

Forschungsdatenmanagement Workshop



Abb. 1: <https://pixabay.com/de/kugel-rund-ball-bin%C3%A4rball-457334/>

Anmerkung

Dieser Foliensatz dient der Unterstützung von Lehrveranstaltungen im Bereich des Forschungsdatenmanagements. Gemäß der Creative Commons Lizenz CC-BY-SA können sie spezifisch zugeschnitten und weiterverwendet werden. Die Autoren hoffen so einen Beitrag zur Verbesserung der Weiterbildung in diesem sich stark entwickelnden Bereich zu leisten.

Entstanden ist der Foliensatz aus dem baden-württembergischen Verbundprojekt „bwFDM-Info“ und dem Projekt „Community-spezifische Forschungsdatenpublikation“ des Kompetenzzentrums Forschungsdaten an der Universität Heidelberg, gefördert durch das Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg.

Autoren: Jochen Apel, Nils Föhles, Jonas Kratzke

Übersicht

- 1) Landesprojekt bwFDM-Info
- 2) Einführung
- 3) Datenerhebung
- 4) Datenhaltung & -verarbeitung
- 5) Datenpublikation
- 6) Datenarchivierung
- 7) Policies, Rechtliche Aspekte
- 8) Fazit & Abschluss



Abb. 2: <https://pixabay.com/de/fernrohr-durchblick-aussicht-blick-122960/>

Landesprojekt bwFDM-Info

Wissensvermittlung zum FDM an den Universitäten des Landes Baden-Württembergs

Gefördert durch das Ministerium für Wissenschaft,
Forschung & Kunst Baden-Württemberg



FDM-Webportal:
forschungsdaten.info

**Datenmanagement-
planungsinstrument (DMPI)**

**Koordination von
eScience-Initiativen**

Information & Beratung



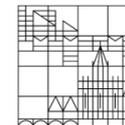
UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

UNIVERSITÄT HOHENHEIM



KIT
Karlsruhe Institute of Technology

Universität
Konstanz



EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN



Einführung

Motivation



Motivation

- 1 Datenexplosion** Immer mehr Daten, welche strukturiert und organisiert werden wollen
- 2 Datenverlust** Fragilität von Daten: Technisches/menschliches Versagen, Naturkatastrophen
- 3 Arbeitskopie** Zur Sicherung eines flüssigen Forschungsprozesses
- 4 Dokumentation** Des guten wissenschaftlichen Arbeitens
- 5 Aufbewahrung** Um etwa Nachhaltigkeitsanforderungen zu genügen

(angelehnt an: Enke & Ludwig 2013)

Forschungsdaten – eine Definition

- „Forschungsdaten sind Daten, die im Zuge wissenschaftlicher Vorhaben z.B. durch Digitalisierung, Quellenforschungen, Experimente, Messungen, Erhebungen oder Befragungen entstehen.“

(Schwerpunktinitiative „Digitale Information“ der Allianz der Wissenschaftsorganisationen)

