

Forschungsdaten aus wissenschaftstheoretischer Sicht

Eckhart Arnold (arnold@badw.de),

Bayerische Akademie der Wissenschaften



Überblick

1. Was sind Forschungsdaten?

2. Warum sollten Forschungsdaten veröffentlicht werden?

- a) Warum sollte man sie digital veröffentlichen?
- b) Warum sollte man sie als Open Access veröffentlichen?

3. Veröffentlichung von Forschungsdaten:

- a) Wie sollte man Forschungsdaten veröffentlichen?
- b) Wann sollte man Forschungsdaten veröffentlichen?
- c) Wann sollte man Forschungsdaten *nicht* veröffentlichen?

1. Was sind Forschungsdaten

Alternative, denkbare Definitionen von „Forschungsdaten“:

- Alle Daten, die in einem Forschungsprojekt *verwendet* werden
- Alle Daten, die in einem Forschungsprojekt anfallen (*erzeugt, gesammelt oder beiläufig entstanden*)
- Alle Daten, die *für die künftige Forschung relevant* sein könnten
- Eine Mischung der drei ersten Kategorien?

Mögliche Fragen:

- Gehören auch Notizzettel zu den Forschungsdaten?
- Gehören Informationen über den Forschungsablauf zu den Forschungsdaten?
(Könnten ja für Wissenschaftshistoriker noch interessant sein...)
- Gehören intermediäre Daten zu den Forschungsdaten?
d.h. Daten, die in der Form, in der sie vorliegen, weder gesammelt wurden, noch unverwandelt ins Forschungsergebnis einfließen

Mögliche Definitionsansätze:

- Keine bzw. Implizite Definition
- Aufzählende oder Merkmalsdefinition
- Empirische Definition
- ***Systematische Definition***

Systematische Definition:

- nach der Rolle im Forschungsprozess
(Entstehungszusammenhang)
- nach der **Bedeutung für die Verifikation der Forschungsergebnisse**
(Geltungszusammenhang)

Geltungszusammenhang wissenschaftlicher Erkenntnis:

1. Rationale Begründung:

- Empirische Prüfbarkeit
- Nachvollziehbare Interpretation
- Logisch folgerichtige Argumentation

2. Reproduzierbarkeit

3. Intersubjektive Kritisierbarkeit

Prinzip der Trennung von Entstehungszusammenhang und Geltungszusammenhang

Was sind Forschungsdaten?

*Forschungsdaten sind alle **für ein gegebenes Projekt** gesammelten oder erzeugten **forschungsrelevanten** Daten – einschließlich der Forschungsergebnisse, soweit sie datenförmig sind.*

Was sind keine Forschungsdaten?

Keine Forschungsdaten sind Daten, die überhaupt nicht oder nur für ein anderes Projekt oder Thema relevant wären.

Aber: Was heißt Forschungsrelevant?

Eine begründungstheoretische Definition:

Forschungsdaten sind alle diejenigen im Rahmen einer wissenschaftlichen Untersuchung gesammelten oder produzierten Daten, die dazu dienen, das Forschungsergebnis zu rechtfertigen.

Kurz: Forschungsdaten sind alle im Begründungszusammenhang der Forschung relevanten Daten.

Merke: Nicht alle Daten, die bei der Forschung anfallen sind Forschungsdaten!

Eine reproduktionsorientierte Definition:

*Forschungsdaten sind alle diejenigen im Rahmen einer wissenschaftlichen Untersuchung gesammelten oder produzierten Daten, **die notwendig sind, um auf demselben Weg dasselbe Ergebnis (re-)produzieren zu können.***

Wiederum: Nicht alle Daten, die bei der Forschung anfallen sind Forschungsdaten!

Etwas Terminologie:

Primärdaten → Intermediärdaten → Ergebnisdaten

z.B. Quellen → Exzerpte → Publikationen

- **Primärdaten** sind Daten, die ohne weitere Verarbeitungsschritte aus der Erhebung/Digitalisierung hervorgegangen sind.
- **Intermediärdaten** sind Daten, die durch die Verarbeitung von Primärdaten entstanden sind, aber noch weiterverarbeitet werden bevor sie in die Forschungspublikationen einfließen.
- **Ergebnisdaten** sind Daten, die nicht mehr weiterverarbeitet werden und in der Form publiziert werden oder in Publikationen einfließen.

