

# Forschungsdatenmanagement

Notwendige, aber nicht hinreichende Voraussetzung  
für den wissenschaftlichen Erkenntnisgewinn

## 09. Veröffentlichen

Dr. habil. Till Biskup

Physikalische Chemie

Universität Rostock

24.05.2024





- 🔑 Die Veröffentlichung von Forschungsdaten und der gewonnenen Erkenntnisse ist ein Kernprinzip der Wissenschaft.
- 🔑 Veröffentlichung dient der direkten (Begutachtung) und späteren (Reproduzierbarkeit) Qualitätskontrolle.
- 🔑 Veröffentlicht werden sollten neben den Erkenntnissen die zugrundeliegenden Daten und Auswertungswerkzeuge.
- 🔑 Voraussetzung für die Veröffentlichung von Daten und Werkzeugen ist ihre Kuration und Dokumentation.
- 🔑 Verantwortlich sind Projektleitung und Forschende, Institutionen für die Bereitstellung geeigneter Plattformen.

# Der Forschungsdatenlebenszyklus

Modell der wissenschaftlichen Methode aus Sicht der Forschungsdaten



### Leitmotiv

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

- ▶ Ein Hauptproblem sind die falschen Anreize im aktuellen Wissenschaftssystem: „publish or perish“.
- ▶ Qualitätssicherung durch Begutachtung ist die beste aller schlechten Lösungen, funktioniert aber auch nur eingeschränkt.
- ▶ Nur ein Bewusstsein für die Ansprüche der Wissenschaft und frühzeitige Vermittlung notwendiger Kompetenzen zum wissenschaftlichen Vorgehen hat Aussicht auf Erfolg.

Wissenschaftliche Aspekte

Organisatorische Aspekte

Verantwortung

Anforderungen und notwendige Werkzeuge

“ *Scientists regard the free interchange of knowledge and insights as essential, and, in consequence, being non-secretive is one of their rules of professional conduct.*

– Edsger Dijkstra

## ▶ Kernprinzip der Wissenschaft

- Nur wenn Erkenntnisse (und ggf. die zugrundeliegenden Daten) veröffentlicht werden, können andere darauf aufbauen.
- Wissenschaft findet immer in einem Kontext statt.

## ▶ Welche Daten sind veröffentlichungswürdig?

- grundsätzlich alle Forschungsdaten im weiteren Sinn
- insbesondere Ergebnisse und Werkzeuge
- vollständige Protokolle der Datenverarbeitung: Transparenz
- ggf. Rohdaten (aber mit Auswertungsprotokoll und Werkzeugen)
- je nach Kontext auch physische Artefakte

E. W. Dijkstra, „On the Role of Scientific Thought“  
in: Selected Writings on Computing: A Personal Perspective. Springer, New York 1982

- ▶ Zielgruppe: andere Wissenschaftler
  - nicht notwendigerweise identischer fachlicher Hintergrund
  - stellt ganz andere Anforderungen hinsichtlich der Dokumentation und Kontextualisierung der zu veröffentlichenden Forschungsdaten
- ▶ Zitierfähigkeit
  - Rückverfolgbarkeit und damit Transparenz
  - Nebeneffekt: wissenschaftliche Anerkennung
- ▶ Qualitätskontrolle
  - direkt: Begutachtung während des Veröffentlichungsprozesses
  - später: Reproduzierbarkeit (im weiteren Sinn) mit dem Ziel der Verallgemeinerbarkeit
- ▶ Voraussetzungen
  - Kuration und Dokumentation
  - Kontextualisierung und explizite Beschreibung der Ergebnisse

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

– Art. 5, Abs. 3 GG

- ▶ Freiheit der Forschung: Veröffentlichung ist kein Zwang
  - Viele Fördermittelgeber und Herausgeber von Zeitschriften wollen die Veröffentlichung erzwingen.
  - Verständnis von Daten (und Forschung) als ökonomische Ressource
- ▶ Sonderfall: moralische Verpflichtung
  - insbesondere bei Studien mit Tieren oder Menschen
  - Veröffentlichung der Daten, egal wie das Ergebnis lautet
  - Tendenz: Präregistrierung der Studien
  - Abwägung von Rechtsgütern