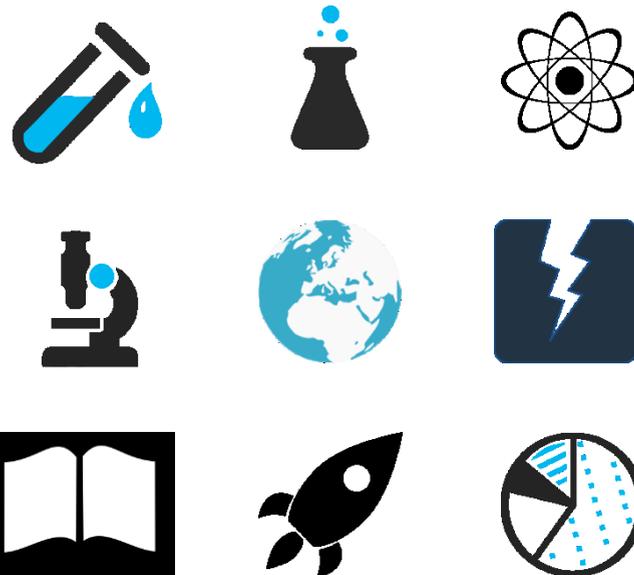


Abb. 3: <https://openclipart.org/>

Forschungsdatenmanagement

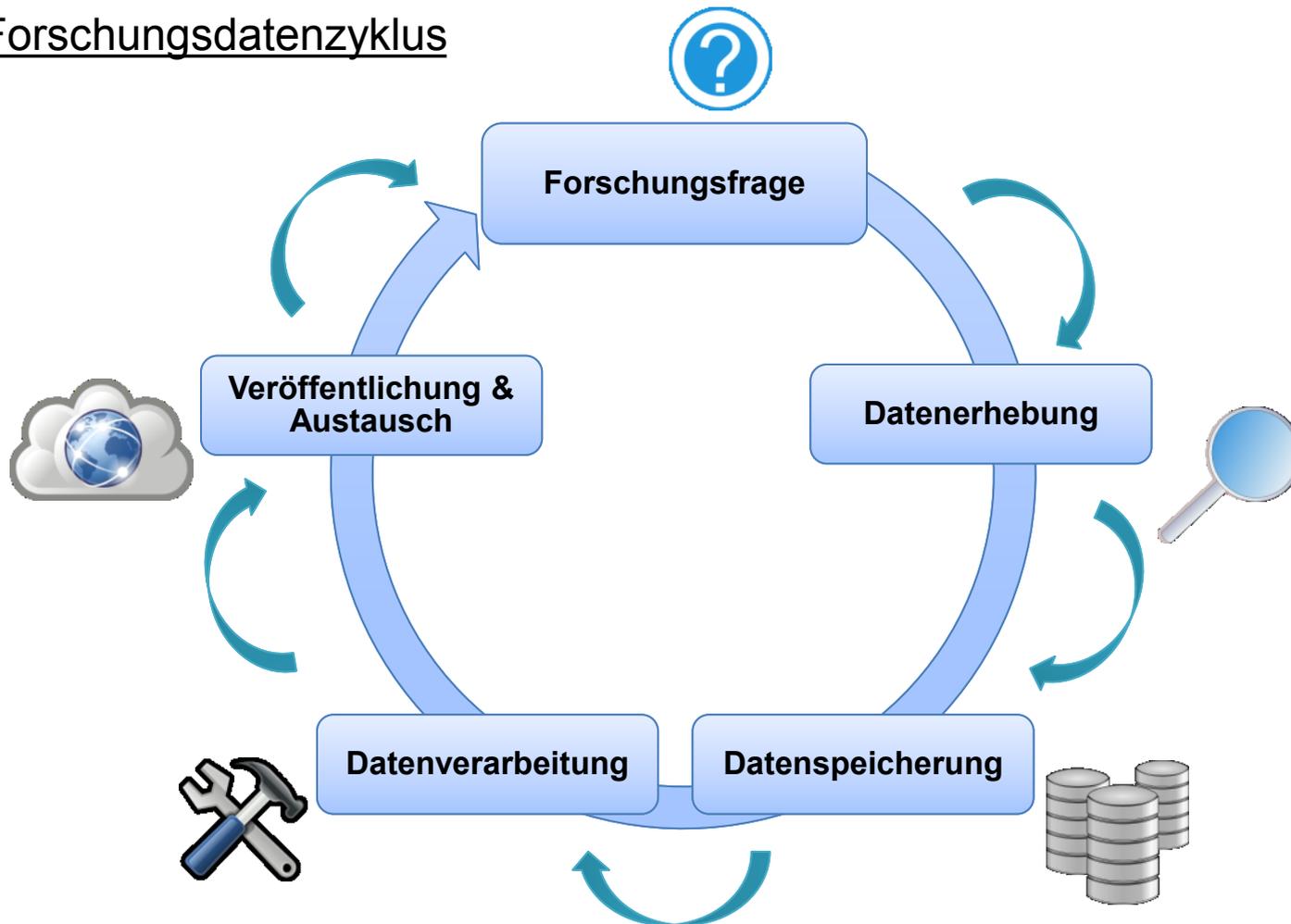
- „Maßnahmen [...], die sicherstellen, dass **digitale Forschungsdaten nutzbar** sind. Was dafür notwendig ist, variiert aber stark mit den **verschiedenen Zwecken**, für die Forschungsdaten genutzt werden sollen.“

