

Forschungsdatenmanagement

Notwendige, aber nicht hinreichende Voraussetzung
für den wissenschaftlichen Erkenntnisgewinn

15. Hindernisse für Forschungsdatenmanagement und mögliche Abhilfe

Dr. habil. Till Biskup

Physikalische Chemie

Universität Rostock

21.06.2024





- 🔑 Forschungsdatenmanagement setzt ein Problembewusstsein für seine Relevanz für die Wissenschaftlichkeit voraus.
- 🔑 Das aktuelle Wissenschaftssystem setzt falsche Anreize mit seinem Fokus auf Quantität statt Qualität.
- 🔑 Reflexion über Forschungsdatenmanagement offenbart mitunter Defizite in der Wissenschaftlichkeit der eigenen Forschung.
- 🔑 Fehlende bzw. fehlerhafte Anforderungsanalyse führt zu unbrauchbaren Werkzeugen und Lösungen.
- 🔑 Umsetzungsprobleme entstehen meist durch mangelnde Kenntnis und Kompetenz beim Einsatz existierender Werkzeuge.

- 1 Ausgangspunkt und Motivation
 - Wissenschaft; Datenmenge und Digitalität
- 2 Aspekte eines forschungsnahen Forschungsdatenmanagements
 - Forschungsdatenlebenszyklus
- 3 Bausteine eines funktionierenden, individuellen Forschungsdatenmanagements
 - Eigenschaften; Prinzipien; notwendige Kompetenzen; Werkzeuge
- 4 **Hindernisse und Probleme**
 - Hindernisse und mögliche Abhilfe; Antimuster
- 5 Funktionierende Lösungen
 - Bewährte Verfahren (aus eigener Anschauung)

Mangelndes Problembewusstsein

Persönliche Freiheit und intellektuelle Unmündigkeit

Fehlende oder fehlerhafte Anforderungsanalyse

Umsetzungsprobleme

These

Wer inhaltlich kein Forschungsdatenmanagement betreibt, dessen Forschung bleibt ohne Beitrag zur Wissenschaft.

- ▶ Forschung kann ohne Beitrag zur Wissenschaft bleiben: Wissenschaft geht es primär um Erkenntnisgewinn.
- ▶ Forschungsdatenmanagement dient der Nachvollziehbarkeit und ist notwendige, aber nicht hinreichende Voraussetzung für den wissenschaftlichen Erkenntnisgewinn.
- ▶ Forschung ohne Forschungsdatenmanagement ist in den wenigsten Fällen brauchbare Forschung, aber sicherlich ohne Beitrag zur Wissenschaft.

Leitmotiv

Die Vermittlung der für das Forschungsdatenmanagement notwendigen Kompetenzen muss in den universitären Curricula verankert werden.

- ▶ Die wissenschaftliche Methodik wird viel zu selten explizit behandelt, ist aber notwendige (wenn auch nicht hineichende) Vorbedingung für jeglichen wissenschaftlichen Erkenntnisgewinn.
- ▶ Die meisten notwendigen Kompetenzen haben nichts mit Digitalität zu tun, sondern mit Struktur, Organisation und letztlich (persönlichem) Projekt-, Zeit- und Ressourcenmanagement.
- ▶ Wissenschaft erfordert verantwortungsvolles Handeln – gegenüber der Gesellschaft und der Wissenschaft selbst.

“ We have learned a lot from experience about how to handle some of the ways we fool ourselves. [...] But this long history of learning how not to fool ourselves—of having utter scientific integrity—is, I'm sorry to say, something that we haven't specifically included in any particular course that I know of. We just hope you've caught on by osmosis.

– Richard P. Feynman

- ▶ Seit 1974 hat sich (fast) nichts geändert . . .
- ▶ Bausteine wissenschaftlicher Integrität
 - Arbeiten nach bestem Wissen und Gewissen
 - permanentes Hinterfragen der eigenen Ergebnisse, Schlüsse und Erkenntnisse
 - Transparenz und Ehrlichkeit

“ *getting something to work—once—isn't that hard.*

Getting it right is another matter entirely. [. . .] Mostly, it takes a passion for the craft and the desire to be a professional.

