

# Das zweite Leben der Forschungsdaten.

Die Gewährleistung von Förderauflagen zu  
Open Access am Beispiel von Horizon 2020.

Sebastian Netscher

GESIS - Leibniz Institute for the Social Sciences

E-Science-Tage 2017: Forschungsdaten managen

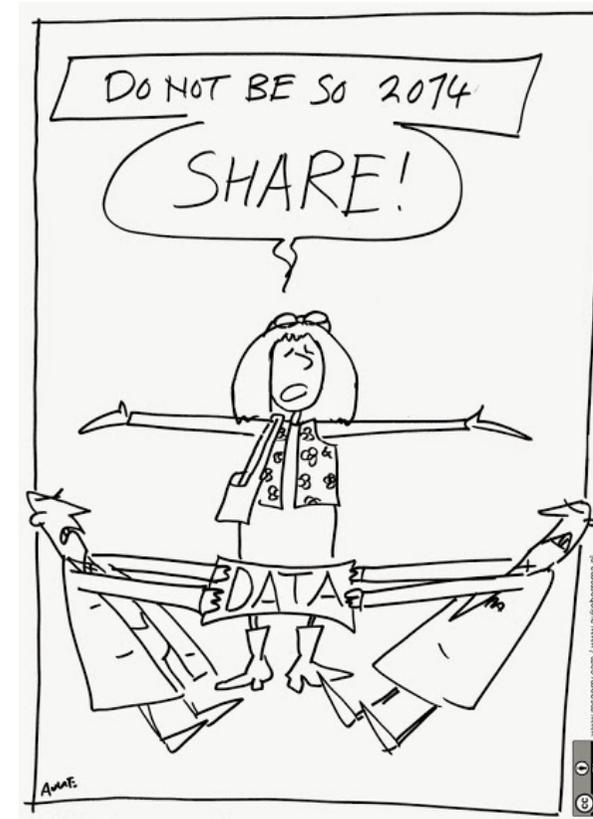
17 April 2017, Heidelberg



# Der Lebenszyklus von Forschungsdaten

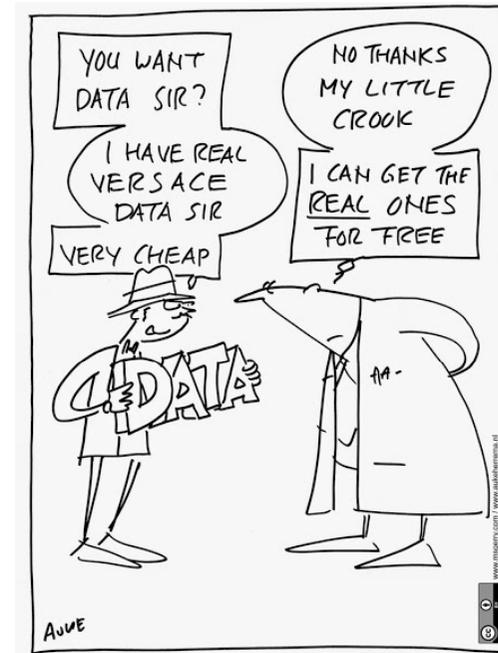


# Open Access und die Verfügbarkeit von Forschungsdaten



# Die Verfügbarkeit von Forschungsdaten

- Reproduktion der Forschungsergebnisse und Replikation der Forschungsdaten
- Nachnutzung der Forschungsdaten in neuen Forschungskontexten
  - ⇒ Open Access: Möglichkeit Dritter  
“to access, mine, exploit, reproduce and disseminate (...) the data”



# Open Access, warum?

- Förderung von Wissenschaft und Forschung
  - ⇒ Ausschöpfen des Potenzials von Forschungsdaten
- Gute wissenschaftliche Praxis
  - ⇒ Transparenz im Forschungsprojekt
- Steigerung der eigenen Reputation
  - ⇒ Zitation der Datenproduzenten
- Effiziente Nutzung öffentlicher Gelder
  - ⇒ Förderauflagen von Drittmittelgebern



