

# Handout "Datenqualität"

## Worum geht es?



Um Datenqualität, die sich in den folgenden vier Dimensionen zeigen kann und jeweils folgende Aspekte betrifft:

- **intrinsische** Datenqualität (= betrifft die Daten selbst)
  - ➔ Glaubwürdigkeit, Fehlerfreiheit, Objektivität
- **kontextuelle** Datenqualität (= zweckabhängig)
  - ➔ Mehrwert, Relevanz, Aktualität, Vollständigkeit, angemessener Umfang
- **repräsentationelle** Datenqualität (= darstellungsbezogen)
  - ➔ eindeutige Auslegbarkeit, Verständlichkeit, einheitliche Darstellung, Übersichtlichkeit
- **Zugriffsqualität** (= systembezogen):
  - ➔ Bearbeitbarkeit, Zugänglichkeit

---

## Warum ist das wichtig?



Eine größtmögliche Datenqualität muss gewährleistet sein, damit die Gestaltung von wissenschaftlichen Modellen oder die Herleitung von wissenschaftlichen Erkenntnissen fundiert und aussagekräftig ist. Bei Software und Skripten muss außerdem gewährleistet sein, dass diese auf jedem beliebigen Rechner ausgeführt werden können.

Es geht aber nicht nur um die Daten selbst. Vielmehr spielt auch die Qualität der beschreibenden Daten (siehe Kapitel 4 „Metadaten und Metadatenstandards“) und der Infrastrukturen (siehe Kapitel 5 „FAIR-Prinzipien und CARE-Prinzipien“), über welche die Daten zur Verfügung gestellt werden, eine wichtige Rolle.

---

## Wie setze ich das um?



Folgende Maßnahmen zur Verbesserung der Datenqualität können Sie bspw. in Ihren Forschungsprozess einbeziehen:

- Implementierung eines Qualitätssicherungskonzepts (insb. bei größeren Projekten)
- First-Time-Right-Prinzip: Prüfroutinen während der Erstellung
- Einheitlichkeit und Standardisierung bei der Datenerstellung (inkl. Dokumentation der Standards)
- Erstellung von Metadaten
- Dublettenprüfung
- Plausibilitätschecks
- Peer-Review durch Kolleg\*innen (vor allem bei Software und Skripten sollten diese auf anderen Rechnern ausgeführt werden, um zu prüfen, ob die Abhängigkeiten stimmen)