

# Forschungsdatenmanagement

Notwendige, aber nicht hinreichende Voraussetzung  
für den wissenschaftlichen Erkenntnisgewinn

## 10. Wiederverwenden

Dr. habil. Till Biskup

Physikalische Chemie

Universität Rostock

24.05.2024





- 🔑 Wissenschaft gründet auf den Vorarbeiten anderer.  
Wiederverwendung ist Voraussetzung für den Erkenntnisfortschritt.
- 🔑 Primärer Fokus ist die interne Nachnutzbarkeit.  
Externe Nachnutzung stellt hohe Ansprüche an die Dokumentation.
- 🔑 Wiederverwendung numerischer Daten kommt fast nie vor:  
Aufbauen auf den Ergebnissen/Erkenntnissen anderer.
- 🔑 Voraussetzungen sind Kuration und Qualitätssicherung,  
offene Datenformate, Metadaten und Konventionen/Standards.
- 🔑 Verantwortlich sind Projektleitung und Forschende,  
Institutionen für die Bereitstellung geeigneter Infrastrukturen.

# Der Forschungsdatenlebenszyklus

Modell der wissenschaftlichen Methode aus Sicht der Forschungsdaten



### Leitmotiv

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

- ▶ Qualitätssicherung durch Begutachtung ist die beste aller schlechten Lösungen, funktioniert aber auch nur eingeschränkt.
- ▶ Forschende sind für die hinreichende Qualitätskontrolle der von ihnen verwendeten Ergebnisse verantwortlich.
- ▶ Die meisten veröffentlichten Forschungsergebnisse sind bestenfalls irrelevant, schlimmstenfalls irreführend und/oder falsch.
- ▶ Die schiere Masse veröffentlichter Forschungsdaten erschwert die Auswahl relevanter Ergebnisse unverhältnismäßig („DoS“).

Wissenschaftliche Aspekte

Organisatorische Aspekte

Verantwortung

Anforderungen und notwendige Werkzeuge

**G I G O**

“ *If I have seen further  
it is by standing on y<sup>e</sup> shoulders of giants.*

– Sir Isaac Newton

- ▶ Wiederverwenden ist ein Kernprinzip der Wissenschaft.
  - entscheidende Frage: was wiederverwenden?
  - Newton: Wiederverwendung von *Erkenntnissen* und Ergebnissen anderer, *nicht* Wiederverwendung der (Roh-)Daten
- ▶ Fokus der Vorlesung: persönliches Forschungsdatenmanagement
  - in erster Linie Wiederverwendung im eigenen Kontext
- ▶ Fataler Fehler von „Big Data“ (und FAIR)
  - verfügbare Daten definieren die Forschungsfrage, nicht umgekehrt
- ▶ vier Kriterien
  - Qualität, Relevanz, Repräsentativität, Rückführbarkeit

### ► Qualität

- Qualität verwendeter Ergebnisse/Erkenntnisse ist oft entscheidend für die Qualität eigener Schlussfolgerungen
- Qualität ist nicht immer offensichtlich, mangelnde Qualität mitunter einfacher festzustellen.
- Qualität ist *nicht* gleichbedeutend mit Richtigkeit/Korrektheit.
- Qualitätsansprüche können sich je nach Fragestellung unterscheiden: dieselben Daten können geeignet sein oder nicht.

### ► Relevanz

- Die meisten verfügbaren Informationen sind für eine gegebene Fragestellung nicht relevant (vgl. Metadaten zur Datenerhebung).
- Die Auswahl der relevanten Informationen ist ein entscheidender Aspekt der Wissenschaft – und Professionalität.
- Relevanz muss immer im Kontext beantwortet werden.
- Kriterien für Relevanz: bekannt, benannt, dokumentiert, diskutiert

### ▶ Repräsentativität

- Idee wissenschaftlicher Modelle, Hypothesen, Theorien: Herleitung allgemein(er) gültiger (kausaler) Zusammenhänge
- Mustererkennung sagt nichts über die Repräsentativität der zugrundeliegenden Daten aus: Mustererkennung ist ein Algorithmus.
- (mögliche) Beschränkung der Repräsentativität verwendeter Daten mindestens benennen und diskutieren (Transparenz, Redlichkeit)
- Datenmenge und Repräsentativität sind komplett unabhängig. Entscheidend sind *unabhängige* Daten/Beobachtungen.

### ▶ Rückführbarkeit

- Ergebnisse/Erkenntnisse müssen auf ihre Quelle rückführbar sein.
- Voraussetzung für Transparenz, Nachvollziehbarkeit und Überprüfbarkeit und damit Qualitätssicherung
- Rückführbarkeit ist notwendiges, aber nicht hinreichendes Kriterium für Datenqualität

Wissenschaftliche Aspekte

Organisatorische Aspekte

Verantwortung

Anforderungen und notwendige Werkzeuge

## ▶ Qualität

- Kontextualisierung: hinreichende Dokumentation
- Konsistenz: keine inneren Widersprüche in den Daten?
- Plausibilität: Sind die Ergebnisse sinnvoll/möglich?
- Voraussetzung: Kenntnis der wissenschaftlichen Zusammenhänge

## ▶ Relevanz

- klare und überprüfbare Kriterien, vorab festlegen
- setzt Vertrautheit mit der wissenschaftlichen Fragestellung voraus

## ▶ Repräsentativität

- Überblick über die möglichen Ergebnisse/Daten
- Kenntnis von Werkzeugen, um Repräsentativität sicherzustellen

## ▶ Rückführbarkeit

- eindeutige Quellenangaben, stabile Verfügbarkeit der Quellen
- Verfügbarkeit aller relevanten Informationen zur Nachvollziehbarkeit

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.