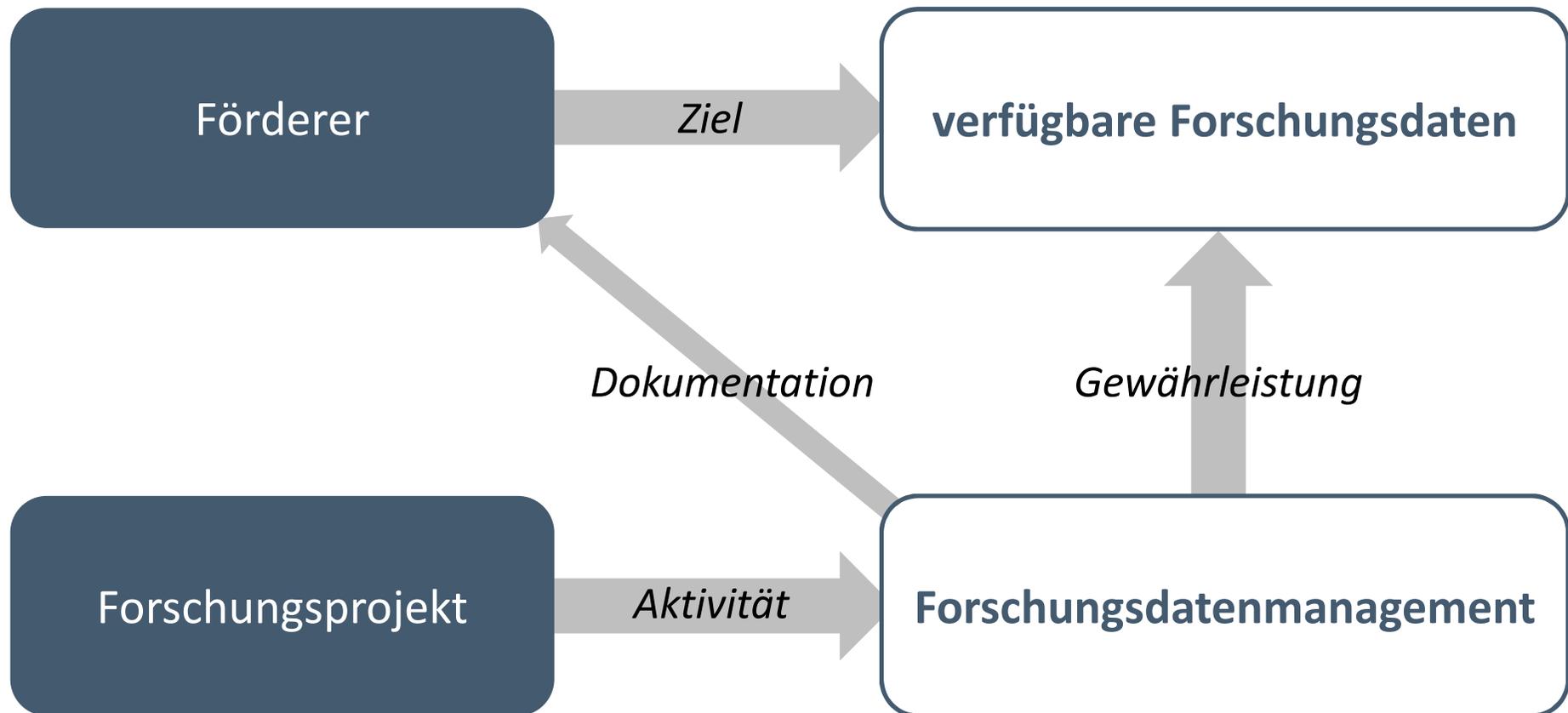
# Förderauflagen zu Open Access

- “Regarding the (...) research data generated (...), the beneficiaries must (...) deposit in a research data repository (...) draw up a ‘data management plan.’” (EU-Horizon 2020, 2016)
- „Die Antragstellenden verpflichten sich, die (...) gewonnenen (...) zur Verfügung zu stellen (...) ein eigenes Forschungsdatenmanagement betreiben“ (BMBF, 2016)



Bundesministerium  
für Bildung  
und Forschung

# Was die Förderaufgabe bedeutet



# Die Prinzipien FAIRer Forschungsdaten

## Findable:

Auffindbarkeit der Forschungsdaten  
⇒ kontrolliertes Vokabular, Metadaten  
(-standards), persistente Identifikatoren etc.



## Accessible:

Zugänglichkeit der Forschungsdaten  
⇒ Zugangs- und Nachnutzungsbedingungen



## Interoperable:

technische Nachnutzbarkeit  
⇒ Dateiformate, genutzte Software etc.

