

Forschungsdatenmanagement und Förderauflagen zur Archivierung von Forschungsdaten in den Sozialwissenschaften

Sebastian Netscher

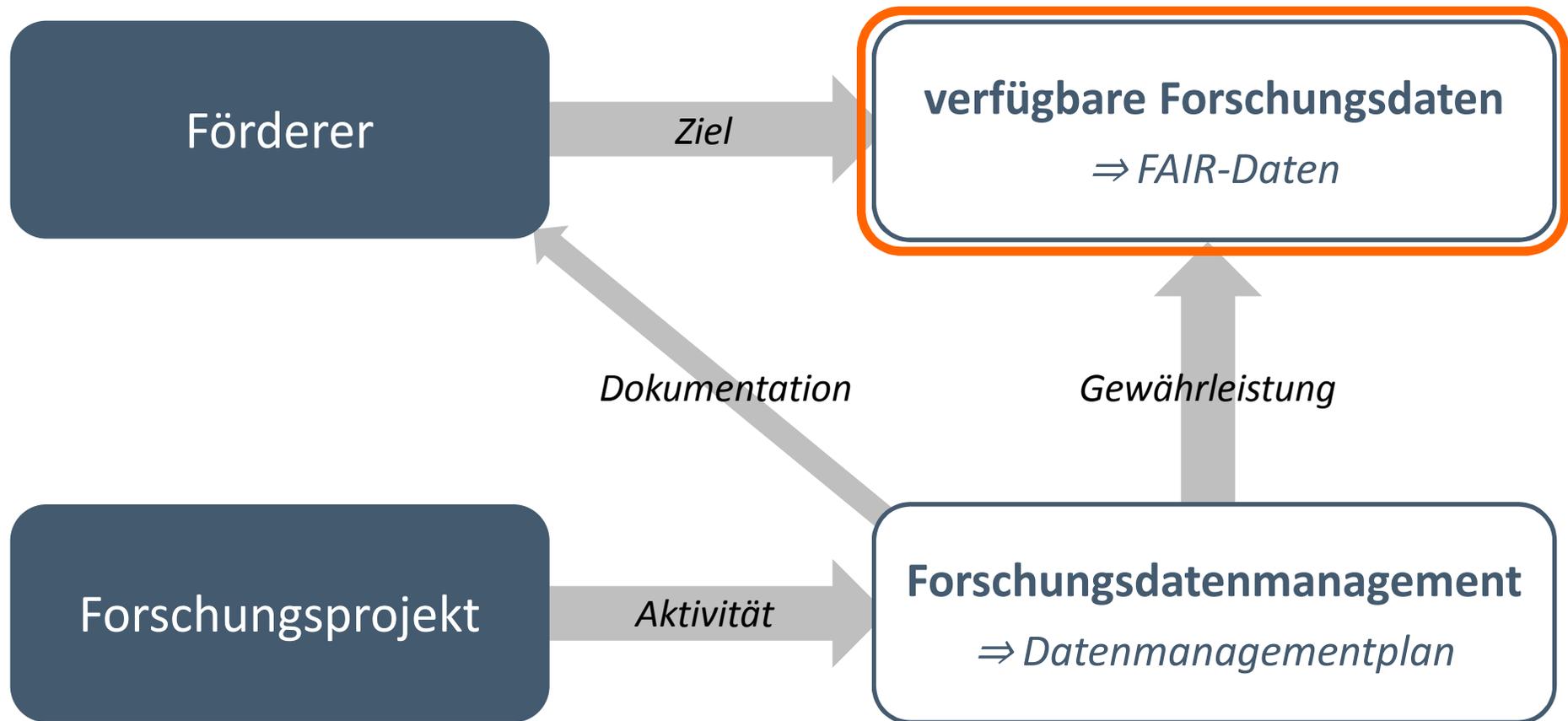
CESSDA Training am Datenarchiv für Sozialwissenschaften,
GESIS - Leibniz Institut für Sozialwissenschaften