(Enke & Ludwig 2013)



Abb. 4: <https://pixabay.com/de/kugel-rund-ball-bin%C3%A4rball-457334/>

Der Forschungsdatenzyklus



FAIR Data Management

F indable (Auffindbar)

- Persistente Adressierung; registrierte Metadaten

A ccessible (Zugänglich)

- Vorhaltung in öffentlichen Datenrepositorien

I nteroperable (Interoperabel)

- Verwendung vereinheitlichter Sprache und breit zugänglicher Formate

R eusable (Wiederverwertbar)

- Klare, offene Lizenzregelungen; Communitystandards

Weitere Informationen auf force11.org

Forschungsdatenmanagementplan

- Systematische Beschreibung des Umgangs mit Forschungsdaten
- Nachvollziehbarkeit und Interpretierbarkeit für die eigene Projektarbeit und Dritte



Einführung

Forschungsdatenmanagementplanung

Unterstützung bei der Erstellung eines Datenmanagementplans

- Erforderlicher Teil von Drittmittelanträgen
- Tool zur Datenmanagementplanung (DMP) unter <https://dmponline.dcc.ac.uk/>
- forschungsdaten.info
- RDMO



(Krähwinkel 2015)

Literatur

Dehnhard, Ina 2014: Data Management and Data Sharing - Eine Einführung. Online unter:

http://www.zpid.de/pub/research/2014_Dehnhard_Datenmanagement_dvs.pdf (letzter Zugriff: 19.04.16).

DFG 2014: Leitfaden für die Antragstellung: Projektanträge. Online unter: http://www.dfg.de/formulare/54_01/54_01_de.pdf (letzter Zugriff 10.08.16)

DOI Foundation 2016: The DOI System. Online unter: <https://www.doi.org/> (letzter Zugriff: 27.05.16)

Enke, Harry & Ludwig, Jens 2013: Leitfaden zum Forschungsdaten-Management - Handreichungen aus dem WissGrid-Projekt. Online unter: http://www.wissgrid.de/publikationen/Leitfaden_Data-Management-WissGrid.pdf (letzter Zugriff: 19.04.16).

Europäische Kommission 2016: H2020 Programm: Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020. Online unter: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf (letzter Zugriff: 10.08.16)

Krähwinkel, Esther 2015: Datenmanagementplan. Online unter: <https://www.uni-marburg.de/projekte/forschungsdaten/service> (letzter Zugriff:19.04.16).

Schwerpunktinitiative „Digitale Information“ der Allianz der Wissenschaftsorganisationen 2014. Online unter: <http://www.allianzinitiative.de/de/handlungsfelder/forschungsdaten/> (letzter Zugriff: 19.04.16).

Wissgrid 2011: Checkliste zum Forschungsdaten-Management. Online unter:

<http://www.wissgrid.de/publikationen/deliverables/wp3/WissGrid-oeffentlicher-Entwurf-Checkliste-Forschungsdaten-Management.pdf> (letzter Zugriff: 28.05.16).

Witt, Michael, et al. 2009: Constructing data curation profiles, *International Journal of Digital Curation* 4:3, 93 - 103. Online unter: <http://www.ijdc.net/index.php/ijdc/article/view/137/165> (letzter Zugriff: 16.08.2016)



Leitfragen:

- Was sind die **Rahmenbedingungen** und allgemeinen Informationen zu dem Forschungsvorhaben?
- Was sind die **Ziele** des Projektes?
- Wer sind die **Projekträger** und **-förderer**?
- Wie sind die **Zuständigkeiten** für das Management von Daten aufgeteilt?
- Wer ist verantwortlich für die **Einhaltung des Datenmanagementplans**?
- Wer sind **institutionelle Ansprechpartner** in FDM-Fragen?
- Werden **Anpassungen** an der Projektstruktur und dem Datenmanagementplan **dokumentiert**?