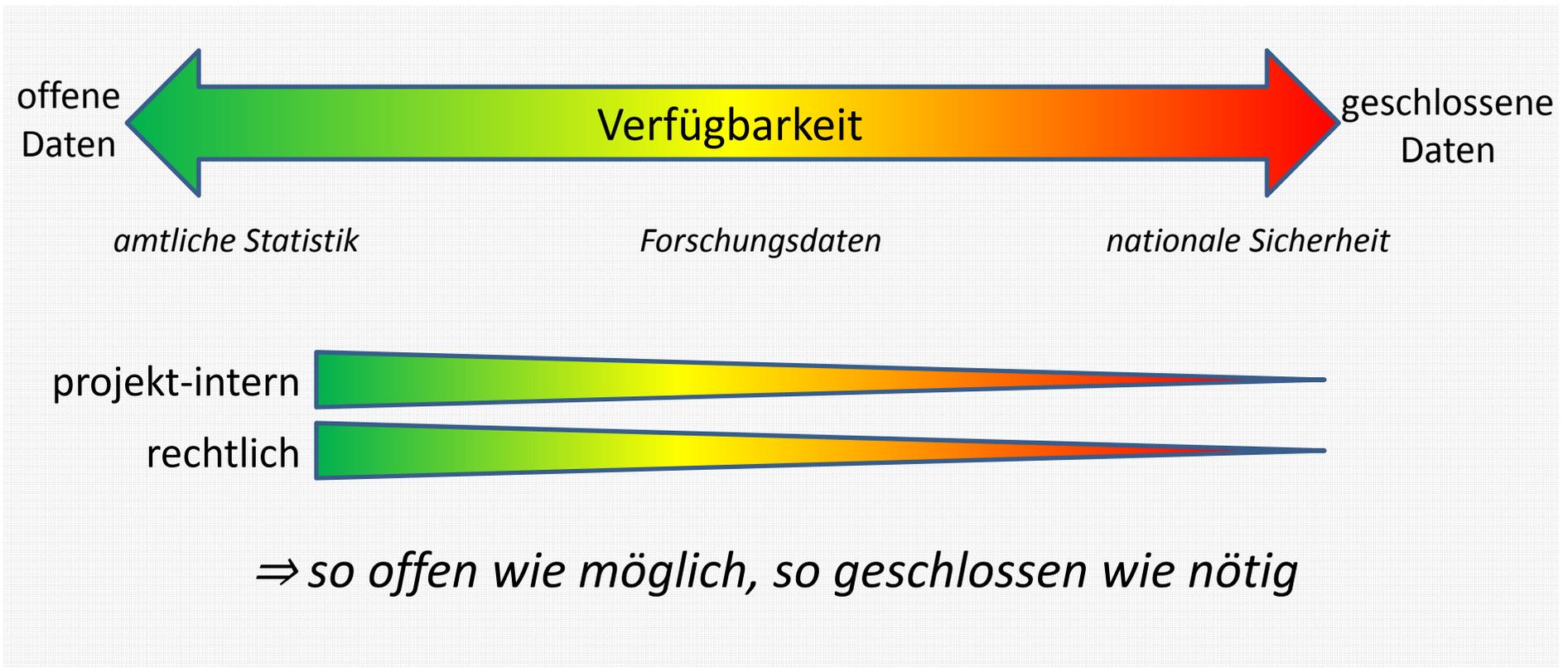


## Re-useable:

analytische Nachnutzbarkeit  
⇒ Verständlichkeit und Interpretierbarkeit



# “Open” ist nicht unbedingt “frei”



# Forschungsdaten- management und verfügbare Forschungsdaten



# Forschungsdatenmanagement (FDM)

- FDM umfasst alle Prozesse und Maßnahmen zum Erhalt und zur (Nach-)Nutzbarkeit von Forschungsdaten
  - orientiert sich am Lebenszyklus von Forschungsdaten
  - verfolgt unterschiedliche Ziele
    - Qualitätssicherung im Forschungsprojekt
    - Replizierbarkeit über das Forschungsprojekt hinaus
    - **Nachnutzbarkeit der Forschungsdaten durch Dritte**



# Bereiche des FDM

## 1. Handhabung der Forschungsdaten



⇒ Dateigenerierung, -dokumentation  
und -organisation

## 2. forschungsethische und rechtliche Aspekte



⇒ Datenschutz (personenbezogene Daten)  
und Urheberrechte (Dritter)