Servicezentrum eSciences, Universität Trier
11 Juli 2017

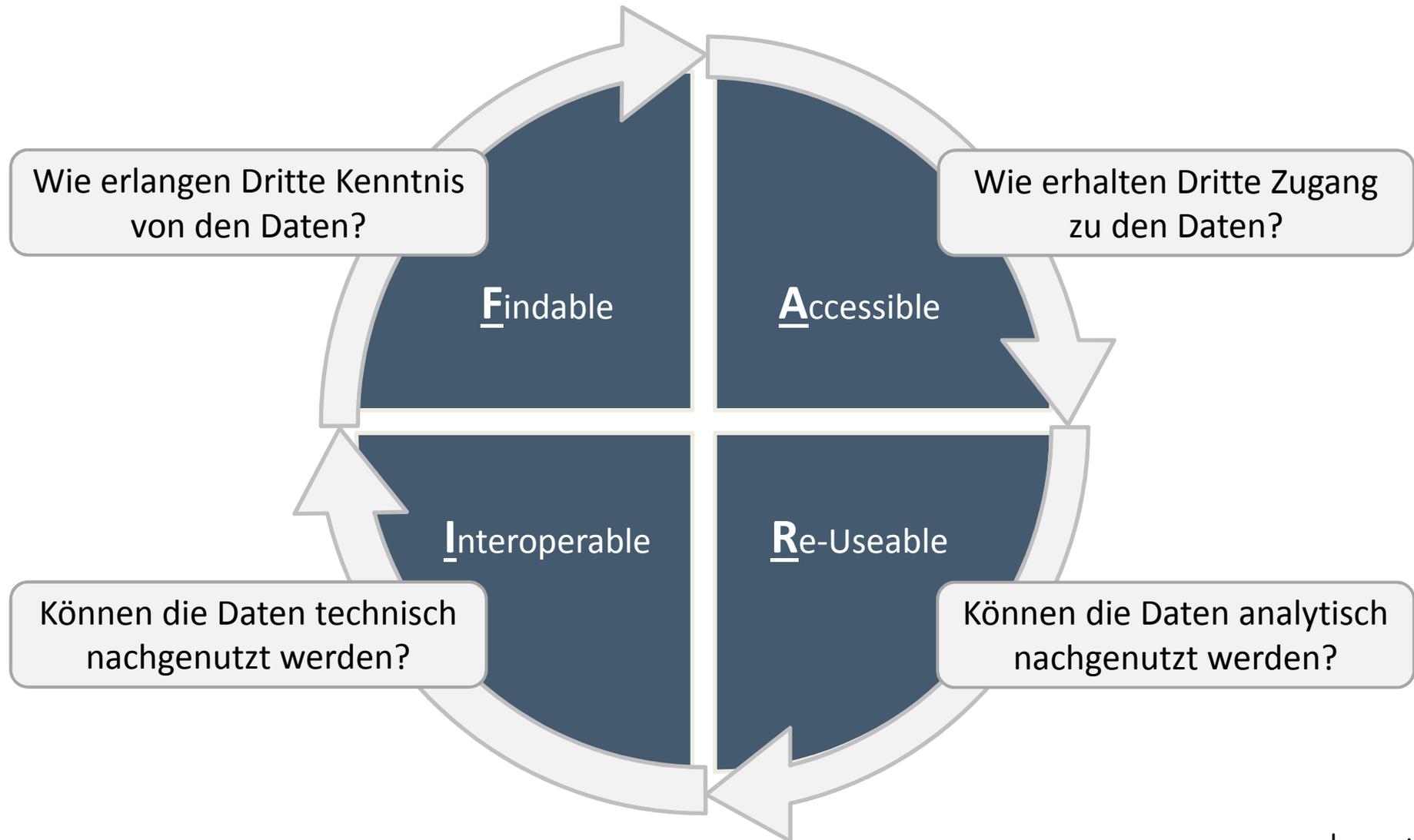
Forschungsdatenmanagement und Förderauflagen

- “Regarding the digital research data generated in the action (...), the beneficiaries must (...) deposit in a research data repository and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate (...) the data(...) laid down in the ‘data management plan’” (EU-Horizon 2020, 2016)
- „Die Antragstellenden verpflichten sich, die im Rahmen des Projekts gewonnenen Daten (...) zur Verfügung zu stellen (...). Um die Weitergabefähigkeit (...) zu gewährleisten, (...) ein eigenes Forschungsdatenmanagement betreiben.“ (BMBF, 2016)