Datenerhebung

Art der Erhebung

- Welche Daten müssen neu erzeugt werden oder können bestehende Daten nachgenutzt werden, um das Projektziel zu erreichen? – Zwei Optionen:

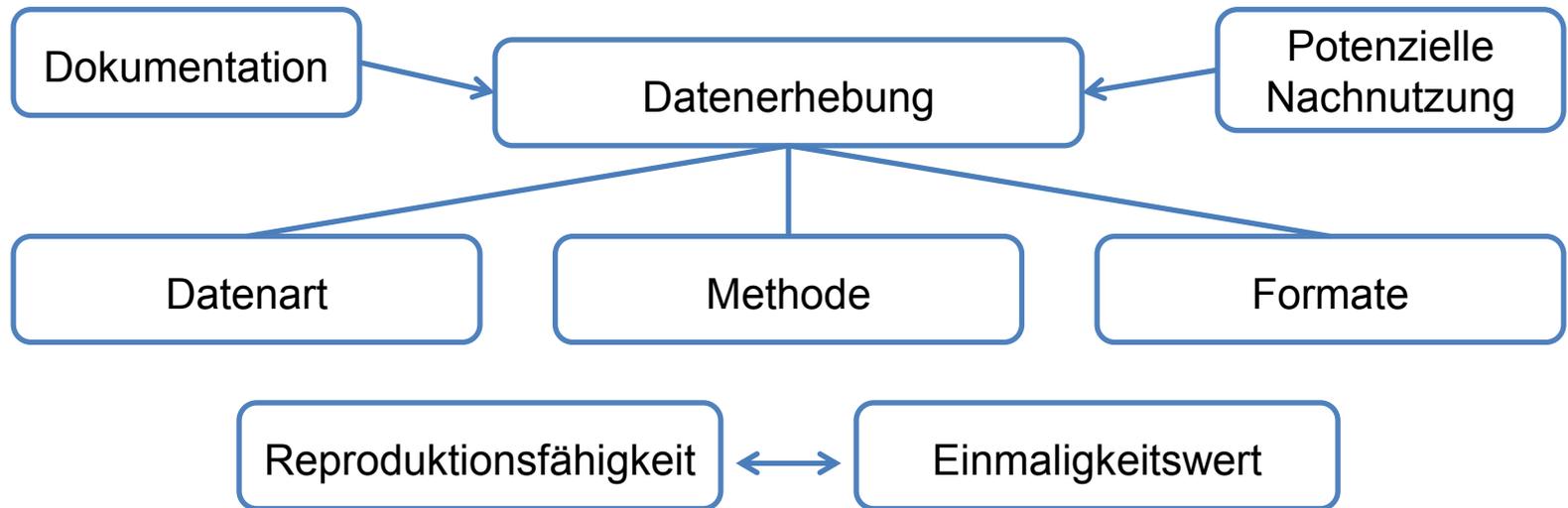
1

Projekt- und fachspezifische
Datenerhebung

2

Nutzung bereits vorhandener
Daten

Projekt- und fachspezifische Datenerhebung

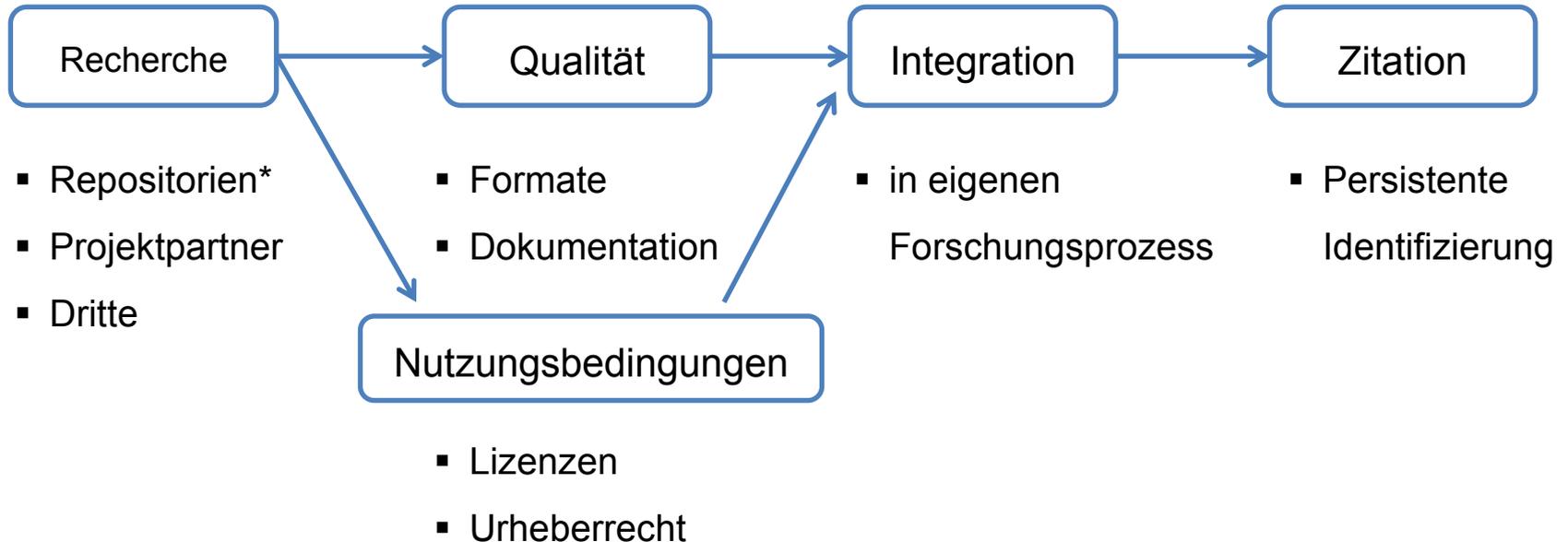




Leitfragen:

- Welche **Methode** wird bei der Datenerhebung verwendet?
- Welche **Art** von Daten wird erzeugt? (Beobachtungs- und Messdaten, prozessierte Daten, Simulationsdaten, Interviews etc.)
- Welche **Rolle** spielen die Daten innerhalb des Projektes? (z.B. Dokumentation, Publikation, Nachnutzung)
- Wie werden die Daten bei der Erhebung **dokumentiert**?
- Kann man **Anforderungen potentieller Nachnutzer** bei der Erhebung berücksichtigen?
- Handelt es sich um einmalige Daten oder können sie **reproduziert** werden?