Wissenschaftliche Aspekte

Organisatorische Aspekte

Verantwortung

Anforderungen und notwendige Werkzeuge

## ▶ Auswahl

- „Datenflut“: nicht alle (Roh-)Daten sollten veröffentlicht werden
- Relevante Daten zurückzuhalten ist genauso wenig hilfreich wie der Verzicht auf eine sinnvolle (und notwendige) Auswahl.
- mangelnde Auswahl ist Ausdruck der Geringschätzung anderer Forschender und mangelnder Professionalität

## ▶ Nachvollziehbarkeit

- umfassende Metadaten für alle bisherigen Schritte, idealerweise maschinenverarbeitbar
- Zielgruppe: andere Wissenschaftler – deshalb noch wichtiger (und tendenziell noch schwieriger zu gewährleisten)

## ▶ Transparenz

- Wege der Entscheidungsfindungen:  
Fokus auf dem Warum (nicht), nicht nur dem Was Wie

- ▶ schriftliche Abhandlungen
  - Bücher
  - Artikel in Fachzeitschriften
- ▶ digitale Daten
  - digitaler Appendix/SI
  - (externes) (Fach-)Repositoryum
  - eigene Webseite
  - Datenjournale
- ▶ digitale Werkzeuge
  - Code-Plattformen (GitHub, ...)
  - Software-Journale (JOSS etc.)
- ▶ physische Artefakte
  - Museen und Sammlungen
  - biologische Proben: z.B. DSMZ

- ▶ Qualitätskontrolle
  - Nicht begutachtete Datenpublikationen sind i.d.R. wenig wert, da Nachvollziehbarkeit nicht systematisch überprüft
  - Begutachtung in ihrer aktuellen Form funktioniert nicht gut, ist aber vermutlich das beste aller schlechten Werkzeuge.
  - Qualität lässt sich i.d.R. nicht algorithmisch überprüfen, sondern bedarf der menschlichen Interaktion.
- ▶ digitale Daten: Datenformate
  - offen, nicht-proprietär, möglichst selbstbeschreibend
  - Daten und maschinenverarbeitbare Metadaten
- ▶ digitale Werkzeuge: Quellcode
  - lesbar, nachvollziehbar, ... – „sauberer Code“
  - Versionierung, eindeutige Versionsnummer

- ▶ Embargo und Zugriffskontrolle
  - legitime Einschränkung des Zugriffs auf die Daten/Ergebnisse
  - ggf. zeitlich beschränkt
- ▶ Schutzrechte
  - Datenschutz, Persönlichkeitsrechte, Datenbankrechte, . . .
  - können der Veröffentlichung (prinzipiell) entgegenstehen
- ▶ Patente
  - Veröffentlichung verwirkt i.d.R. Patentierbarkeit
- ▶ Lizenzen
  - Übertragung von Nutzungsrechten
- ▶ Urheberrecht und Autorschaft
  - Software: Urheberrecht i.d.R. beim Arbeitgeber
  - Rechte an Daten: oft nicht trivial – Autorschaft?

Wissenschaftliche Aspekte

Organisatorische Aspekte

Verantwortung

Anforderungen und notwendige Werkzeuge

- ▶ primäre Verantwortung
  - wissenschaftliche Projektleitung/Gruppenleitung
  - individuelle Forschende
- ▶ Unterstützung
  - Institutionen
  - Fachgesellschaften, Fördermittelgeber etc.

### Leitmotiv

Forschungsdatenmanagement ist primär die Verantwortung der individuellen Forschenden.

- ☞ Veröffentlichung von Ergebnissen ist Kernaspekt der Wissenschaft
- ☞ Institutionen/Fachgesellschaften sind lediglich unterstützend tätig.

- ▶ wissenschaftliche Projektleitung/Gruppenleitung
  - Qualitätssicherung der veröffentlichten Ergebnisse
  - Einschätzung der Relevanz von Ergebnissen
  - Kontextualisierung im größeren Maßstab
  - Qualitätskontrolle der schriftlichen (Begleit-)Publikation
  - darf nicht zum Hemmnis bei Veröffentlichungen auf Kosten der abhängigen Forschenden werden, weder zeitlich noch inhaltlich
  
- ▶ individuelle Forschende
  - Auswahl der Daten
  - ggf. Konversion digitaler Daten in offene Formate
  - Kuratation der Daten: hinreichende Dokumentation
  - Kontextualisierung der Daten und Ergebnisse
  - mindestens Entwurf einer schriftlichen (Begleit-)Publikation
  - Redlichkeit und Transparenz:  
veröffentlichte Ergebnisse sollten repräsentativ sein

