

Forschungsdatenmanagement

Notwendige, aber nicht hinreichende Voraussetzung
für den wissenschaftlichen Erkenntnisgewinn

03. Was ist Forschungsdatenmanagement?

Dr. habil. Till Biskup

Physikalische Chemie

Universität Rostock

26.04.2024





- ❏ Forschungsdatenmanagement ist „alter Wein in neuen Schläuchen“. Wir haben nur verlernt, mit unseren Daten umzugehen.
- ❏ Forschungsdatenmanagement ist eine Frage der Einstellung, weniger eine Frage der verfügbaren Werkzeuge.
- ❏ Forschungsdatenmanagement ist notwendige, aber keine hinreichende Voraussetzung wissenschaftlichen Erkenntnisgewinns.
- ❏ Forschungsdaten umfassen neben den Daten und Metadaten auch die Werkzeuge und sind grundsätzlich publikationswürdig.
- ❏ Die primäre Verantwortung für Forschungsdatenmanagement liegt bei den individuellen Forschenden.

These

Forschungsdatenmanagement ist nichts Neues,
sondern ein Kernaspekt der wissenschaftlichen Methode.

- ▶ Nachvollziehbarkeit ist konstituierend für die Wissenschaftlichkeit. Das erfordert einen systematischen und strukturierten Umgang mit Forschungsdaten als dem Ausgangspunkt des Erkenntnisgewinns.
- ▶ Vielen Forschenden fehlt das Bewusstsein für die Bedeutung von Forschungsdatenmanagement. Sie betreiben entsprechend oft selbst kein (ausreichendes) Forschungsdatenmanagement.
- ▶ Die Konsequenz: geringe Qualität der meisten Veröffentlichungen und mangelnde Nachvollziehbar- und Reproduzierbarkeit.

These

Forschungsdatenmanagement hat nichts mit Digitalität zu tun,
Digitalität macht Forschungsdatenmanagement nur drängender.

- ▶ Die eigentliche Neuerung der zunehmenden Digitalität: Das Erzeugen nahezu beliebiger Datenmengen wird stark vereinfacht.
- ▶ Quantität ohne Qualitätssicherung konterkariert die Wissenschaft.
- ▶ Digitalität und datengetriebene Wissenschaft als Treiber des Forschungsdatenmanagements zu verstehen, greift (viel) zu kurz.
- ▶ Forschungsdaten liegen nicht notwendigerweise digital vor. Die unnötige Einschränkung auf digital verfügbare Daten führt zu einer Verarmung und Einschränkung wissenschaftlicher Erkenntnis.

Forschungsdatenmanagement: eine Annäherung in Definitionen

Forschungsdaten: ein weites Feld

Verantwortlichkeiten für Forschungsdatenmanagement

Umsetzung von Forschungsdatenmanagement

Forschungs-daten-management

- ▶ drei Begriffe, die zu definieren wären:
 - Forschung
 - (Forschungs-)Daten
 - Management
- ▶ Verständigung über die Begriffe beugt Misskonzeptionen vor.
 - „Forschungsdatenmanagement“ meist nicht bzw. missverstanden

Historische (und irreführende) Einordnung des Begriffs

- ▶ meist mit zunehmender Digitalisierung verknüpft/daraus motiviert
 - „Big Science“, „Big Data“, „e-Science“, „Viertes Paradigma“
- ▶ Hauptmotivation: Wiederverwendung von Daten durch Dritte
 - oft schwierig/unmöglich: viele unterschiedliche und gute Gründe
 - einseitige hochproblematische „Lösung“: FAIR-Prinzipien

Forschung

Systematisches Vorgehen, um einer bestimmten Fragestellung nachzugehen oder Phänomene zu erklären oder Experimente unter kontrollierten Bedingungen durchzuführen, das der wissenschaftlichen Methodik folgt. Wissenschaft setzt Forschung voraus. Allerdings kann Forschung ohne Beitrag zur Wissenschaft (ohne Erkenntnisgewinn) bleiben, vgl. Feynmans Begriff „Cargo Cult Science“.

- ▶ Forschung und Wissenschaft sind voneinander zu unterscheiden.
 - Wissenschaft setzt Forschung voraus.
 - Forschung kann ohne Beitrag zur Wissenschaft bleiben.
- ▶ Forschung sollte der wissenschaftlichen Methodik folgen.
 - wurde unter der Überschrift „Wesen der Wissenschaft“ diskutiert

R. P. Feynman: Cargo Cult Science. *Engineering and Science* 37(7):10–13, 1974

Forschungsdaten (Arbeitsdefinition)

Daten, die im Zuge wissenschaftlicher Vorhaben im Rahmen von Forschung z.B. durch Digitalisierung, Quellenforschungen, Experimente, Messungen, Erhebungen oder Befragungen entstehen.

