzgebung zieht an (Lieferkettengesetz, EU data act, usw.)

NFDI4ing



FAZIT & AUSBLICK

- FDM ist ein Dauerthema, kein Modetrend
- FAIR ist eine etablierte Richtschnur, open data eine Spielart davon
- Das Feld professionalisiert sich zusehends aus verschiedenen Richtungen, z.B.
 - Data Stewards
 - Data Librarians
 - Data Scientists
- Die FDM-Reifegrade sind hochgradig unterschiedlich in den Fächern, Institutionen, Personen
- Abgrenzungen von FDM verschwimmen (Data science, Künstliche Intelligenz)

WEITERE LINKS

- <https://forschungsdaten.info> (DIE deutschsprachige Seite zum FDM, redaktionell betreut)
- Deutsche Forschungsgemeinschaft (DFG) zum [Umgang mit Forschungsdaten](#)
- [Rat für Informationsinfrastrukturen \(RfII\)](#)