Die Bedeutung der Förderauflagen



Die FAIR-Prinzipien von Forschungsdaten

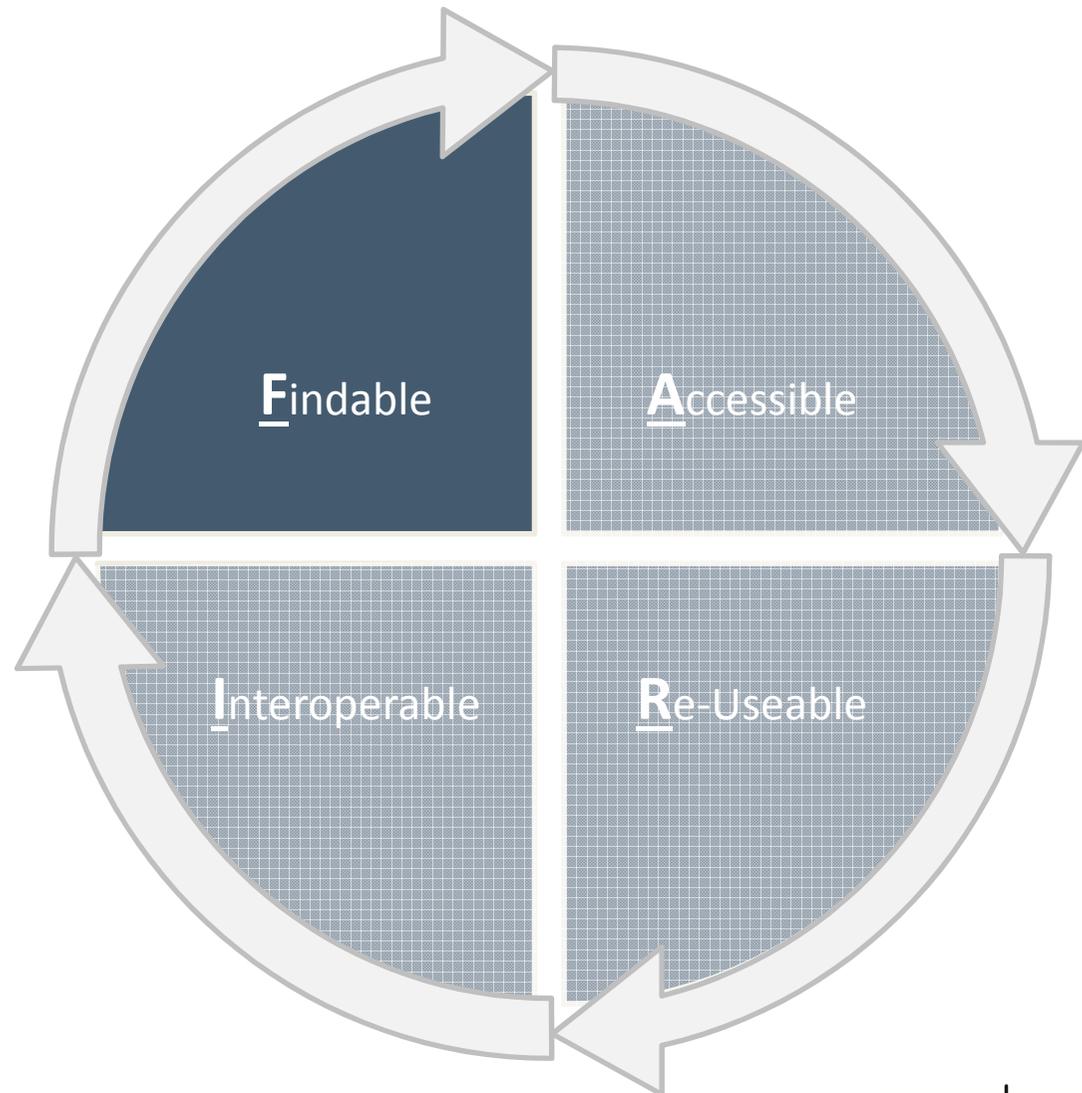


FAIR-Prinzipien: Auffindbarkeit

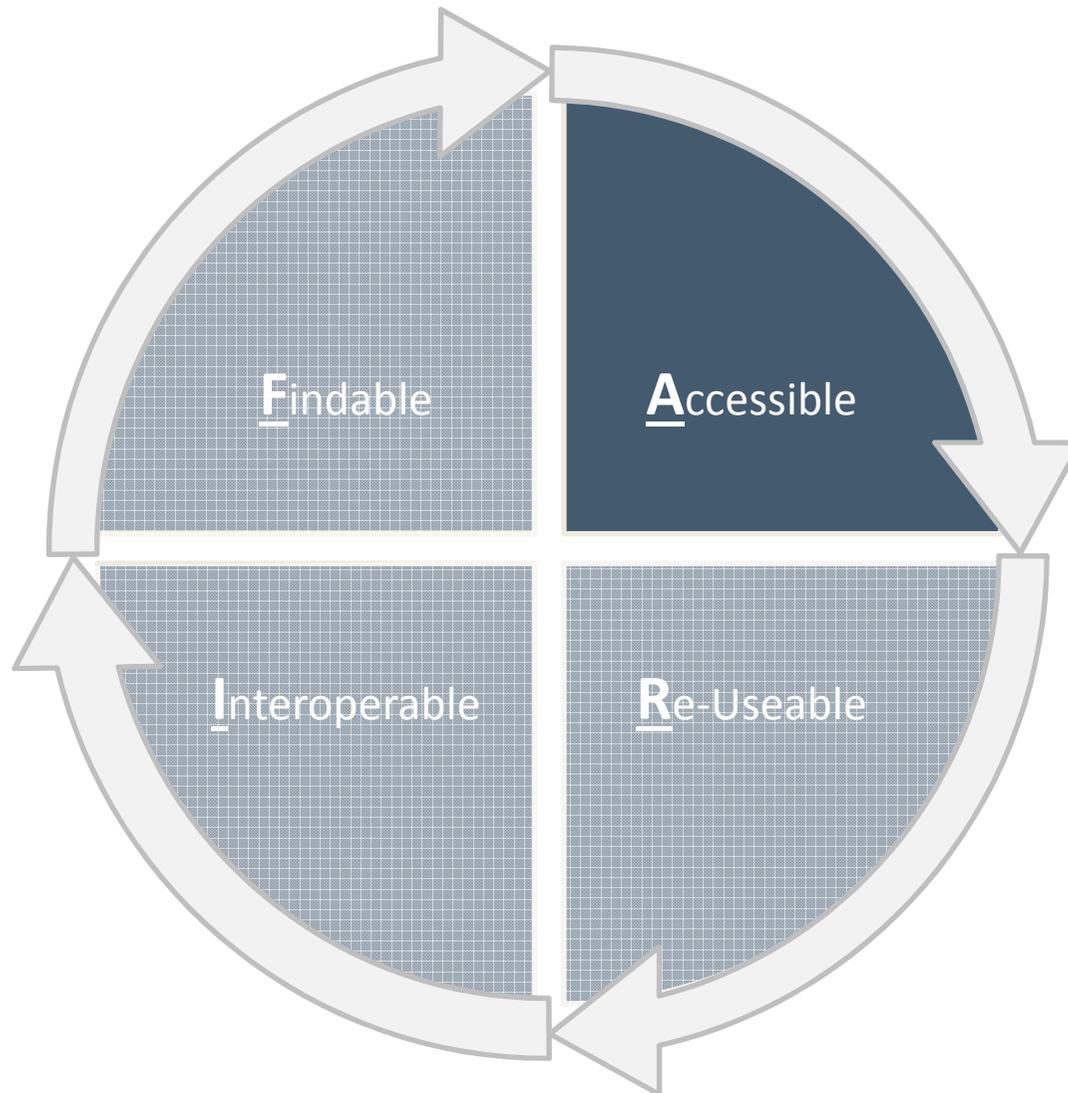
Auffindbarkeit

*Wie erlangen Dritte
Kenntnis von den Daten?*

- Zitation in Publikationen
 - Persistente Identifikatoren, z. B. DOIs
 - Datennachweis in Suchportalen
- ⇒ kontrolliertes Vokabular und Metadaten(-standards)



FAIR-Prinzipien: Zugänglichkeit



Zugänglichkeit

*Wie erhalten Dritte
Zugang zu den Daten?*

- Zugangs- und
Nachnutzungsbedingungen
 - projektinterne
Einschränkungen
 - rechtliche
Einschränkungen
- ⇒ so offen wie möglich, so
geschlossen wie nötig