- ▶ Forschungsdaten sind nicht notwendigerweise digital.
- ▶ Daten werden erst durch eine Forschungsfrage zu Forschungsdaten.
- ▶ Forschungsdaten als solche zu erkennen/zu identifizieren, ist selbst Teil des wissenschaftlichen Vorgehens.
- ☛ Eine Antwort auf die Frage „Was sind Forschungsdaten?“ ist hochgradig nichttrivial und kontextabhängig.
- ☛ Details im folgenden Abschnitt

Management (Tätigkeit)

Planung, Organisation und Verwaltung (und ggf. Leitung) unter ökonomischen Gesichtspunkten. Auch wenn im Kontext der Wissenschaft ökonomische Aspekte oft nur eine untergeordnete Rolle spielen, ist ein ökonomischer Ressourceneinsatz kein Widerspruch zur (und keine Einschränkung der) wissenschaftlichen Vorgehensweise.

- ▶ Strukturen und Formalisierung helfen dabei, Daten und Wissen aufzubereiten und daraus Erkenntnisse zu gewinnen.
 - Organisation ist Mittel zum Zweck, kein Selbstzweck.
- ▶ Erkenntnisgewinn ist immer ein kreativer Prozess, keine Routine.
 - Erkenntnis setzt ein erkennendes Subjekt voraus.
 - Zu viel Struktur und Formalisierung kann Erkenntnis verhindern.

Forschungsdatenmanagement

Umgang mit Forschungsdaten über ihren gesamten Lebenszyklus hinweg mit dem Fokus auf Nachvollziehbarkeit und Nachnutzbarkeit; wird meist auf die digitale Welt bezogen, ist letztlich aber nichts anderes als sauberes wissenschaftliches Arbeiten; notwendige, aber nicht hinreichende Bedingung für den wissenschaftlichen Erkenntnisgewinn.

- ▶ Forschungsdatenmanagement ist eine Frage der Einstellung, weniger eine Frage der verfügbaren Werkzeuge.
- ▶ Ein angemessener Umgang mit Forschungsdaten ist eine Frage der Professionalität als Wissenschaftler.
- ▶  Forschungsdatenlebenszyklus wird noch ausführlich thematisiert.

Häufige Misskonzeptionen und Missverständnisse

- ▶ FDM betrifft nur digitale Daten.
 - FDM ist unabhängig von Medium und Art der Daten.
- ▶ FDM ist ein von oben verordnetes „notwendiges Übel“.
 - FDM ist Mittel zum Zweck (Forschung), nicht Selbstzweck.
- ▶ FDM ist Garant für qualitativ hochwertige Forschung.
 - Feynman: „Cargo Cult Science“
- ▶ FDM schränkt die Forschungsfreiheit ein.
 - Forschende tragen Verantwortung für den Umgang mit ihren Daten.
- ▶ FDM lässt sich vom ursprünglichen Forschungskontext trennen.
 - Korrekter Umgang mit Forschungsdaten erfordert Fachwissen.
- 👉 FDM wird oft falsch verstanden und betrieben, ist kontraproduktiv und schadet in dieser Form der Wissenschaft mehr, als es nützt.

These

Forschungsdatenmanagement ist „alter Wein in neuen Schläuchen“ – wir haben über die Digitalisierung (und den massiven Zuwachs an Forschenden und Daten) nur weitgehend verlernt, ordentlich und der Wissenschaft angemessen mit unseren Daten umzugehen.

These

Forschungsdatenmanagement ist nichts anderes als sauberes wissenschaftliches Arbeiten nach bestem Wissen und Gewissen.

Wer (inhaltlich) kein Forschungsdatenmanagement betreibt, betreibt auch keine Wissenschaft.

Forschungsdatenmanagement garantiert aber nicht Wissenschaftlichkeit.

Forschungsdatenmanagement: eine Annäherung in Definitionen

Forschungsdaten: ein weites Feld

Verantwortlichkeiten für Forschungsdatenmanagement

Umsetzung von Forschungsdatenmanagement

“ *Forschungsdaten sind eine wesentliche Grundlage für das wissenschaftliche Arbeiten. Die Vielfalt solcher Daten entspricht der Vielfalt unterschiedlicher wissenschaftlicher Disziplinen, Erkenntnisinteressen und Forschungsverfahren. Zu Forschungsdaten zählen u.a. Messdaten, Laborwerte, audiovisuelle Informationen, Texte, Surveydaten, Objekte aus Sammlungen oder Proben, die in der wissenschaftlichen Arbeit entstehen, entwickelt oder ausgewertet werden. Methodische Testverfahren, wie Fragebögen, Software und Simulationen können ebenfalls zentrale Ergebnisse wissenschaftlicher Forschung darstellen und sollten daher ebenfalls unter den Begriff Forschungsdaten gefasst werden.*