– Robert C. Martin

- ▶ Ein Ergebnis ist der notwendige *Ausgangspunkt* – nicht das Ziel.
 - Dass etwas *nicht* funktioniert (hat), ist ein valides Ergebnis.
 - Voraussetzung: Sorgfalt bei der Arbeit, konsequentes Hinterfragen
- ▶ nächster Schritt: verstehen, warum/wie das Ergebnis zustandekam
 - setzt Nachvollziehbarkeit voraus
 - Wie weiß ich, ob ich nachvollziehbar gearbeitet habe?
Indem ich die Fragen, die ich jetzt stelle, beantworten kann.
- ▶ Kennzeichen der Professionalität
 - sich seiner Verantwortung bewusst sein und entsprechend handeln

Leitmotiv

Die Qualität eines Großteils veröffentlichter Forschungsergebnisse wird den Ansprüchen der Wissenschaft nicht gerecht.

- ▶ Ein Hauptproblem sind die falschen Anreize im aktuellen Wissenschaftssystem: Quantität statt Qualität („publish or perish“).
- ▶ Qualitätssicherung durch Begutachtung ist die beste aller schlechten Lösungen, funktioniert aber auch nur sehr eingeschränkt – weil die Gutachter meist selbst Teil des Systems sind.
- ▶ Nur ein Bewusstsein für die Ansprüche der Wissenschaft und frühzeitige Vermittlung notwendiger Kompetenzen zum wissenschaftlichen Vorgehen hat Aussicht auf Erfolg.

These

Die meisten veröffentlichten Forschungsergebnisse sind bestenfalls irrelevant, schlimmstenfalls irreführend und/oder falsch.

- ▶ Häufige Gründe für mangelnde Qualität
 - Mangelnde Recherche und fehlender Kontext
 - Mangelnde Nachvollziehbarkeit
 - Mangelnde Qualität der erhobenen Daten aufgrund fehlender Einarbeitung oder Kompetenz
- ▶ Probleme
 - Qualität von Publikationen ist meist nur mühsam bestimmbar.
 - Unmengen von Veröffentlichungen zu sortieren bindet Ressourcen.
 - Die Qualität der Ergebnisse beeinflusst die eigene Forschung.

- ▶ Mangelnde Recherche und fehlender Kontext
 - Literaturrecherche ist mühsam, aber unabdingbar (und nie fertig).
 - Schriftliche Kontextualisierung ist Kernaufgabe der Forschenden.
 - Recherche und Kontext bewahren vor unnötiger Replikation.
- ▶ Mangelnde Nachvollziehbarkeit
 - Material- und Methodenteile werden nicht wertgeschätzt.
 - Veröffentlichungen beschreiben fast nie hinreichend genau, was gemacht wurde.
- ▶ Mangelnde Qualität der erhobenen Daten aufgrund fehlender Einarbeitung oder Kompetenz.
 - Geräte/Verfahren werden meist als „black box“ verwendet.
 - Einarbeitung wird oft nicht wertgeschätzt oder ermöglicht.
 - rechtzeitige Qualitätskontrollen fehlen weitestgehend bzw. sind Verantwortung der zu unerfahrenen Forschenden.
 - Daten *unbekannter* Qualität sind unbrauchbar.

These

Forschungsdatenmanagement ist eine politische Worthölse, die den notwendigen intellektuellen Tiefgang vermissen lässt.

- ▶ Den meisten, die sich „professionell“ mit FDM befassen, fehlt die Erfahrung und damit die Kompetenz/Qualifikation, um ihren Aufgaben sinnvoll gerecht werden zu können.
- ▶ Die politisch geforderten Forschungsdateninfrastrukturen offenbaren eine schockierende Unkenntnis sowohl der Historie als auch der realen Gegebenheiten.
- ▶ Das Rad ständig neu zu erfinden und aus Fehlern anderer (aus Unkenntnis und/oder Ignoranz) nicht zu lernen, führt nicht zu brauchbaren Lösungen drängender Probleme.

“ *Scholarship and data have long and deeply intertwined histories. Neither are new concepts. What is new are efforts to extract data from scholarly processes and to exploit them for other purposes. Costs, benefits, risks, and rewards associated with the use of research data are being redistributed among competing stakeholders. The goal of this book is to provoke a much fuller, and more fully informed, discussion among those parties. At stake is the future of scholarship.*

– Christine L. Borgman

- ☛ Das Herausreißen von Forschungsdaten aus ihrem Kontext ist ein großes Problem für den Erkenntnisgewinn.
- ☛ Die wenigsten Beteiligten sind bereit, darüber nachzudenken.

C. L. Borgman: *Big Data, Little Data, No Data: Scholarship in the Networked World*. MIT Press, Cambridge, MA, 2015. S. xix; Hervorhebung nicht im Original.