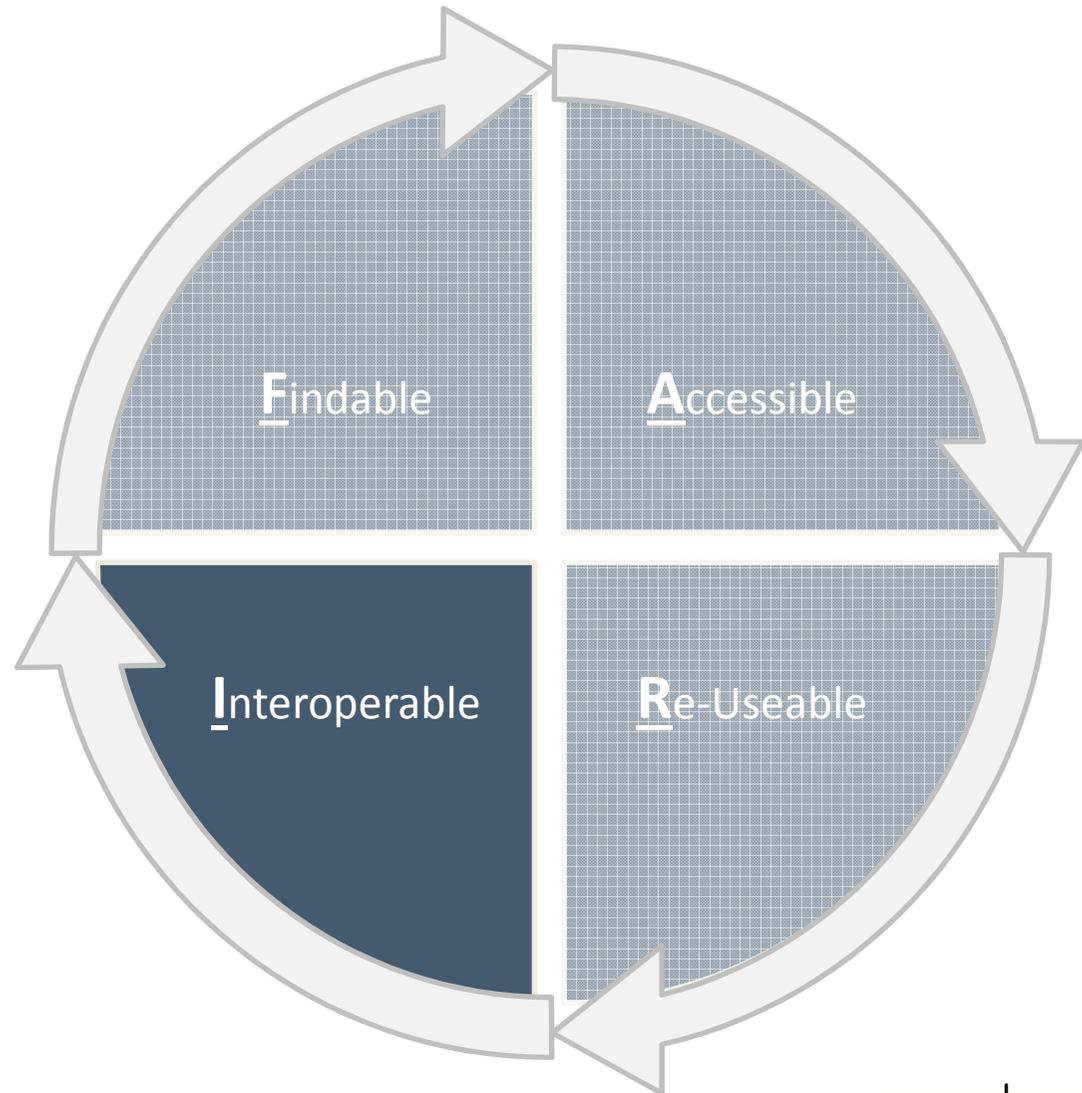
FAIR-Prinzipien: Interoperabilität

Interoperabilität

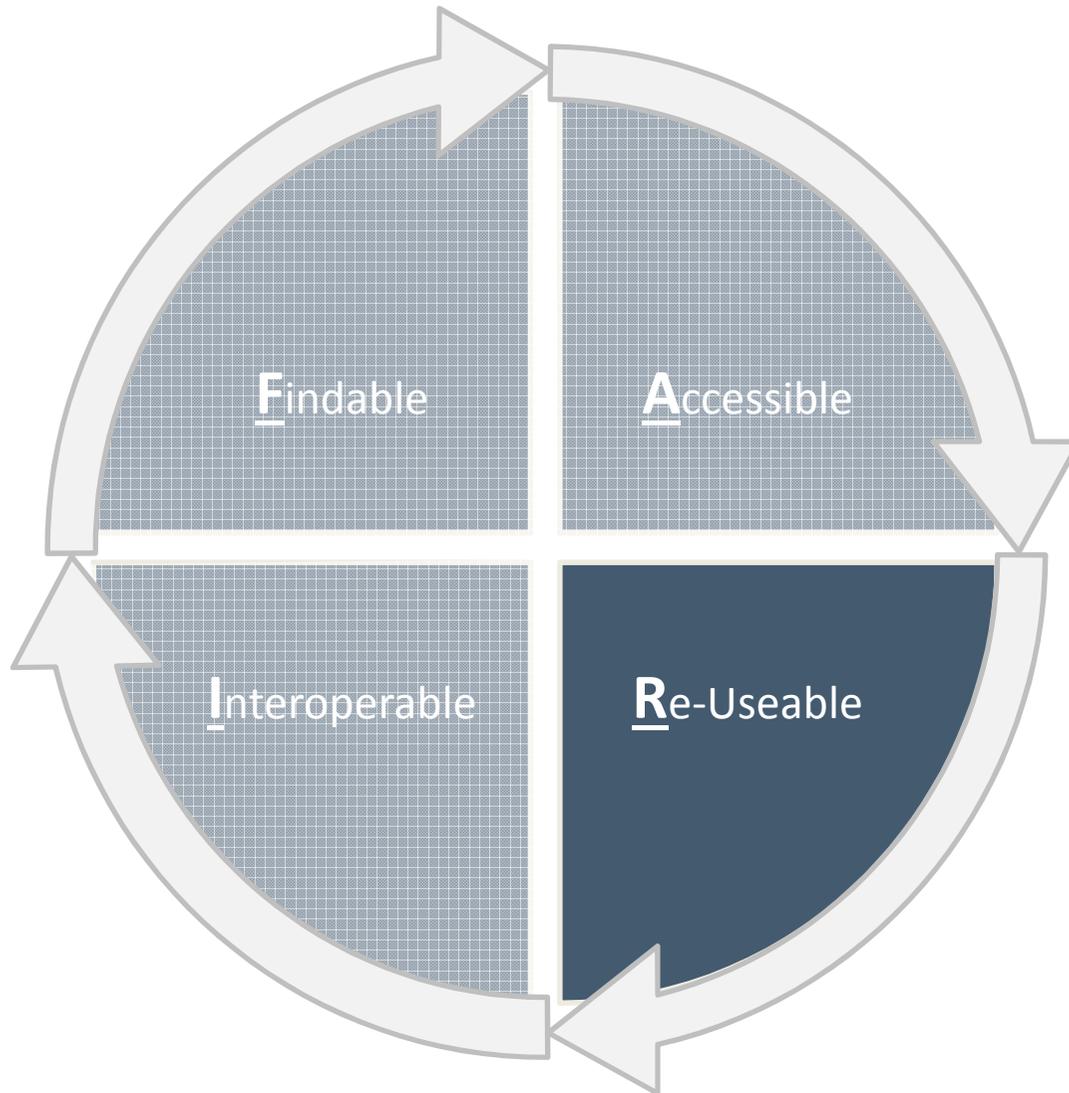
Können die Daten technisch nachgenutzt werden?

- Aufbau und Struktur des Datensatzes
- Dateiformate, Software und Versionen