– DFG

Drei Kategorien von Forschungsdaten

- ▶ Daten im herkömmlichen Sinn
 - „u.a. Messdaten, Laborwerte, audiovisuelle Informationen, Texte, Surveydaten“
 - können analog oder digital vorliegen
- ▶ Objekte
 - „Objekte aus Sammlungen oder Proben“
 - physische Objekte
- ▶ Werkzeuge
 - „Fragebögen, Software und Simulationen“
- ☛ Forschungsdaten sind Ausgangspunkt der Wissenschaft, nicht das fertige Ergebnis oder die daraus gezogene Erkenntnis.
- ☛ Inklusive Definition stellt uns vor große Herausforderungen im Umgang mit Forschungsdaten: Proben? Software?

“ *The recognition that an observation, artifact, or record constitutes data is itself a scholarly act. [...]*
The concept of data is difficult to define, as data may take many forms, both physical and digital.

– Christine L. Borgman

- ▶ Zuschreibung ist nicht universell
 - Des einen Rauschen sind des anderen Daten (Bsp.: Astronomie).
 - Forschungsdaten werden zu Forschungsdaten im Kontext einer wissenschaftlichen Fragestellung.
- ▶ Forschungsdaten als solche zu erkennen erfordert Kompetenz
 - Vertrautheit mit Fachdisziplin und Fragestellung
 - Denken jenseits überkommener Traditionen – Bsp.: Werkzeuge werden viel zu selten als Forschungsdaten erkannt und akzeptiert.

Forschungsdatenmanagement: eine Annäherung in Definitionen

Forschungsdaten: ein weites Feld

Verantwortlichkeiten für Forschungsdatenmanagement

Umsetzung von Forschungsdatenmanagement

Leitmotiv

Forschungsdatenmanagement ist primär die Verantwortung der individuellen Forschenden.

- ▶ Forschungsdatenmanagement erfordert durch reale langjährige Erfahrung erworbene Kompetenzen sowohl in der jeweiligen Fachdomäne als auch mit dem Umgang relevanter Werkzeuge.
- ▶ Der Ruf nach professionellen Forschungsdatenmanagern (Data Steward, Data Architect) und Datenkuratoren ist fehlgeleitet und nicht zielführend.
- ▶ Forschende müssen sich die notwendigen Kompetenzen selbst aneignen, wollen sie den Anforderungen der Wissenschaft genügen.

Individuelle Verantwortlichkeiten

- ▶ Gruppenleitung
- ▶ eigenverantwortlich Forschende
- ▶ abhängig Forschende

Institutionelle Verantwortlichkeiten

- ▶ Leitung der Institution
 - ▶ Koordination (Verwaltung)
 - ▶ Infrastruktur-Einrichtungen: IT, Bibliothek
-
- ☛ Jede Ebene hat je eigene Verantwortlichkeiten und Zuständigkeiten.
 - ☛ Individuelle Forschende haben die primäre Verantwortung für FDM.
 - ☛ Kommunikation untereinander und über Grenzen ist essentiell.

- ▶ **Gruppenleitung**
 - Verantwortung, für die Umsetzung von FDM zu sorgen
 - zwei Seiten: FDM einfordern und ermöglichen
- ▶ **(Eigen-)verantwortlich Forschende**
 - Personenkreis: Doktoranden im 2./3. Jahr, PostDocs
 - Verantwortung, FDM umzusetzen
 - Verantwortung, Strukturen (mit) zu entwickeln und zu verbessern
- ▶ **abhängig Forschende**
 - Personenkreis: BSc bis Doktoranden im 1./2. Jahr
 - Verantwortung, FDM umzusetzen und sich an die vorgegebenen Regeln zu halten
 - Verantwortung, die vorgegebenen Regeln und Strukturen nach eigenen Möglichkeiten kritisch zu hinterfragen
- 👉 **Hinweis: Freiheit der Forschung (Art. 5 GG) greift erst sehr spät.**

“ *Kunst und Wissenschaft, Forschung und Lehre sind frei.
Die Freiheit der Lehre entbindet nicht von der
Treue zur Verfassung.*

– Art. 5, Abs. 3 GG

- ▶ Deutsche Besonderheit, erklärbar aus der Geschichte.
- ▶ Art. 5 GG gilt nur für *unabhängig* Forschende.

Wer gehört zu den unabhängig Forschenden?

- ▶ Unabhängig heißt vermutlich „nicht weisungsgebunden“.
- ▶ Spätestens mit abgeschlossener Habilitation, vermutlich aber auch schon als Leitung einer Nachwuchsgruppe.
- ☛ Im Zweifel das Justizariat (z.B. des DHV) fragen.