“ I've messed up early enough with my own data.

– t.b.

Voraussetzungen für intellektuellen Tiefgang im FDM

- ▶ eigene Erfahrung
 - eigenes Scheitern an der Komplexität und an den Anforderungen
 - Reflexion über die Gründe des Scheiterns als Ausgangspunkt
 - Eine Doktorarbeit reicht nicht aus.
- ▶ Kenntnis des Kontextes und der Vorarbeiten
 - Grundkompetenz und Voraussetzung für wissenschaftliches Arbeiten
 - Der NFDI vergleichbare Programme sind bereits gescheitert.
- ▶ Wissen über Werkzeuge aus anderen Disziplinen
 - Viele Probleme sind nicht spezifisch für FDM.
 - Es gibt gute Werkzeuge, nicht nur in der Softwareentwicklung.

Mangelndes Problembewusstsein

Persönliche Freiheit und intellektuelle Unmündigkeit

Fehlende oder fehlerhafte Anforderungsanalyse

Umsetzungsprobleme

These

Freiheit der Forschung und Wissenschaft ist nicht die Freiheit, nicht zu planen bzw. unstrukturiert vorzugehen, sondern die Freiheit, die Forschungsfrage selbst zu definieren und immer wieder zu ändern.

- ▶ Freiheit geht immer mit Verantwortung einher.
 - Je mehr Freiheit, desto mehr Verantwortung.
- ▶ Mangelnde Planung ist unprofessionell und meist eine Ausrede.
 - wird weder der Verantwortung gegenüber der Wissenschaft noch gegenüber der Gesellschaft gerecht
- ▶ Gründe für das Ändern der Forschungsfrage
 - Unvorhersehbarkeit von Ergebnissen
 - zunehmendes Verständnis der Fragestellung

Argument

- ▶ Angehenden Wissenschaftler/innen zu viele Vorgaben zu machen hindert sie daran, selbstständig zu werden und sich zu entfalten.

Entgegnung

- ▶ Anforderungen aus dem Wesen der Wissenschaft
 - Nachvollziehbarkeit
 - Qualität der Ergebnisse nach bestem Wissen und Gewissen
 - sorgfältiges und systematisches Vorgehen
- ▶ Verantwortung der Betreuenden
 - Ansprüche der Wissenschaft vorleben
 - Ansprüche der Wissenschaft explizit formulieren und vermitteln
 - Ansprüche der Wissenschaft einfordern
- ☛ Der Weg ist zweitrangig – wenn den Ansprüchen genügt wird.

“ *Aufklärung ist der Ausgang des Menschen aus seiner selbst verschuldeten Unmündigkeit. Unmündigkeit ist das Unvermögen, sich seines Verstandes ohne Leitung eines anderen zu bedienen. Selbstverschuldet ist diese Unmündigkeit, wenn die Ursache derselben nicht am Mangel des Verstandes, sondern der EntschlieÙung und des Muthes liegt, sich seiner ohne Leitung eines andern zu bedienen. Sapere aude! Habe Muth dich deines eigenen Verstandes zu bedienen! ist also der Wahlspruch der Aufklärung. Faulheit und Feigheit sind die Ursachen, warum ein so großer Theil der Menschen [...] gerne Zeitlebens unmündig bleiben [...]. Es ist so bequem, unmündig zu sein. [...] Ich habe nicht nöthig zu denken, wenn ich nur bezahlen kann; andere werden das verdrieÙliche Geschäft schon für mich übernehmen.*

– Immanuel Kant

I. Kant, *Berlinische Monatsschrift* 4:481–494, 1784

- ▶ äußere Form statt Inhalt
 - vgl. R. Feynman: „Cargo Cult Science“
 - Bsp.: Datenmanagementplan, Metadaten
 - Forschungsdatenmanagement als lästige Pflichtübung zu verstehen zeugt von mangelndem Verständnis der Wissenschaft.
- ▶ mangelnde Bereitschaft zur inhaltlichen Auseinandersetzung
 - Verständnis der Fragestellung ist eine Grundvoraussetzung für die Wahl von Werkzeugen und die Entwicklung von Strategien.
 - ohne inhaltliche Auseinandersetzung keine Fachkompetenz
 - ohne Verständnis kein kompetenter Einsatz der Werkzeuge und damit keine Qualitätssicherung der Ergebnisse
- ▶ mangelnde Bereitschaft zur Anpassung von Werkzeugen
 - Werkzeuge sind keine Lösungen. Lösungen entstehen durch den kompetenten und situationsangepassten Gebrauch von Werkzeugen.
 - Anpassung von Werkzeugen erfordert weiteren Kompetenzerwerb

