

Forschungsdatenmanagement

Notwendige, aber nicht hinreichende Voraussetzung
für den wissenschaftlichen Erkenntnisgewinn

06. Erheben

Dr. habil. Till Biskup

Physikalische Chemie

Universität Rostock

03.05.2024





- 🔑 Datenerhebung inkl. zugehöriger Metadaten muss systematisch erfolgen und hinreichend dokumentiert werden.
- 🔑 Qualitätskontrolle bereits während der Datenerhebung ist essentiell für einen sinnvollen Ressourceneinsatz.
- 🔑 Organisatorische Aspekte umfassen Schemata für Dateinamen und Verzeichnisse genauso wie ELNs und PIDs.
- 🔑 Rechtliche Aspekte wie Datenschutz und Persönlichkeitsrechte müssen bei der Datenerhebung konsequent beachtet werden.
- 🔑 Praktisch verantwortlich sind die einzelnen Forschenden, die Projektleitung für die Etablierung der Prozesse.

Der Forschungsdatenlebenszyklus

Modell der wissenschaftlichen Methode aus Sicht der Forschungsdaten



Wissenschaftliche Aspekte

Organisatorische Aspekte

Verantwortung

Anforderungen und notwendige Werkzeuge

“ *If experimentalists don't calibrate their equipment, check their reagents' [sic!] purity, and take careful notes, what they're doing isn't considered science.*

– Greg Wilson

- ▶ Die charmant naive Sichtweise eines Informatikers...

🎵 Leitmotiv

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

- ▶ Es fehlt uns meist an Bewusstsein und Disziplin (und Willen), alle relevanten Informationen (nicht Daten) zu erheben und mit zu veröffentlichen.

- ▶ Vorgehen vom Bekannten zum Unbekannten
 - Wissenschaft findet immer in einem Kontext statt.
 - Reproduzieren vorheriger Ergebnisse mit eigenem Equipment als Ausgangspunkt für die eigene Forschung (vgl. R. Feynman)
- ▶ Systematisches und korrektes Vorgehen
 - korrekter Umgang mit Geräten/Experimenten (Kalibrierung, Testmessungen, Parameterbestimmung, ...)
 - unter Berücksichtigung des aktuellen Stands der Wissenschaft
 - Messung: Beobachtung unter kontrollierten Bedingungen
- ▶ Qualitätskontrolle während der Datenaufnahme
 - Konsistenzprüfung
 - Kriterien vorab festgelegt, aus der Erfahrung gewonnen
- ▶ Metadaten und Dokumentation: Kontext der Datenerhebung
 - Daten ohne Kontext sind wertlos – und sollten gelöscht werden.

These

Die Menge verfügbarer Informationen ist während der Datenaufnahme maximal, eine sinnvolle Reduktion die eigentliche Herausforderung.

- ▶ Datenerhebung findet immer in einem Kontext statt.
 - Einzelfallentscheidung, was zur Spezifität des Kontextes gehört
- ▶ Metadaten sollten sechs Fragen beantworten:
 - Wer hat
 - was
 - mit wem
 - wann,
 - wie und
 - warum gemacht?

Wissenschaftliche Aspekte

Organisatorische Aspekte

Verantwortung

Anforderungen und notwendige Werkzeuge

- ▶ Vorgehen vom Bekannten zum Unbekannten
 - relevante vorherige Ergebnisse identifizieren
 - Strategien zum Reproduzieren mit dem eigenen Equipment
- ▶ Systematisches und korrektes Vorgehen
 - Protokolle für den Umgang mit Geräten/Experimenten (Kalibrierung, Testmessungen, Parameterbestimmung, ...)
 - Forschende entsprechend einlernen
- ▶ Qualitätskontrolle während der Datenaufnahme
 - Kriterien für die Konsistenzprüfung festlegen
 - auf der Erfahrung anderer aufbauen
- ▶ Metadaten und Dokumentation: Kontext der Datenerhebung
 - Datenstrukturen, Schemata für Metadatenerhebung
 - ggf. Formalisierung von Protokollen zur Datenerhebung