## 3. Längerfristige Sicherung der Forschungsdaten



⇒ Replikation und Nachnutzung

# 1. Handhabung von Forschungsdaten



- Generierung der Forschungsdaten
  - Sammlung und Aufbereitung der Rohdaten
  - Bereinigung der Forschungsdaten
- Dokumentation
  - Datengenerierung
  - Forschungsdaten
- Datenadministration
  - Organisation von Daten und Dateien
  - Sicherung der Daten (Backups)

**accessible**



**interoperable**



**re-useable**



**findable**



## 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



## 3. Längerfristige Sicherung



- Replikation der Forschungsdaten
  - ⇒ längerfristige (interne) Sicherung der Forschungsdaten
  - ⇒ Verfügbarkeit zu Replikationszwecken
- Verfügbarkeit der Forschungsdaten
  - ⇒ Bereitstellung der Forschungsdaten

**findable**



**accessible**



**interoperable**



**re-useable**



# Der Datenmanagementplan (DMP)

- zweckorientierte Dokumentation des FDM
  - Dokumentation
    - immer projekt-spezifisch
    - lebendes Dokument
      - „needs to be updated over the course of the project“
  - zweckorientiert
    - Entwicklung und Beschreibung einer Strategie zur
      - projekt-internen Qualitätssicherung
      - Replikation der Forschungsdaten
      - Verfügbarmachung der Forschungsdaten für Dritte