Beobachtungssätze

- ▶ Menschen wissen i.d.R. nicht, was sie tun.
- ▶ Menschen denken i.d.R. nicht darüber nach, was sie tun.
- ▶ Menschen lesen i.d.R. keine Anleitungen/Dokumentation.
- ▶ Du und ich – wir sind auch (nur) Menschen

Das Ergebnis

- ▶ Werkzeuge: unvollständig, schlecht dokumentiert, schlecht wartbar
- ▶ Ergebnisse: nicht nachvollziehbar und von unbekannter Qualität
- ☛ Wir haben so gut wie nie die Chance, neu anzufangen – und wenn, dann fehlt uns i.d.R. die Erfahrung (aus den bereits gemachten Fehlern und der Reflexion darüber), um es besser zu machen.

These

Beschäftigung mit Forschungsdatenmanagement hält einen Spiegel vor und offenbart mitunter Defizite der eigenen Forschungstätigkeit.

- ▶ Forschungsdatenmanagement verstanden als sauberes wissenschaftliches Arbeiten mit dem Ziel der Nachvollziehbarkeit ist eine permanente Herausforderung.
- ▶ In der Realität sind viele Forschungsarbeiten nicht nachvollziehbar – und damit letztlich unwissenschaftlich und unbrauchbar.
- ▶ Das Wissenschaftssystem ist nicht auf Wissenschaftlichkeit bedacht, sondern auf Quantität und Neuheit. Etwas „ordentlich“ zu machen wird i.d.R. nicht honoriert.

“ *If experimentalists don't calibrate their equipment, check their reagents' [sic!] purity, and take careful notes, what they're doing isn't considered science. In contrast, computationalists don't even learn how to assess their software's quality in any systematic way, and very few would be able to recreate and rerun the programs they used to produce last year's papers. As a result, most computational science is irreproducible and of unknown quality.*

– Greg Wilson

- ▶ Experimentell Forschende sind nicht besser . . .
- ▶ Mangelnde Nachvollziehbarkeit führt zu *unbekannter* Qualität.
- ▶ Hinreichende Qualität ist Voraussetzung für Weiterverwendung.
- 👉 Wissenschaftlichkeit korreliert direkt mit Nachvollziehbarkeit.

Mangelndes Problembewusstsein

Persönliche Freiheit und intellektuelle Unmündigkeit

Fehlende oder fehlerhafte Anforderungsanalyse

Umsetzungsprobleme

These

Forschende sind meist nicht in der Lage, ihre Ansprüche an ein Forschungsdatenmanagement zu formulieren.

- ▶ Gründe sind ein mangelndes Bewusstsein für die Thematik, Unkenntnis der vorhandenen Werkzeuge und fehlende Kompetenz hinsichtlich ihrer Verwendung.
- ▶ Die Folge ist die Nutzung ungeeigneter Werkzeuge.
- ▶ Die Lösung sind nicht Forschungsdatenzentren oder Verantwortliche für Forschungsdatenmanagement, sondern die Vermittlung der notwendigen Kompetenzen in den universitären Curricula (und durch die Betreuenden).

These

Das falsche Werkzeug richtig einzusetzen ist genauso wenig zielführend, wie das richtige Werkzeug falsch zu verwenden.

- ▶ Voraussetzung der Anforderungsanalyse
 - hinreichendes Verständnis der Fragestellung
 - Abstraktionsvermögen
 - Kenntnis vorhandener potentiell relevanter Werkzeuge
- ▶ Aspekte
 - Auflistung der sich aus der Fragestellung ergebenden Ansprüche
 - Erhebung, welche Arten von Werkzeugen gebraucht werden
 - Priorisierung: nicht alles ist erreichbar, nicht alles verfügbar
- ▶ Bedeutung: sinnvoller Einsatz begrenzter Ressourcen
 - Werkzeuge müssen entwickelt/angepasst, definitiv erlernt werden.

These

Zentrale (einheitliche) Lösungen sind zum Scheitern verurteilt, weil sie die komplexe Realität nicht berücksichtigen (können).