- Begleitmaterialien (Dokumentation und Software)

⇒ disziplinspezifische Standards



FAIR-Prinzipien: Nachnutzbarkeit

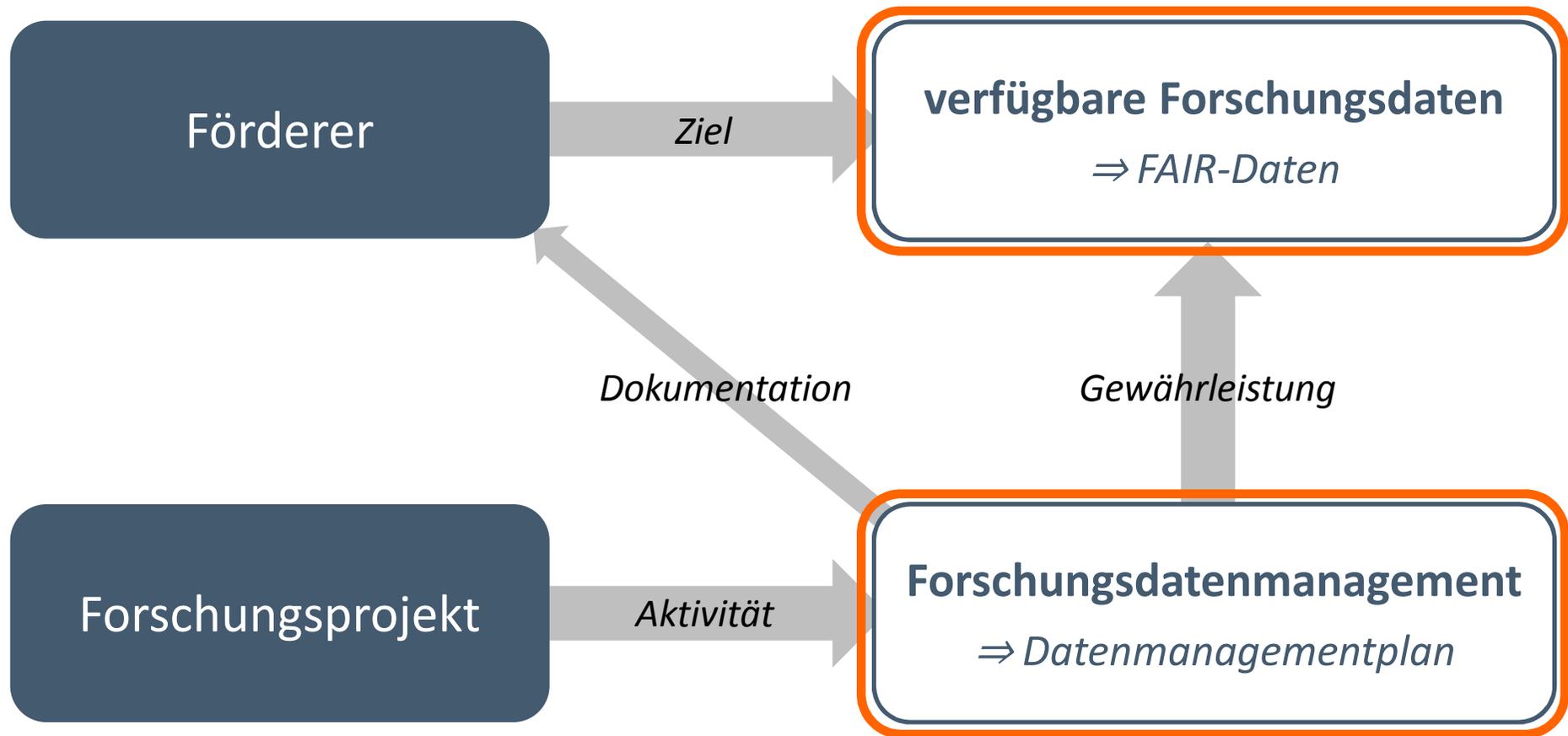


Nachnutzbarkeit

Können die Daten analytisch nachgenutzt werden?