- ▶ Datenspeicherung
 - (lokales) Repositorium für „warme“ Forschungsdaten
 - Schemata für Datei-/Verzeichnisnamen:
intuitiv, modular, eindeutig
 - erweiterbare und nachvollziehbare Verzeichnisstrukturen
- ▶ Datenschutz
 - Grundprinzip der Datensparsamkeit: Pseudonymisierung oder Anonymisierung ist im Zweifelsfall nicht dauerhaft umsetzbar
- ▶ Dokumentation: Laborbuch und Metadaten
 - unabhängig davon, ob analog oder digital
 - systematisch, nachvollziehbar für andere,
vollständig bzgl. der für die Auswertung notwendigen Informationen
- ▶ eindeutige Bezeichner (PIDs) für Proben etc.
 - einfachster Fall: durchnummerieren
 - Zuordnungstabelle redundant sichern

Wissenschaftliche Aspekte

Organisatorische Aspekte

Verantwortung

Anforderungen und notwendige Werkzeuge

- ▶ wissenschaftliche Projektleitung/Gruppenleitung
 - Ermöglichen und Einfordern von Datenerhebung die den Ansprüchen der Wissenschaft genügt
 - Qualifizierung der Mitarbeitenden
- ▶ individuelle Forschende, die die Daten erheben
 - gewissenhaftes Arbeiten gemäß den Ansprüchen der Wissenschaft
 - eigenverantwortliche Qualifikation und Wissenserwerb

Leitmotiv

Forschungsdatenmanagement ist primär die Verantwortung der individuellen Forschenden.

- ☞ Datenerhebung ist ureigene Aufgabe der Forschenden.
- ☞ Institutionelle Verantwortlichkeiten sind quasi nicht erkennbar.

- ▶ Vorgehen vom Bekannten zum Unbekannten
 - wissenschaftlichen Kontext bereitstellen und aufzeigen
 - Reproduzieren bekannter Ergebnisse aktiv einfordern
- ▶ Systematisches und korrektes Vorgehen
 - Etablierung von Abläufen, Strukturen und Konventionen
 - regelmäßige Überprüfung, ob die Abläufe eingehalten werden und noch angemessen sind bzw. angepasst werden müssen
- ▶ Qualitätskontrolle während der Datenaufnahme
 - Kriterien zur Qualitätssicherung festlegen
 - Qualität der erhobenen Daten (mit Werkzeugen) überprüfen
- ▶ Metadaten und Dokumentation: Kontext der Datenerhebung
 - Strukturen und Werkzeuge etablieren
 - Einsatz der Werkzeuge motivieren und einfordern
 - Qualität der Dokumentation regelmäßig überprüfen

- ▶ Vorgehen vom Bekannten zum Unbekannten
 - eigenständige Recherche des notwendigen konkreten Kontextes
 - gewissenhafte erfolgreiche Reproduktion bekannter Ergebnisse
- ▶ Systematisches und korrektes Vorgehen
 - etablierte Abläufe, Strukturen und Konventionen einhalten
 - kritisches Hinterfragen aus der Praxis
 - Strukturen (eigenständig) weiterentwickeln, zur Diskussion stellen
- ▶ Qualitätskontrolle während der Datenaufnahme
 - Kriterien und Werkzeuge zur Qualitätskontrolle einsetzen
 - Datenaufnahme ggf. wiederholen
 - (fragliche) Qualität der Daten ggf. in Metadaten vermerken
- ▶ Metadaten und Dokumentation: Kontext der Datenerhebung
 - etablierte Strukturen und Werkzeuge einsetzen
 - kritisch auf Vollständigkeit und Handhabbarkeit hinterfragen, ggf. (eigenständig) weiterentwickeln und zur Diskussion stellen

Wissenschaftliche Aspekte

Organisatorische Aspekte

Verantwortung

Anforderungen und notwendige Werkzeuge

“ *Civilization advances by extending the number of important operations which we can perform without thinking about them.*
– Alfred North Whitehead