- ▶ Leitung der Institution
 - Erstellung und Verabschiedung von Leitlinien
 - Ermöglichung der Strukturen und Infrastrukturen, die es für ein funktionierendes FDM braucht
 - Motivation zum FDM durch authentisches Interesse daran
- ▶ Koordination
 - Hilfestellung bei der Umsetzung des FDM seitens der Forschenden
 - Bereitstellung von Informationen und Strukturen
 - aktives Werben für die Vorteile und Notwendigkeiten von FDM
 - Mittler zwischen Leitung, Infrastruktur und Forschenden
- ▶ Infrastruktur-Einrichtungen: IT, Bibliothek
 - Bereitstellung der notwendigen Infrastruktur
 - aktive Beratung und ggf. Entwicklung/Anpassung von Lösungen

Verantwortlichkeiten jenseits der Institutionen

- ▶ Fachverbände
 - Erarbeitung fachspezifischer Konventionen und Standards z.B. für Daten- und Metadatenformate
 - Etablierung einschlägiger Fach-Repositoryen etc.
- ▶ Forschungsförderer
 - Einfordern von FDM in den Anträgen, inkl. sinnvoller Begutachtung
 - Überprüfung, ob die FDM-Maßnahmen auch umgesetzt wurden
 - Förderung der Entwicklung und Etablierung von FDM-Werkzeugen jenseits großer Forschungsverbände oder der NFDI.
 - Langfristige Förderung und Bereitstellung von Infrastrukturen
- ▶ (über-)staatlich geförderte Institutionen
 - Bereitstellung globaler Infrastrukturen und Standards
 - Beispiele: EOSC, Zenodo, OAIS, DCAT-AP

Forschungsdatenmanagement: eine Annäherung in Definitionen

Forschungsdaten: ein weites Feld

Verantwortlichkeiten für Forschungsdatenmanagement

Umsetzung von Forschungsdatenmanagement

Leitfrage

- ❓ Was kann ich als individuelle(r) Forschende(r) im Hier und Jetzt tun, um zu einer Verbesserung der Wissenschaftlichkeit (meiner eigenen Forschung) beizutragen?

- ▶ Ziel: Sicherstellung/Verbesserung der Wissenschaftlichkeit
 - Es geht nicht um Häkchen und Metriken.
 - Extrinsische Anforderungen nutzen, um sich selbst die richtigen (wichtigen) Fragen zu stellen – und zu beantworten.

- ▶ persönliche Verantwortung
 - Es geht um *meine* eigene Forschung.
 - Es geht um das, was *ich* tun kann.

“ *Do what you can, with what you've got, where you are.*

– Theodore Roosevelt

- ▶ Bewusstsein für die Relevanz von FDM schärfen
 - Es geht um die Wissenschaft(lichkeit) – nicht weniger.
 - FDM ist Voraussetzung, aber kein Garant für Wissenschaftlichkeit.
- ▶ notwendiges Wissen aneignen
 - FDM ist nicht neu: Es gibt genug gute (alte) Literatur dazu.
 - Auf diejenigen hören, die langjährige praktische Erfahrung haben.
- ▶ notwendige Kompetenzen aneignen
 - Nur kompetent eingesetzte Werkzeuge werden zu Lösungen.
- ▶ FDM im persönlichen Kontext umsetzen
 - Schrittweise vorgehen: von einfach zu komplex, von bekannt und bewährt zu unbekannt und unerprobt.

Drei Leitlinien für eine realisierbare Umsetzung

- ▶ dezentral
 - automatisierte Synchronisierung mit zentralen Diensten möglich
 - ▶ modular – und damit flexibel und erweiterbar
 - Verknüpfungen statt ein Gesamtsystem
 - mit aktuellen Ressourcen umsetzbar
 - ▶ kompatibel/dialogfähig (*interoperable*)
 - beste Repräsentation für den jeweiligen Anwendungsfall
 - Schnittstellen mitdenken
 - offene Standards
 - maschinenlesbar
- 👉 Wird im dritten Teil der Vorlesung ausführlich behandelt:
„Bausteine eines funktionierenden, individuellen FDM“



- ❏ Forschungsdatenmanagement ist „alter Wein in neuen Schläuchen“. Wir haben nur verlernt, mit unseren Daten umzugehen.
- ❏ Forschungsdatenmanagement ist eine Frage der Einstellung, weniger eine Frage der verfügbaren Werkzeuge.
- ❏ Forschungsdatenmanagement ist notwendige, aber keine hinreichende Voraussetzung wissenschaftlichen Erkenntnisgewinns.
- ❏ Forschungsdaten umfassen neben den Daten und Metadaten auch die Werkzeuge und sind grundsätzlich publikationswürdig.
- ❏ Die primäre Verantwortung für Forschungsdatenmanagement liegt bei den individuellen Forschenden.