

Von offenen zu **FAIR**en Daten

Eckhart Arnold (arnold@badw.de), IT / Digital Humanities-Dept.



FAIR - Gute wissenschaftlicher Praxis bei der Bereitstellung von Daten:

- **Findable**
Permalinks; Metadaten; Verzeichnung in Suchsystemen
 - **Accessible**
Maschinelles Zugriff (Massendownload); Dauerhafte Metadaten-Bereitstellung
 - **Interoperable**
Standardunterstützung (so weit möglich); Verlinkung
 - **Reusable**
Urheber- und Lizenzinformationen; (in der Regel uneingeschränktes) Recht zur Weitergabe und Nachnutzung; Fachspezifische Anforderungen; *Einfachheit*
-

Fair geht über „offene Daten“ hinaus, indem es nicht nur die Nutzungsrechte, sondern auch technische Anforderungen der Datenbereitstellung regelt:

Rechte:

- CC-BY(-SA) – Lizenz:
 - Jede/r darf die Daten weiterbenutzen
 - Jede/r darf sie weitergebenAusnahmen möglich, aber nur begründet und nicht im Belieben des Urhebers

Technische Anforderungen:

- Maschinenlesbarkeit
 - Massenabruf
 - Verknüpfung (wo sinnvoll), Verzeichnung, Dokumentation und (wenn möglich) Standardkonformität, *Einfachheit*
-

Welche Empfehlungen kann man Forschungsprojekten geben, für die das Thema neu ist, und die möglichst rasch FAIR werden wollen?

FAIR – rasch und schmerzlos

- 1. *Permalinks***
- 2. *Massendownload***
- 3. *Creative-Commons-Lizenz (CC-BY oder CC-BY-SA)***

Alle drei Punkte lassen sich rasch und aufwandsarm umsetzen. Permalinks und Massendownload können mit Systemen wie gitlab realisiert werden.

Warum FAIR sein?

- *Daten können (anders als Texte) nur mit Hilfe technischer Systeme adäquat rezipiert werden.*
 - > nicht nur Bereitstellung zur Ansicht, sondern auch maschinenlesbare Form und Komplettabruf
 - *Daten können durch Aggregation, Kombination, Verknüpfung, Neu-Analyse wertvolle neue Erkenntnisse liefern.*
 - > möglichst geringe rechtliche Einschränkungen bei der Nachnutzung
 - *Der Nachvollzug (Reproduktion) und die Kritik datengestützter wissenschaftlicher Befunde ist nur bei freiem Zugriff auf die Daten möglich.*
 - > möglichst wenig rechtliche oder technische Barrieren bei der Nachnutzung
-

Wo ist FAIR inzwischen verankert?

- DFG-Leitlinien zur Sicherung guter wissenschaftlicher Praxis, <https://t1p.de/z8lt> S. 18ff.
- Die „Wissenschaftspolitische Stellungnahme zum Akademienprogramm“ des Wissenschaftsrats, <https://t1p.de/8932> S. 37ff.
- In den Empfehlungen von Fachverbänden: Beispiel: Schönbrodt, Gollwitzer, Abele-Brehm: Der Umgang mit Forschungs-daten im Fach Psychologie: Konkretisierung der DFG-Leitlinien, <https://t1p.de/2lev>

Aus Gründen der Nachvollziehbarkeit, Anschlussfähigkeit der Forschung und Nutzbarkeit hinterlegen Wissenschaftlerinnen und Wissenschaftler, wann immer möglich, die der Publikation zugrunde liegenden Forschungsdaten und zentralen Materialien – den FAIR-Prinzipien („Findable, Accessible, Interoperable, Re-Usable“) folgend – zugänglich in anerkannten Archiven und Repositorien.

DFG-Leitlinien zur Sicherung guter wissenschaftlicher Praxis,
<https://t1p.de/z8lt> S. 19. (Erläuterungen zu „Leitlinie 13: Herstellung von öffentlichem Zugang zu Forschungsergebnissen“)

Die Entwicklung rund um die NFDI und die EOSC bewertet die Union als große Chance zur Standardisierung und Zertifizierung ihrer Dienste und Forschungsdaten unter den Gesichtspunkten der FAIR-Prinzipien. (S. 40)

Mit sogenannten FAIR-Impulsen sollen mindestens zwei Akademien mit unterschiedlich weit ausgebauten DH-Infrastrukturen zu FAIR-Aspekten über wenige Monate mit Blick auf ein konkretes Ergebnis zusammenarbeiten. Als Ergebnis eines solchen Impulses sollen z. B. Forschungsdaten in Discovery-Systemen auf unterschiedlichen Ebenen – von der Union der Akademien über die NFDI bis zur EOSC (European Open Science Cloud) integriert werden. (S. 41)

Die „Wissenschaftspolitische Stellungnahme zum Akademienprogramm“ des Wissenschaftsrats, <https://t1p.de/8932> S. 37ff.