# Eine Strategie zur Generierung verfügbarer Forschungsdaten



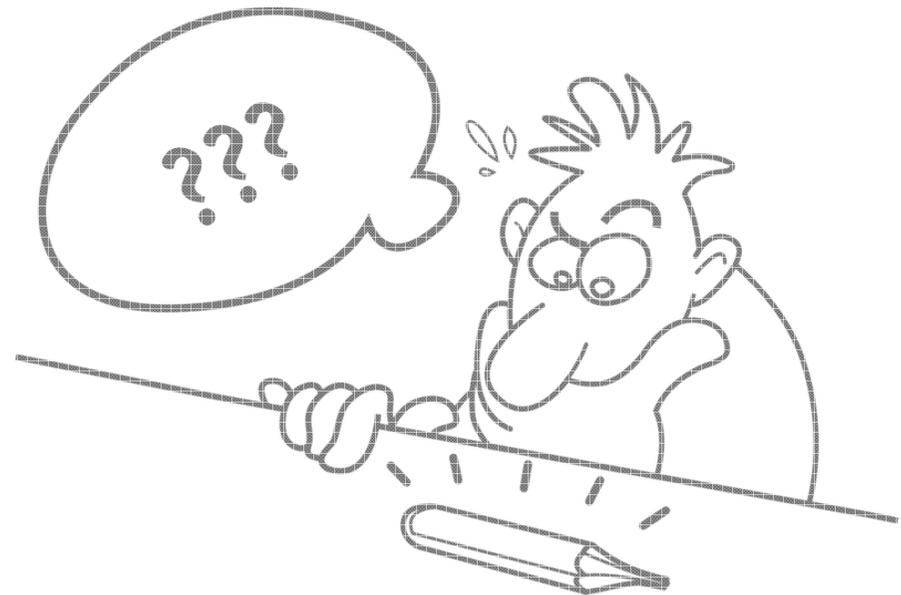
# Entwicklung einer Strategie

## 1. Planung des FDM

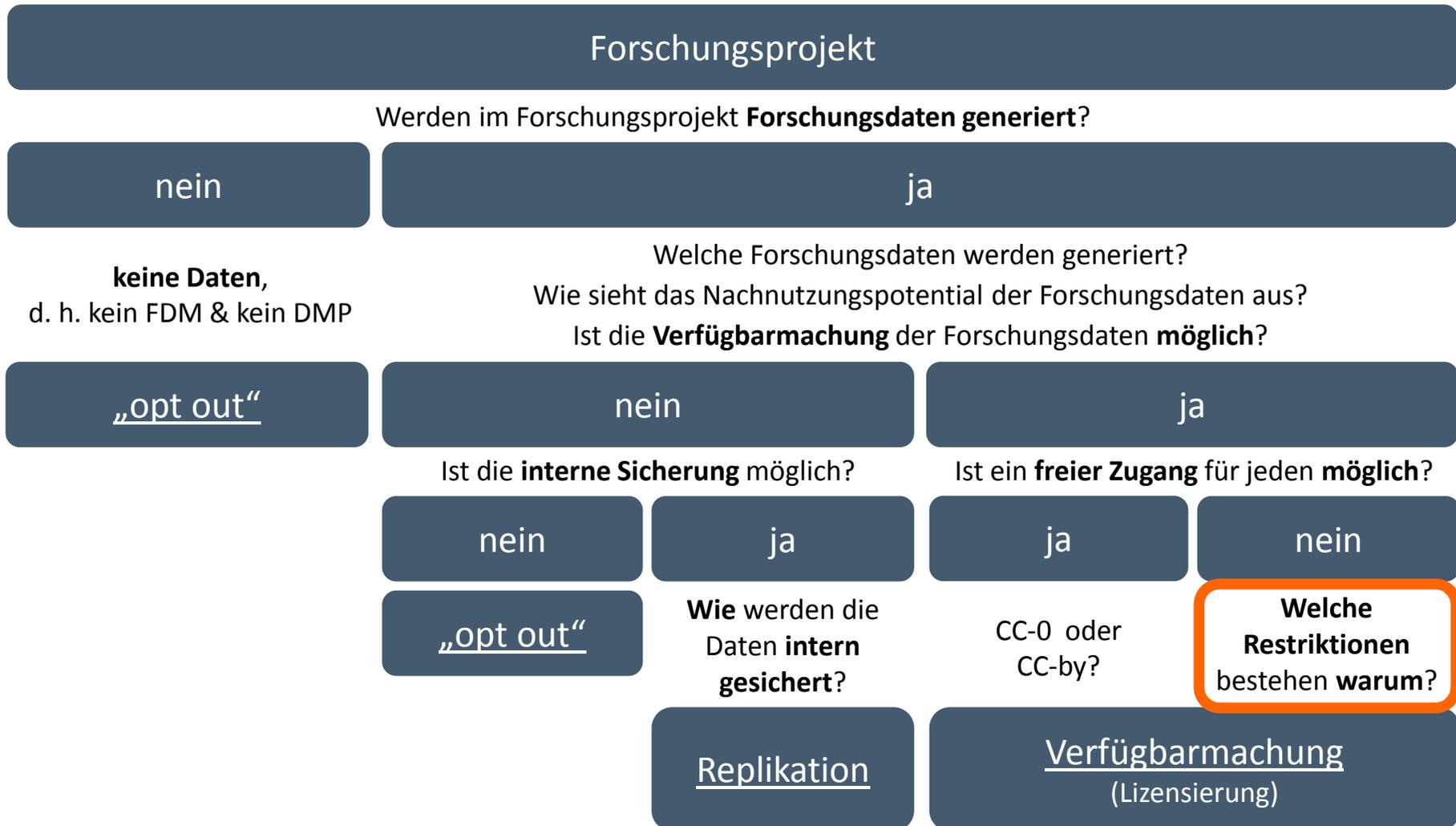
- Forschungsdaten & Verfügbarkeit
- Einschränkungen der Verfügbarkeit
- Archivierung & Verfügbarkeit

## 2. Erstellung eines DMP

- vom erste Entwurf zum Abschlussbericht
- Aufbau des DMP



# 1. Planung: Forschungsdaten & Verfügbarkeit



# 1. Planung: Einschränkungen der Verfügbarkeit

- Gründe für Einschränkungen
  - projektinterne Einschränkungen, z. B. Erreichen der Projektziele, Abschluss von Forschungsarbeiten und Qualifikationen
  - rechtliche Einschränkungen, z. B. aufgrund des Datenschutzes, des Urheberrechts oder sonstiger rechtliche Vorgaben
- Handhabung der Einschränkungen
  - (zeitliches) Embargo
  - Zugangs- und Nachnutzungsbedingungen (Lizenz)



# 1. Planung: Archivierung & Verfügbarkeit

Langfristige Sicherung und Bereitstellung der Forschungsdaten

Welche Forschungsdaten sollen **langfristig gesichert** bzw. **bereitgestellt** werden?

Wie sollen die Daten **langfristig gesichert** bzw. **bereitgestellt** werden?

durch das Forschungsprojekt

Wie und **wie lange** soll die Daten **verfügbar** sein?