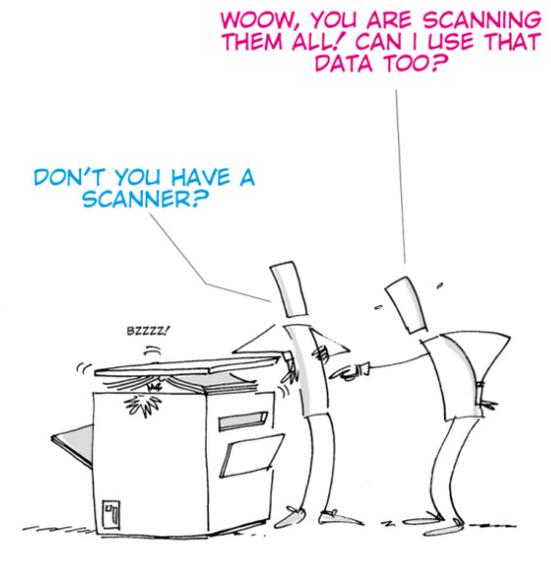
- Dokumentation der Daten
 - Inhaltliche Verständlichkeit der Daten
 - Interpretierbarkeit von Ergebnissen
- ⇒ kontrolliertes Vokabular und Metadaten(-standards)

Die Bedeutung der Förderauflagen



Forschungsdatenmanagement (FDM)

- FDM umfasst alle Prozesse und Maßnahmen zum Erhalt und zur (Nach-)Nutzbarkeit von Forschungsdaten
 - Qualitätssicherung im Forschungsprojekt
 - Replizierbarkeit über das Forschungsprojekt hinaus
 - Nachnutzbarkeit der Forschungsdaten durch Dritte



Bereiche des FDM

1. Handhabung der Forschungsdaten

⇒ Organisation, Dateiaufbereitung
und -dokumentation

2. Forschungsethische und rechtliche Aspekte

⇒ Datenschutz (personenbezogene
Daten) und Urheberrechte (Dritter)

3. Längerfristige Sicherung der Forschungsdaten

⇒ Replikation und Nachnutzung



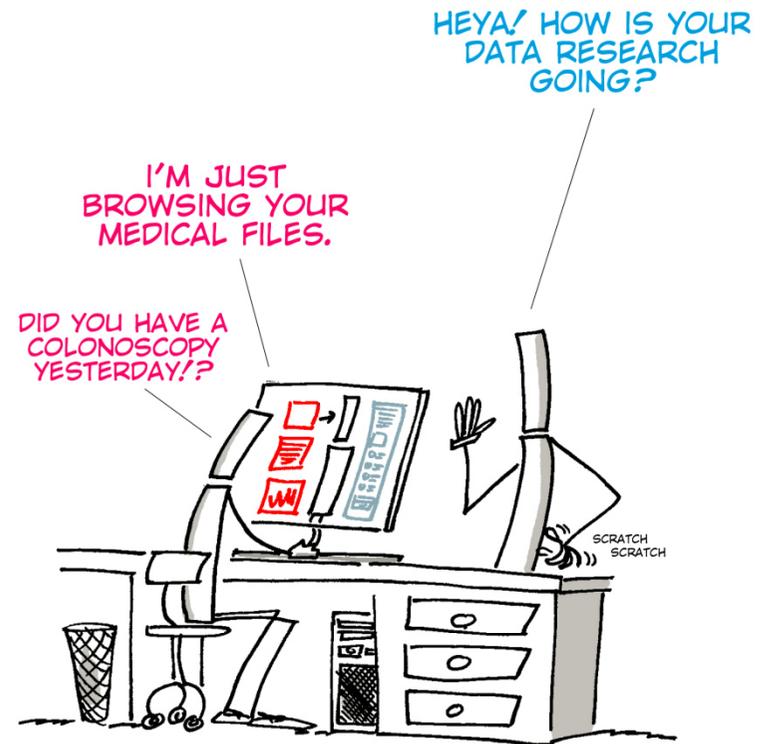
1. Handhabung von Forschungsdaten

- Datenorganisation
 - Ordnerstruktur, Benennung und Versionierung der Dat(ei)en
 - Sicherung der Dat(ei)en (Backups)
- Datenaufbereitung
 - Validierung der Daten
 - Prüfung auf fehlende Antworten, inkonsistentes Antwortverhalten etc.
 - Anonymisierung
- Datendokumentation
 - Studiendokumentation
 - Dokumentation der Forschungsdaten



2. Ethische und Rechtliche Aspekte

- Datenschutz (personenbezogene Informationen)
 - informierte Einwilligung
 - Anonymisierung
 - Schutz vor unautorisiertem Zugriff
- Urheberrecht (Recht am geistigen Eigentum)
 - an den Forschungsdaten
 - Urheberrechte Dritter