FAIR gilt in beide Richtungen:

Datenbereitstellende können ein Embargo für die Datennachnutzung definieren. Das bedeutet, dass die bisher noch nicht in Publikationen verwendeten Daten zwar – wie von der DFG gefordert – auf einem Repositorium gespeichert werden, allerdings für eine gewisse Karenzzeit nicht für Dritte zugänglich sind. (S. 27, unter 7.3 „Rechte der Datenbereitstellenden“)

Nachnutzende sollten daher Kontakt zu den Datenbereitstellenden aufnehmen, um eine möglichst valide Nutzung der Daten zu ermöglichen und Missverständnisse zu vermeiden. ... Wer Daten anderer für eigene Forschung nutzt, muss diese Daten adäquat zitieren. (S. 28. „Pflichten der Datennutzenden“)

Schönbrodt, Gollwitzer, Abele-Brehm: Der Umgang mit Forschungsdaten im Fach Psychologie: Konkretisierung der DFG-Leitlinien, <https://t1p.de/2lev>

Beispiel: Repertorium Academicum Germanicum (RAG): <https://rag-online.de>

„Das RAG ist ein Langzeitprojekt, das die biografischen, sozialen und kulturellen Daten der universitären Gelehrten des Alten Reiches bis 1550 in einer Datenbank erfasst und auswertet. ... Das RAG hat das Ziel, die Geschichte der kulturellen Reichweiten einer geistigen Führungs- und Impulsgruppe der Vormoderne zu entwickeln und mit rund 60'000 Personen und ihren 360'000 Lebens- und Karrierestationen einen umfassenden Einblick in die mittelalterlichen Ursprünge der modernen Wissensgesellschaft zu gewinnen.“

Auf rag-online.de kann man:

- in der Datenbank recherchieren
- sich vorgegebene „Szenarien“ in grafisch visualisierter Form anzeigen lassen

Was das RAG nicht zulässt:

- Herunterladen der kompletten Daten nicht möglich
- Einspeisen der Daten und Re-Analyse mit alternativen Systemen (z.B. „neo4j“ statt „nodegoat“) nicht vorgesehen.
- Keinerlei Information über die Nutzungsrechte
- In mindestens einem Fall hat sich der Projektleiter ausdrücklich die Nachnutzung der Daten durch einen missliebigen Fachkollegen verboten.

Folge: Eine wissenschaftliche Kritik der von Projektmitarbeiter/innen veröffentlichten und auf diese Daten gestützten Analysen ist nur sehr eingeschränkt möglich!

- **Was sind Forschungsdaten?**

- Alle Daten, die im Laufe eines Forschungsprozesses anfallen?
- Alle Daten, die zum Nachvollzug bzw. zur Rechtfertigung eines gelieferten Ergebnisses notwendig sind? (Wissenschaftstheoretisches Kriterium)
- Alle Daten, die für künftige wissenschaftliche Forschungen interessant sein könnten?

- **Sind Texte auch Forschungsdaten?**

Sollten die FAIR-Prinzipien auch für die Veröffentlichung von Texten, z.B. Editionen, Transkriptionen, Wörterbucheinträge gelten? Sollten sie auch für die Veröffentlichung von Fachartikeln gelten?

- **Wann sollten (welche) Forschungsdaten veröffentlicht werden?**

- sobald sie anfallen?
- wenn sie redaktionell bearbeitet worden sind?
- nachdem die „Datensammler/innen“ Gelegenheit hatten, eigene Analysen dazu veröffentlichen?
- nach Abschluss eines Forschungsprozesses?
- nach dem Tod aller beteiligten Wissenschaftler/innen?

- **Unter welcher Lizenz sollten Daten veröffentlicht werden?**

- **Zu welchem Zusatzaufwand führt die Berücksichtigung von FAIR-Prinzipien? Wenn überhaupt?**

- Ist das immer der Fall, manchmal, wann?
 - Wie kann damit umgehen? D.h. den Zusatzaufwand minimieren, vermeiden, Verzögerungen rechtfertigen?
-