Diskussion der Definition

Pro:

- Verständlich und praxistauglich
- Liefert zugleich ein klares Abgrenzungskriterium
- Wissenschaftstheoretisch einleuchtend

Contra:

- Nachnutzung in der Definition nicht berücksichtigt
- Daher möglicherweise zu restriktiv
- Status von Quell- / Archivdaten in der Definition nicht unmittelbar erfasst

2. Warum sollten Forschungsdaten veröffentlicht werden?

Warum sollten Forschungsdaten zugänglich gemacht werden?

- Wenn die Forschungsdaten nicht unzugänglich sind, lässt sich das Forschungsergebnis nicht prüfen!

Aber warum gleich veröffentlichen?

- Weil man so – ganz im Sinne der Wissenschaft – anderen Forscher/innen eine maximale Chance bietet, mit ihrer Kritik anzusetzen.
- Weil es die Schwelle für die Nachnutzung verringert.

Typische Vorbehalte gegen die Veröffentlichung von Forschungsprimär und -intermediärdaten:

- **Angst vor Kritik:** „Daten nicht in hinreichend präsentablen Zustand“
- **Unsicherheit:** „Uhrheberrechtsfragen könnten ungeklärt sein“
- **Kontrolle:** „Wir möchten gerne wissen, wer die Daten nutzt“
- **Angst vor Konkurrenz:** „Wir haben viel Zeit investiert, diese Daten zusammen zu tragen...“

Vorbehalte gegen die Digitalisierung von Forschungsdaten sind oft lediglich Vorbehalte gegen deren Veröffentlichung!

„Digitale Veröffentlichung“ bedeutet:

1. im Internet verfügbar
2. in menschenlesbarer Form *und*
3. **in maschinenlesbarer Form**
4. ggf. vernetzt

„Digitale Veröffentlichung“ verlangt *nicht*:

- dass auch Werkzeuge zur Nutzung bereit gestellt werden.
- dass stets alle oben aufgeführten Merkmale erfüllt sind.

Vorteile der digitalen Veröffentlichung von Forschungsdaten:

- ermöglicht die *Auswertung mit Algorithmen*
- ermöglicht die *Vernetzung*
- drastisch bessere *Verbreitung*

Veröffentlichung von Forschungsdaten als **Open Access**:

- erleichtert bzw. ermöglicht oft erst die digitale Nachnutzung
- Wie schon zuvor: Erleichtert erheblich die kritische Prüfung wissenschaftlicher Erkenntnis

Ohne Open Access lassen sich die Vorzüge der Digitalisierung kaum ausschöpfen!

3. Veröffentlichung von Forschungsdaten: Welche, wie, wann und wann nicht?

Welche Forschungsdaten sollten veröffentlicht werden?

Alle Primär-, Intermediär- und Ergebnisdaten, die zur Rechtfertigung der Forschungsergebnisse notwendig sind.

Veröffentlichung oder Zugänglichmachung?

*Veröffentlichung besser...
...aber auch eine Ressourcenfrage*

Wie sollten Forschungsdaten veröffentlicht werden?

Mindeststandard:

- Grundlegende Metadaten
- Permalink und Zitierhinweis
- Gebräuchliche oder dokumentierte Datenformate

Goldstandard:

- Standarddatenformate und Trivialdatenformate
- Dokumentation
- Ggf. tiefe Verknüpfungen

Inwiefern sollten (alternative) Nachnutzungsmöglichkeiten berücksichtigt werden?

- Vorteil: Könnte die Wirksamkeit (Impact) eines Forschungsprojekts erhöhen.
- Nachteil: Zusätzlicher Aufwand.

Faustregel:

- Aufbereitung für alternative Nachnutzungsmöglichkeiten fällt in die Verantwortung der daran Interessierten
- Aber Nachnutzungsmöglichkeiten sollten nicht verbaut werden.

Wann sollten Forschungsdaten offen gelegt werden?

- **Von Anfang an?** (sobald sie anfallen)
- Erst bei Abschluss eines Projekts oder von Meilensteinen?

Wann sollten Forschungsdaten nicht veröffentlicht werden?

Nur sofern besondere Gründe es verbieten, z.B.:

- Rechte Dritter
- Persönlichkeitsschutz
- Öffentliche Sicherheit
- Etc.

Schlussfolgerungen:

- Grundsätzlich: Verbreitung von Daten und Ergebnissen fällt in die Verantwortung der Forscher/innen
- Forschungsdatenmanagement schließt Forschungsdatenveröffentlichung ein.
- Welche Daten Forschungsdaten sind, kann von ihrer Rolle im Geltungszusammenhang abhängig gemacht werden.
- Digitale Veröffentlichung als Open Access sollte die Regel sein.