Nutzung bereits vorhandener Daten



* Beispiele für Daten-Repositorien: re3data.org, govdata.de



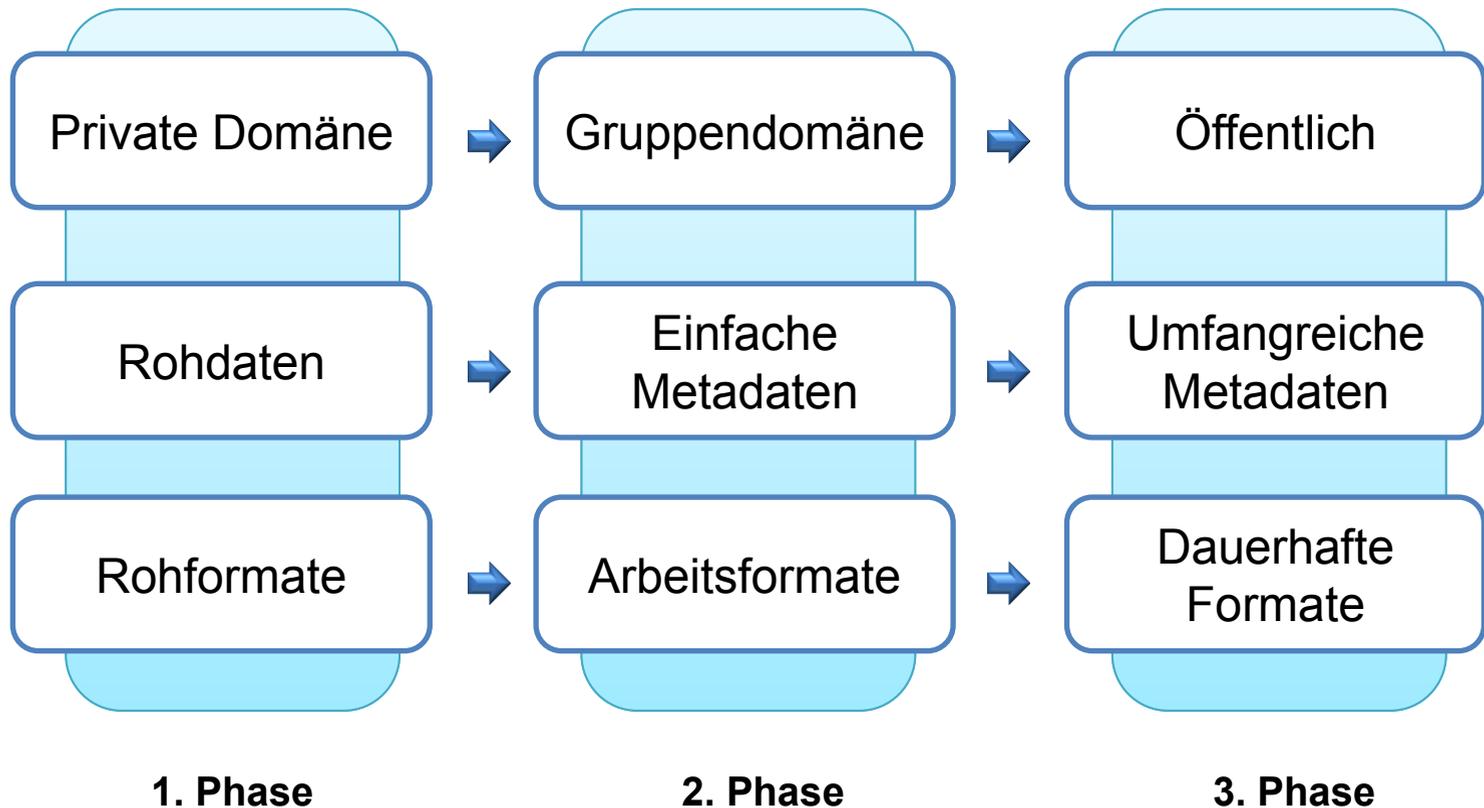
Nutzung bereits vorhandener Daten

- Wie werden **nachnutzbare Daten** recherchiert und **identifiziert**?
- Unter welchen **Lizenzen** stehen diese Daten?
- Ist die **Qualität** der Daten ausreichend? (z.B. geeignete **Formate**, ausreichende **Dokumentation**)
- Wie wird die **Integration** zu Projekt-internen Daten organisiert?
- Sind **nicht-digitalisierte Quellen** einzubeziehen?



Datenhaltung & -verarbeitung

Drei-Phasenmodell der Datenhaltung





Speicher

Leitfragen:

- Wie soll auf die Daten **zugegriffen** werden?
- Welche **Maßnahmen gegen ein versehentliches Löschen** oder eine **Manipulation** der Daten werden getroffen?
- Welche **Gründe** gibt es für eine **Aufbewahrung der Daten**?
- Welche **Daten** sollen oder müssen **aufbewahrt** werden, welche Daten sollen oder können **gelöscht** werden?
- **Wer** ist für die Aufbewahrung der Daten verantwortlich?
- Ist die **Datenablage dokumentiert**?

Metadaten

- Daten über Daten
- Standardisierte Beschreibung, möglichst maschinenlesbar
- Kategorien:

Deskriptiv	Administrativ	Technisch	Strukturell
Autor Titel Abstract Keywords ...	Urheberrechts- & Lizenzinfos DOI ...	Formatinfos Dateigrößen Spezif. Parameter ...	Verhältnis von Datenkomponenten ...