- ▶ Wissenschaft ist sowohl viel zu komplex als auch viel zu individuell, als dass generelle Lösungen ohne tiefgreifende Anpassungen an die realen Gegebenheiten nutzbar wären.
- ▶ Der Ruf nach einheitlichen Werkzeugen und Austauschformaten verkennt die grundlegend individuellen Ansprüche jeglicher Forschung jenseits allgemeiner (und damit trivialer) Kriterien.
- ▶ (Digitale) Lösungen haben nur dann einen realen Mehrwert, wenn sie die Komplexität der Fragestellung berücksichtigen und entsprechend flexibel, modular und erweiterbar sind.

Mangelndes Problembewusstsein

Persönliche Freiheit und intellektuelle Unmündigkeit

Fehlende oder fehlerhafte Anforderungsanalyse

Umsetzungsprobleme

Problem

- ▶ Das Wissen über grundlegende Werkzeuge und Lösungen für das Forschungsdatenmanagement fehlt bei vielen Forschenden.

Mögliche Abhilfe

- ▶ Recherchieren und sich umhören, was andere machen:
Die meisten Probleme und Anforderungen sind weiter verbreitet.
- ▶ Liste bewährter Werkzeuge pflegen

Verantwortlichkeiten

- ▶ individuelle Forschende
 - eigenständige Recherche, auch jenseits des eigenen Fachgebiets
- ▶ Betreuende und Projektleitung
 - Stichwortgeber, Überblick über mögliche Werkzeuge

Problem

- ▶ Viele Forschende nehmen sich nicht die Zeit, sich in die Verwendung von Werkzeugen einzuarbeiten.

Mögliche Abhilfe

- ▶ Bewusstsein für die Notwendigkeit schärfen
- ▶ Verfügbarkeit und Qualität der Dokumentation sicherstellen
- ▶ zielgruppenangepasste Schulungsangebote etablieren

Verantwortlichkeiten

- ▶ individuelle Forschende
 - Bereitschaft zur Einarbeitung, reale Zeitinvestition
- ▶ Betreuende und Projektleitung
 - Einarbeitung fordern und Zeit und Ressourcen bereitstellen

Kompetenzniveaus

- 0 nicht kennen
- 1 kennen
- 2 verwenden
- 3 beherrschen
- 4 (weiter-)entwickeln

notwendig für wissenschaftliches Arbeiten

- ▶ Niveau 3, oft genug Niveau 4

typischerweise vorhanden

- ▶ Niveau 0 bis 2

Problem

- ▶ Forschungsdatenmanagement wird im Gesamtkontext betrachtet, die schiere Größe der Aufgabe scheint nicht zu bewältigen.

Mögliche Abhilfe

- ▶ Aufteilen in kleinere Aufgaben (Modularisierung)
- ▶ Priorisierung von Maßnahmen u.a. nach Kosten–Nutzen-Verhältnis
- ▶ die Aspekte umsetzen, deren Lösung bekannt und erreichbar ist

Verantwortlichkeiten

- ▶ individuelle Forschende
 - nicht aufgeben, aber Anforderungen im Blick behalten
- ▶ Betreuende und Projektleitung
 - Motivation, Hilfe bei Modularisierung und Priorisierung

Problem

- ▶ Formalisierung, Automatisierung und Digitalisierung von Abläufen fehlen in der Routine. Die Folge: manuelle Mehrarbeit und Fehler.

Mögliche Abhilfe

- ▶ Abläufe durchdenken, formalisieren, automatisieren, digitalisieren
- ▶ digitale Werkzeuge plattformunabhängig verfügbar machen
- ▶ bei Problemen mit digitalen Lösungen mindestens formalisieren

Verantwortlichkeiten

- ▶ individuelle Forschende
 - Abläufe mindestens durchdenken und formalisieren
- ▶ Betreuende und Projektleitung
 - digitale Werkzeuge nach Möglichkeit zur Verfügung stellen



- 🔑 Forschungsdatenmanagement setzt ein Problembewusstsein für seine Relevanz für die Wissenschaftlichkeit voraus.
- 🔑 Das aktuelle Wissenschaftssystem setzt falsche Anreize mit seinem Fokus auf Quantität statt Qualität.
- 🔑 Reflexion über Forschungsdatenmanagement offenbart mitunter Defizite in der Wissenschaftlichkeit der eigenen Forschung.
- 🔑 Fehlende bzw. fehlerhafte Anforderungsanalyse führt zu unbrauchbaren Werkzeugen und Lösungen.
- 🔑 Umsetzungsprobleme entstehen meist durch mangelnde Kenntnis und Kompetenz beim Einsatz existierender Werkzeuge.