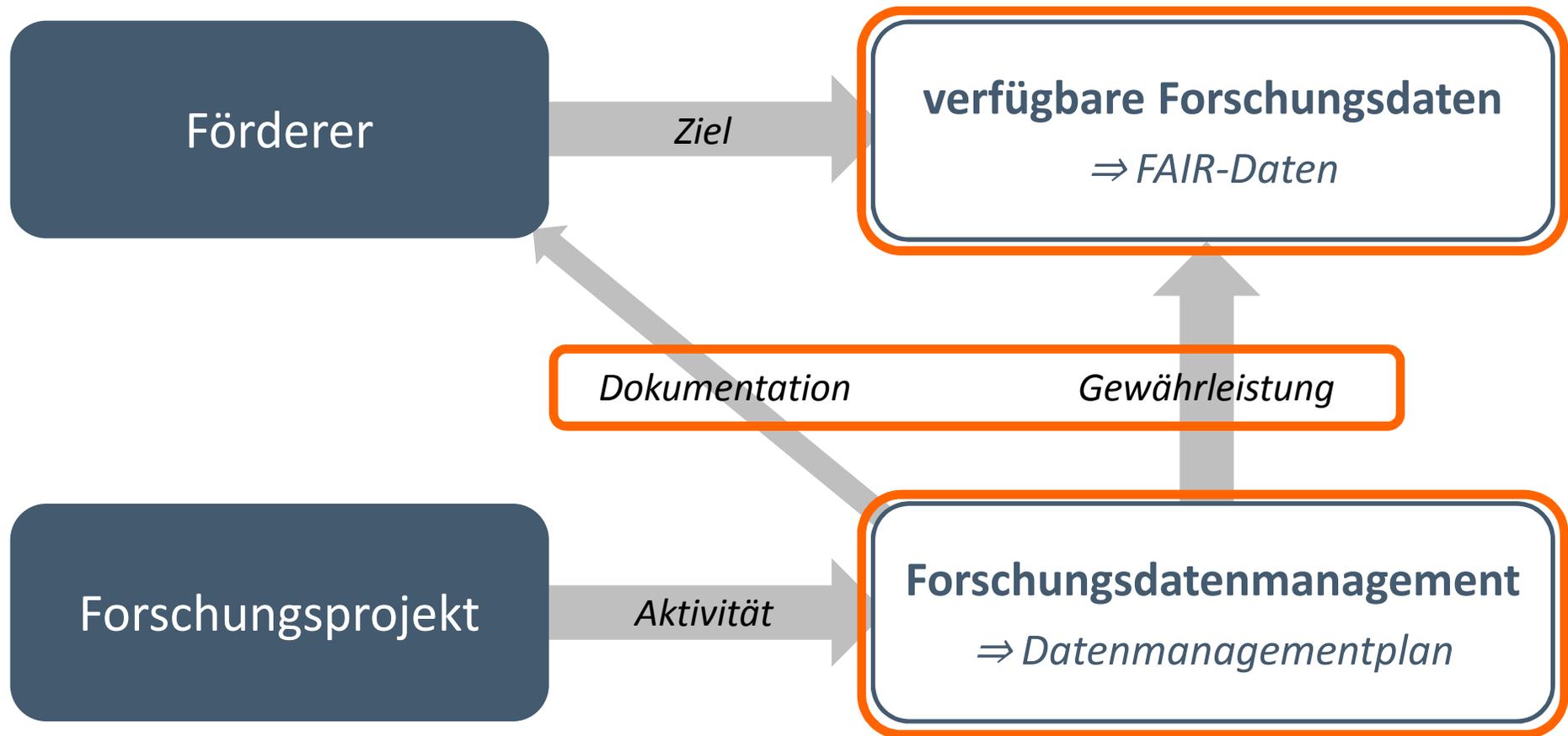


Der Datenmanagementplan (DMP)

- Zweckorientierte Dokumentation des Forschungsdatenmanagements
 - immer projekt-spezifisch
 - lebendes Dokument
 - zweckorientiertes Dokument
 - projekt-internen Qualitätssicherung
 - Replikation der Forschungsdaten
 - Verfügbarmachung der Forschungsdaten für Dritte



Die Bedeutung der Förderauflagen



Strategie zur Generierung verfügbarer Daten

1. Planung des FDM

- Forschungsdaten und Verfügbarkeit
- Einschränkungen der Verfügbarkeit
- Archivierung und Verfügbarkeit

2. Erstellung eines DMP

- vom erste Entwurf zum Abschlussbericht
- Aufbau des DMP



Forschungsprojekt

Werden im Forschungsprojekt **Forschungsdaten generiert**?

nein

keine Daten,
d. h. kein FDM & kein DMP

ja

Welche Forschungsdaten mit welchem Nachnutzungspotential?
Ist die **Verfügbarmachung** der Forschungsdaten **möglich**?

„opt out“

nein

Ist die **interne Sicherung** möglich?

nein

„opt out“

Wie werden die
Daten **intern
gesichert**?

Replikation

ja

CC-0 oder
CC-by?