► Dokumentation und Metadatenaufnahme

- maschinenlesbar und *menschenschreibbar*
- niederschwellig und einfach bedienbar/nutzbar
- Vorlagen und Schemata: nicht nachdenken müssen, welche Information erhoben werden muss (letztlich Formulare)
- flexibel erweiterbar (unter Erhalt der Kompatibilität)
- Kommentarfeld für Informationen, die bislang nicht vom Schema strukturiert erfasst werden
- Zweck der Datenerhebung als eigenes Feld

- ▶ Datenablage
 - eindeutiger Ort für gegebene Daten
 - modulare, erweiterbare, nachvollziehbare hierarchische Struktur
 - Schemata für Benennung
 - sinnvolle Sortierbarkeit (Namensschemata)
 - Konventionen und technische Lösungen, um Metadaten gemeinsam mit den Daten abzulegen
- ▶ eindeutige Bezeichner
 - eindeutig im gegebenen Kontext
 - hinreichend kurz und menschenlesbar (ggf. Kurzform)
 - eindeutiger Ort für die Ablage zugehöriger Informationen
- ▶ Protokolle für die Datenerhebung
 - hinreichend detaillierte Beschreibung des Vorgehens
 - praxisnah und einfach abzuwickeln (unabhängig ihrer Komplexität)

- ▶ Schema für Metadaten
 - erspart das Nachdenken, welche Metadaten relevant sind
 - sorgt für Maschinenlesbarkeit
- ▶ Format für Metadatenerhebung
 - Erhebung *während* der Datenaufnahme
 - plattformunabhängig, nicht notwendigerweise digital
- ▶ Strukturen und Protokolle für die Datenerhebung
 - ermöglichen die notwendige Datenqualität
 - Verfahrensanweisung (*standard operating procedure*, SOP)
- ▶ Konventionen für die strukturierte Datenablage
 - Namen für Verzeichnisse und Dateien, Verzeichnishierarchien
- ▶ Mechanismen zur Qualitätskontrolle
 - möglichst früh (unmittelbar nach Erhebung): Aufwandsminimierung
 - Konsistenzüberprüfung, spezifische Qualitätskriterien

- ▶ Einsatz möglichst früh: während/direkt nach der Datenerhebung
 - Korrekturen mit minimalem Ressourceneinsatz möglich
 - Bsp.: Kontrolle *vor* dem Probenausbau
 - Bsp.: optische/akustische Kontrolle *während* der Datenaufnahme
- ▶ Werkzeuge zur Darstellung der Charakteristika primärer Daten
 - auch Menschen, die Daten nur erheben, sollen eigenständig entscheiden können, ob die Datenaufnahme erfolgreich war.
- ▶ Voraussetzung zur Erstellung dieser Werkzeuge und Mechanismen
 - intellektuelle Durchdringung des Gesamtvorgangs
 - Fähigkeit, dieses Verständnis in Werkzeuge zu gießen
- ▶ Beispiele entsprechender Werkzeuge
 - einfache Metriken („Wenn der Peak nicht da kommt ...“)
 - komplexe Auswertungsverfahren und -Software
 - grafische Anzeigeprogramme mit integrierter Auswertung



- 🔑 Datenerhebung inkl. zugehöriger Metadaten muss systematisch erfolgen und hinreichend dokumentiert werden.
- 🔑 Qualitätskontrolle bereits während der Datenerhebung ist essentiell für einen sinnvollen Ressourceneinsatz.
- 🔑 Organisatorische Aspekte umfassen Schemata für Dateinamen und Verzeichnisse genauso wie ELNs und PIDs.
- 🔑 Rechtliche Aspekte wie Datenschutz und Persönlichkeitsrechte müssen bei der Datenerhebung konsequent beachtet werden.
- 🔑 Praktisch verantwortlich sind die einzelnen Forschenden, die Projektleitung für die Etablierung der Prozesse.