Wie werden die Daten **auffindbar** und **zugänglich**?

durch ein Archiv / Repositoryum

Archive machen Daten  
**auffindbar** und **zugänglich**

Archiv / Repositoryum

Disziplin, Region, Datentyp, Dauer der Archivierung etc.

re3data.org  
REGISTRY OF RESEARCH DATA REPOSITORIES

OpenAIRE

# 1. Planung: Archivierung der Forschungsdaten

- *“Open access (...) via a research data repository (...) an online archive for research data”*

- Archive machen Forschungsdaten



– auffindbar, z. B. über persistente Identifikatoren, Publikation von Metadaten in Suchmaschinen etc.



– zugänglich, z. B. über Lizenzen (Definition von Zugangs- und Nachnutzungsbedingungen)

- Archive unterstützen die Generierung interoperabler und nachnutzbarer Daten



## 2. Erstellen eines DMP

- erster Entwurf: Strategie zur Generierung verfügbarer Forschungsdaten
  - EU: innerhalb der ersten 6 Projektmonaten
  - BMBF: Teil des Förderantrags
- aktualisierter DMP in den Zwischenberichten: Sicherstellung der Umsetzung der Strategie
- Abschlussbericht: Erklärung, wann die Daten archiviert werden bzw. wurden
  - ⇒ *„Forschungsdaten sind auffindbar über ..., zugänglich durch ..., interoperabel und nachnutzbar entsprechend disziplinspezifischer Standards“*

## 2. Erstellen eines DMP: 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

# Horizon 2020 & DMP-Vorlagen

EUROPEAN COMMISSION  
Directorate-General for Research & Innovation

**H2020 Programme:  
Guidelines on FAIR Data  
Management in Horizon  
2020.**

**Annex 1: FAIR Data  
Management at a Glance:  
Issues to Cover in Your  
Horizon 2020 DMP.**

**Version 3.0.  
26 July 2016**

**Initial Example DMP**

Project Name Example plan (Horizon 2020 DMP) - Initial Example DMP  
Project Identifier Example-DMP-ID-1  
Principal Investigator / Researcher Kerstin Helbig  
Project Data Contact +49 (0)30 2093-70072, Kerstin.Helbig@hu-berlin.de  
Description The example project is based on ABC and aims to investigate research questions XYZ. Data will be collected for the purpose of quantitative analysis and evidence-based conclusions drawing.  
Funder European Commission (Horizon 2020)

For each data set specify the following:  
Data set reference and name  
Data set description  
Example test dataset will be generated in a laboratory by the project research team. The collected dataset will be processed using various approaches. Apart from the research team, the data will be shared with other research groups working on similar questions in the area of ABC. As no comparable data is available for secondary analysis at the moment, it is planned to make our dataset publicly available in a disciplinary repository as soon as possible to support open access publications.

**Humboldt Universität  
zu Berlin:  
Muster-DMP  
Horizon 2020.  
Version 3.0.**

**Standards and metadata**  
We intend to store our dataset in a publicly accessible repository using descriptive metadata that is suitable for the discipline. Metadata for example test dataset will be stored within it separate ABC file in a standardised way by using the XXX schema, which is suitable for the discipline. Keywords will be added by using the YYY thesaurus. Files and folders will be named and structured by using a name convention that includes the project name and ID.

**Data sharing**  
We aim to publish our software code along with publishing our dataset in a disciplinary repository. Analysis will be performed using freely available open source software tools.

**Archiving and preservation (including storage and backup)**  
As specified by the "rules of good scientific practice" we aim to preserve data for at least ten years. Approximated end volume of example test dataset is X GB. Associated costs for dataset preparation for archiving will be covered by the project itself. Long-term preservation will be provided and associated costs covered by a selected disciplinary repository. During the project data will be saved daily with backup on a separate server. Backup will be checked at intervals of two weeks.

This document was generated by DMPonline (<http://dmponline.dcc.ac.uk>) 1 of 1

Vielen Dank für Ihre Aufmerksamkeit

gesis

Leibniz-Institut  
für Sozialwissenschaften

Mitglied der  
*Leibniz*  
Leibniz-Gemeinschaft

cessda

# Literatur

- Force 11 (2016): The FAIR Data Principles.
- Bundesanzeiger (2016): Bekanntmachung des BMBF. 29.03.2016.
- 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 Juli 2016.
- Horizon 2020 Programme (2016): Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020. Version 3.1. 25 August 2016.
- Horizon 2020 Programmme (2016): Guidelines on the Implementation of Open Access to Scientific Publications and Research Data in Projects supported by the European Research Council under Horizon 2020. August 2016.