- ▶ Institutionen
  - Bereitstellung geeigneter Plattformen: Repositorien etc.
  - langfristige Sicherung der Verfügbarkeit der Infrastruktur
  - lokale Institution: ggf. Unterstützung bei der Auswahl geeigneter Plattformen
- ▶ Fachgesellschaften
  - Etablierung von Standards für offene Formate und Metadatenschemata
  - Unterstützung/ggf. Etablierung einschlägiger Fachrepositorien
- ▶ Fördermittelgeber
  - Finanzierung der zusätzlichen Kosten für Open-Access-Publikationen
  - langfristige Finanzierung der notwendigen Infrastruktur

Wissenschaftliche Aspekte

Organisatorische Aspekte

Verantwortung

Anforderungen und notwendige Werkzeuge

- ▶ Zitierfähigkeit
  - Transparenz: Rückverfolgbarkeit von Aussagen auf Quellen
  - essentielles Kriterium der Wissenschaft, das gerade droht, komplett verloren zu gehen
  - Nebeneffekt: wissenschaftliche Anerkennung (mitunter sehr spät)
- ▶ Qualitätskontrolle
  - Nachvollziehbarkeit, ggf. Reproduzierbarkeit
  - Überprüfung der Datenverarbeitung
  - Interpretationen ändern sich, Ergebnisse sollten Bestand haben
- ▶ Transparenz und Nachvollziehbarkeit
  - Offenlegung betrachteter und verworfener Alternativen
  - ggf. Nennung nicht berücksichtigter Ergebnisse
  - vollständiges Protokoll der Datenverarbeitung
  - hinreichende Beschreibung der Datenerhebung

- ▶ langfristige Verfügbarkeit veröffentlichter Ergebnisse
  - Wissenschaft ist generationenübergreifend, nicht kurzfristig
  - Relevanz von Ergebnissen lässt sich nicht vorhersagen
  - Langfristig bedeutet sehr viel länger als zehn Jahre.
  - Der Wert von Daten und Ergebnissen ergibt sich mitunter erst durch weiteren Kontext und zusätzliche Erkenntnisse.
  
- ▶ Auffindbarkeit
  - andere Forschende müssen eine reale Chance haben, die veröffentlichten Ergebnisse zu finden
  - Sprache ist ggf. die geringste Barriere

- ▶ Zitierfähigkeit
  - Zitationsformate jenseits etablierter schriftlicher Dokumente (Daten, digitale Werkzeuge)
  - eindeutige, stabile Kennungen (PIDs): DOI, ARK, ...
- ▶ Qualitätskontrolle
  - Schemata und etablierte Formate für Metadaten
  - Checklisten für Minimalstandards
  - ggf. institutionelle Hilfe bei der Datenkuration
  - Begutachtung von Datenpublikationen
- ▶ Transparenz und Nachvollziehbarkeit
  - schriftliche Kontextualisierung der publizierten Daten
  - Metadaten für Datenerhebung und -Verarbeitung (idealerweise maschinenverarbeitbar und mit etablierten Schemata)

- ▶ langfristige Verfügbarkeit veröffentlichter Daten
  - für gedruckte Textpublikationen hinreichend gelöst
  - für digitale Artefakte eine ungelöste Herausforderung  
betrifft auch nur noch elektronisch publizierte Zeitschriften
  - offene Datenformate
  - Strategien zum Transfer auf jeweils neue Speichermedien
  - physische Artefakte: langfristiger, sicherer Aufbewahrungsort
- ▶ Auffindbarkeit
  - spezialisierte Suchmaschinen  
(NCBI PubMed, Web of Science, Google Scholar, ...)
  - Kataloge (analog und digital)
  - gute Überblicksartikel, Literaturstudien, Monographien



- 🔑 Die Veröffentlichung von Forschungsdaten und der gewonnenen Erkenntnisse ist ein Kernprinzip der Wissenschaft.
- 🔑 Veröffentlichung dient der direkten (Begutachtung) und späteren (Reproduzierbarkeit) Qualitätskontrolle.
- 🔑 Veröffentlicht werden sollten neben den Erkenntnissen die zugrundeliegenden Daten und Auswertungswerkzeuge.
- 🔑 Voraussetzung für die Veröffentlichung von Daten und Werkzeugen ist ihre Kuration und Dokumentation.
- 🔑 Verantwortlich sind Projektleitung und Forschende, Institutionen für die Bereitstellung geeigneter Plattformen.