- 👉 Wiederverwenden von Ergebnissen ist Kernaspekt der Wissenschaft
- 👉 Institutionen/Fachgesellschaften sind lediglich unterstützend tätig.

- ▶ Qualität
  - Kriterien für Dokumentation, Konsistenz, Plausibilität aufstellen
  - umfassende Kenntnis relevanter wissenschaftlicher Zusammenhänge
- ▶ Relevanz
  - überprüfbare Kriterien vorab festlegen
  - Vertrautheit mit der wissenschaftlichen Fragestellung
- ▶ Repräsentativität
  - umfassender Überblick über die möglichen Ergebnisse/Daten
  - Kenntnis von Werkzeugen, um Repräsentativität sicherzustellen
- ▶ Rückführbarkeit
  - Quellenangaben und Referenzen einfordern
  - alle für die Nachvollziehbarkeit relevanten Informationen einfordern

- ▶ Qualität
  - Kriterien anwenden, Daten *vor* Nutzung auf Qualität prüfen
  - Kenntnis relevanter wissenschaftliche Zusammenhänge aneignen
- ▶ Relevanz
  - Kriterien anwenden, Daten entsprechend aussuchen/filtern
  - mit wissenschaftlicher Fragestellung hinreichend vertraut machen
- ▶ Repräsentativität
  - Bewusstsein für die Probleme mangelnder Repräsentativität
  - Werkzeuge zur Sicherstellung der Repräsentativität einsetzen
- ▶ Rückführbarkeit
  - für alle ausgewählten Daten/Ergebnisse Referenzen angeben
  - Nachvollziehbarkeit überprüfen

- ▶ Qualität
  - Kriterien entwickeln: Was macht die Qualität von Daten aus?
  - Qualitätskontrolle verwendeter Daten einfordern
- ▶ Relevanz
  - Etablierung von Metadatenschemata zur Beschreibung von Daten und Ergebnissen, die eine akkurate Auswahl/Filterung erlauben
- ▶ Repräsentativität
  - Entwicklung und Etablierung entsprechender Kataloge
  - Bereitstellung von Infrastrukturen, um Daten vorzuhalten
- ▶ Rückführbarkeit
  - Quellenangaben und vollständige Dokumentation einfordern
  - notwendige Infrastrukturen entwickeln, etablieren, vorhalten

Wissenschaftliche Aspekte

Organisatorische Aspekte

Verantwortung

Anforderungen und notwendige Werkzeuge

- ▶ Qualität
  - Überprüfbare Kriterien: Was bedeutet Datenqualität?
  - Kuration: Dokumentation, Nachvollziehbarkeit, Transparenz
- ▶ Relevanz
  - klare, überprüfbare Kriterien
  - Auszeichnung von Daten/Ergebnissen als relevant
- ▶ Repräsentativität
  - Verfügbarkeit einer repräsentativen Stichprobe
  - Benennung evtl. Einschränkungen
- ▶ Rückführbarkeit
  - Nachvollziehbarkeit von der Veröffentlichung bis zur ursprünglichen Beobachtung inkl. aller Zwischenschritte
  - langfristige Verfügbarkeit der relevanten Informationen

- ▶ Qualität
  - Werkzeuge zur möglichst automatisierbaren Qualitätskontrolle
  - Konventionen/Standards für Metadaten(schemata) und Datenformate
- ▶ Relevanz
  - Werkzeuge zur möglichst automatischen Vorauswahl
  - Strukturen zur lokalen Ablage der relevanten Ergebnisse
- ▶ Repräsentativität
  - einschlägige umfassende und durchsuchbare Kataloge
  - Repositorien zum Zugriff auf die Daten
- ▶ Rückführbarkeit
  - Zitierfähigkeit
  - dauerhafte Kennungen (PID) und langfristige Verfügbarkeit



- 🔑 Wissenschaft gründet auf den Vorarbeiten anderer.  
Wiederverwendung ist Voraussetzung für den Erkenntnisfortschritt.
- 🔑 Primärer Fokus ist die interne Nachnutzbarkeit.  
Externe Nachnutzung stellt hohe Ansprüche an die Dokumentation.
- 🔑 Wiederverwendung numerischer Daten kommt fast nie vor:  
Aufbauen auf den Ergebnissen/Erkenntnissen anderer.
- 🔑 Voraussetzungen sind Kuration und Qualitätssicherung,  
offene Datenformate, Metadaten und Konventionen/Standards.
- 🔑 Verantwortlich sind Projektleitung und Forschende,  
Institutionen für die Bereitstellung geeigneter Infrastrukturen.