Verfügbarmachung

ja

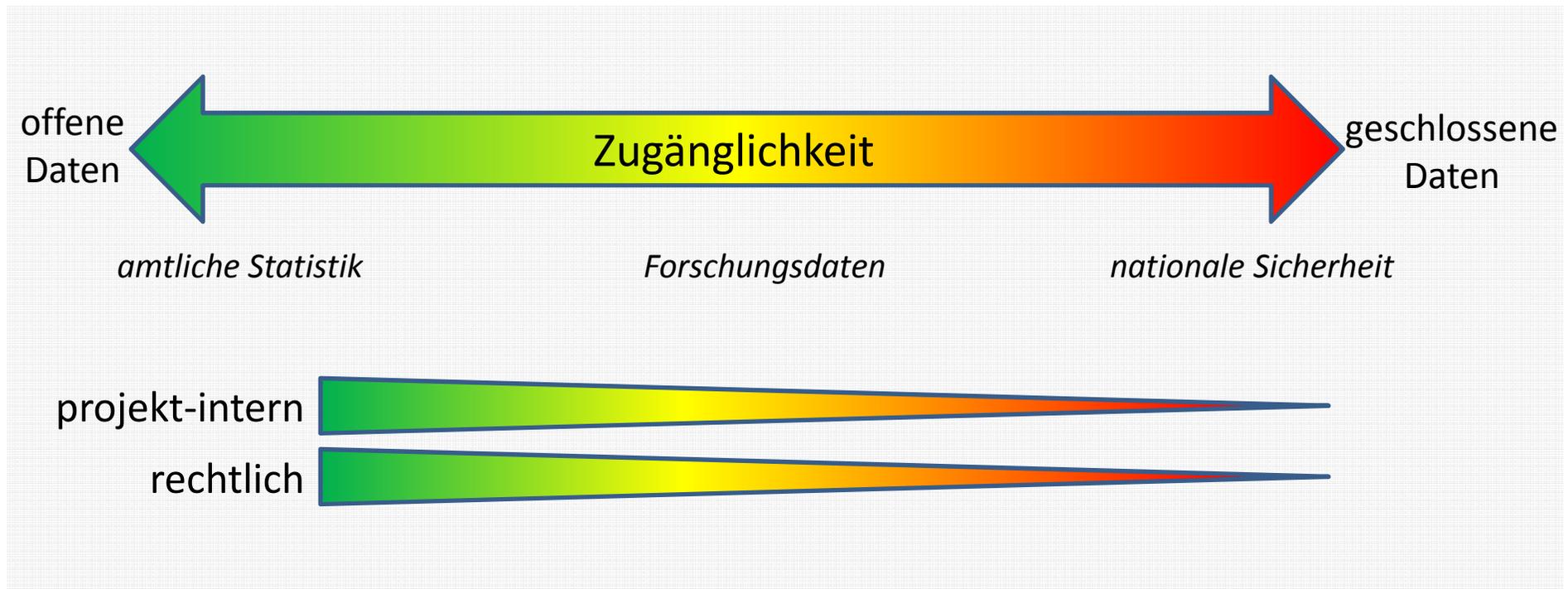
Ist ein **freier Zugang** für jeden **möglich**?

ja

Welche
Restriktionen
bestehen **warum**?

nein

Zugänglichkeit und Restriktionen



⇒ so offen wie möglich, so geschlossen wie nötig

Forschungsprojekt

Werden im Forschungsprojekt **Forschungsdaten generiert**?

nein

keine Daten,
d. h. kein FDM & kein DMP

„opt out“

ja

Welche Forschungsdaten mit welchem Nachnutzungspotential?
Ist die **Verfügbarkeit** der Forschungsdaten **möglich**?

nein

Ist die **interne Sicherung** möglich?

nein

„opt out“

Wie werden die
Daten **intern
gesichert**?

Replikation

ja

ja

Ist ein **freier Zugang** für jeden **möglich**?

ja

CC-0 oder
CC-by?

Verfügbarkeit

Welche Daten sollen
wie bereitgestellt werden?

Projekt

Repository

Auffindbarkeit und **Zugänglichkeit**

Forschungsdaten-Repositoryen

- Machen Forschungsdaten
 - auffindbar, z. B. über persistente Identifikatoren, Publikation von Metadaten in Suchmaschinen etc.
 - zugänglich, z. B. über Lizenzen (Definition von Zugangs- und Nachnutzungsbedingungen)
- Unterstützen bei der Generierung interoperabler und nachnutzbarer Forschungsdaten
- Ein geeignetes Repository finden
 - ⇒ Disziplin, Region, Datentyp, Dauer der Archivierung etc.

Erstellen eines DMP

- Erster Entwurf: Strategie zur Generierung verfügbarer Forschungsdaten
 - EU: innerhalb der ersten 6 Monaten
 - BMBF: Teil des Förderantrags
- Aktualisierter DMP in den Zwischenberichten: Sicherstellung der Umsetzung der Strategie
- Abschlussbericht: Erklärung, wann die Daten archiviert werden bzw. wurden



Aufbau des DMP

1. Beschreibung der Forschungsdaten:
⇒ Projektkontext, Art der Daten, Datengenerierung, Verfügbarkeit der generierten Daten etc.
2. Einschränkungen in der Verfügbarkeit der Daten
⇒ projektinterner und/oder rechtlicher Aspekte
3. Verfügbarkeit und Gewährleistung des FAIR-Prinzips
⇒ (Dauer der) Auffindbarkeit und Zugänglichkeit, Interoperabilität und Nachnutzbarkeit
4. Verantwortlichkeiten und Ressourcen für das FDM

Literatur

- Büttner, S., H.C. Hobohm & L. Müller (2011): Handbuch Forschungsdatenmanagement. Bad Honnef: Bock + Herchen.
- Bundesanzeiger (2016): Bekanntmachung des BMBF. 29.03.2016.
- CESSDA Training (2015): CESSDA User Guide on Data Management.
- Corti, L., V. van den Eynden, L. Bishop & M. Woollard (2014): Managing and Sharing Research Data. A Guide to Good Practice. London: Sage Publication Ltd.
- Force 11 (2016): The FAIR Data Principles.
- Horizon 2020 Programme (2016): Annotated Model Grant Agreement. Version 2.1.1. 1. July 2016.
- Horizon 2020 Programme (2016): Guidelines on FAIR Data Management in Horizon 2020. Version 3.0. 26 July 2016.
- Horizon 2020 Programme (2016): Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020. Version 3.1. 25 August 2016.
- Humboldt Universität zu Berlin (2017): Muster Datenmanagementplan für Horizon 2020. Version 3.0.
- Jensen, U. (2012): Leitlinien zum Management von Forschungsdaten. Sozialwissenschaftliche Umfragedaten. GESIS-Technical Reports 2012|07.