Links:

- <http://www.dlib.indiana.edu/services/metadata/index.shtml>
- <http://www.dcc.ac.uk/resources/metadata-standards>

Data Curation Profile

- Hilfe bei der Dokumentation von Datenveränderungen im Laufe des Datenlebenszyklus
- Drei zentrale Elemente:

1

Beschreibung des Datensatzes und Informationen über dessen Lebenszyklus

2

Dokumentation aller relevanten Maßnahmen, die getätigt wurden, um den Datensatz zu organisieren, zu teilen, zu verändern und zu erhalten

3

Angaben zu offenen und ungelösten Problemen hinsichtlich der getroffenen Maßnahmen

Quelle: Witt, Michael, et al. 2009: Constructing data curation profiles, International Journal of Digital Curation 4:3, 93 – 103, online abrufbar unter: <http://www.ijdc.net/index.php/ijdc/article/view/137/165> (letzter Zugriff: 10.07.2016).



Datenhaltung

Leitfragen:

- Welche Informationen sind zur **Dokumentation** der Forschungsdaten notwendig?
- Welche Zusatzinformationen sind für eine **langfristige Lesbarkeit** erforderlich?
- Wie sollen Metadaten **abgelegt** und **gespeichert** werden?
- Zu welchem **Zeitpunkt** soll die Dokumentation geschehen?
- Gibt es **Vorgaben** oder **Standards** für die Datendokumentation?
- Wie sollen Dateien **geordnet, verwaltet und versioniert** werden?
- Gibt es **Benennungsregeln** für die Dateien?
- Wie werden **Veränderungen und Aktualisierungen** der Daten dokumentiert?

Wissenschaftliche Datenformate

Anforderungen an Formate

- Umfassend dokumentiert
- Lesbar (maschinell & personell)
- Robust gegenüber Dateisystemfehlern

US Library of Congress:

http://www.digitalpreservation.gov/formats/fdd/browse_list.shtml

UK data archive:

<http://www.data-archive.ac.uk/create-manage/format/formats-table>



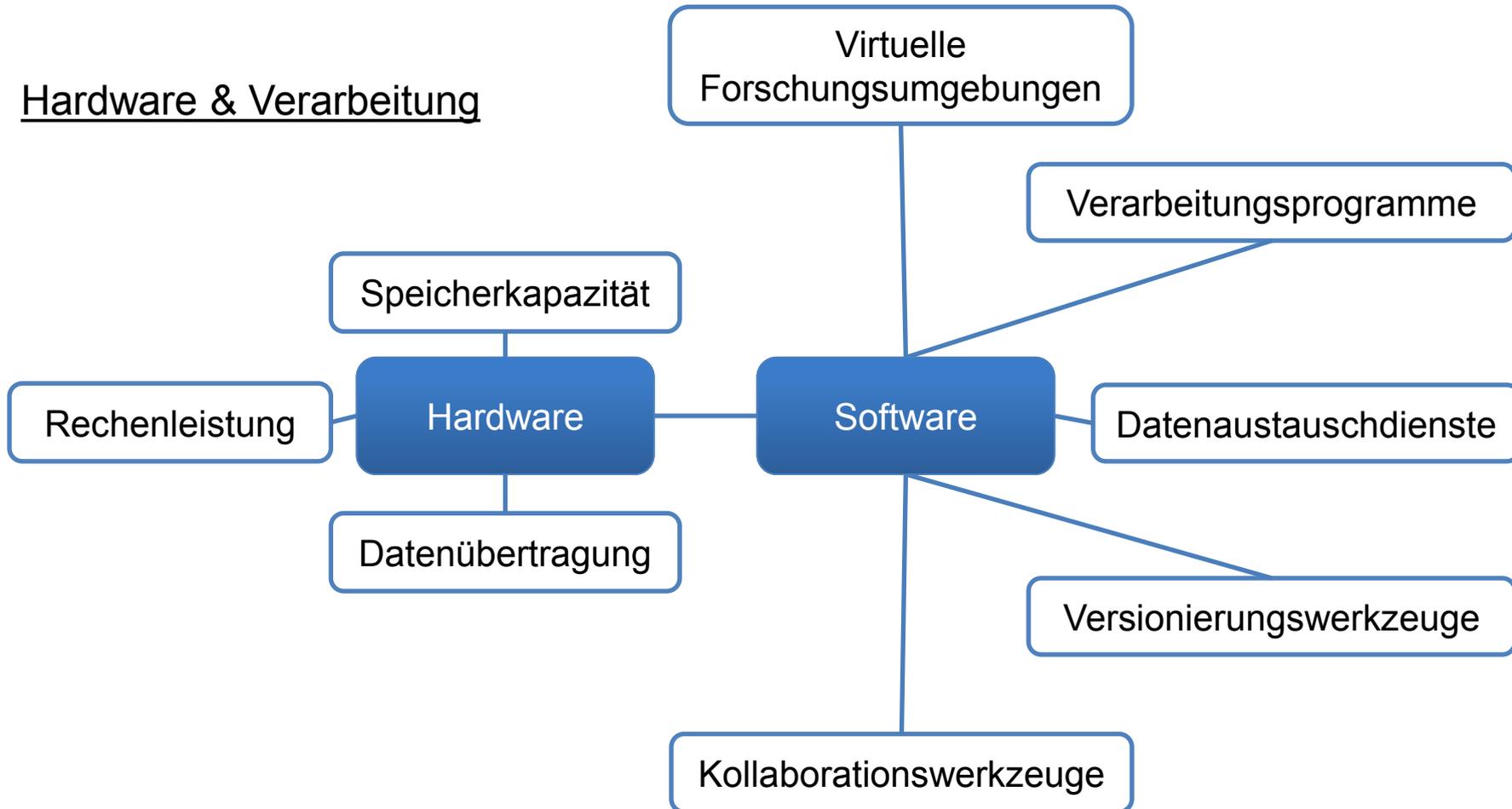
Datenformate

Leitfragen:

- In welchen **Formaten** werden die Daten erzeugt?
- Gibt es verbreitete **Standards**, die bei der **Wahl des Formats** zu beachten sind?
- Gibt es verwendbare **offene Formate** als Alternative zu **proprietären Formaten**?

Datenhaltung & -verarbeitung

Hardware & Verarbeitung





Hardware

Leitfragen:

- **Wo** werden die Daten gespeichert? (z.B. Server, Festplatten, Datenbanken)
- Was ist die erwartete **Datenmenge** für die Dauer des Forschungsprojektes?
- Ergeben sich besondere Anforderungen an die **technische Infrastruktur**? (z.B. erhöhter Bedarf an Speicherplatz, Backup, Datenübertragungsleistung)



Software

Leitfragen:

- Mit welcher **Software** werden die Daten verarbeitet?
- Wie gut sind Software und vorgesehene **Datenformate** kompatibel?
- Was sind die **Lizenzkosten** der Software? Gibt es freie (Open-Source) Alternativen?

Kollaboration

- Datenaustausch
- Zusammenarbeit
- Versionierung

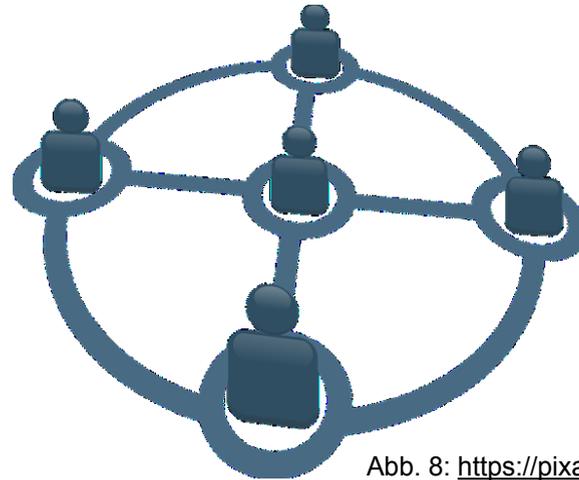


Abb. 8: <https://pixabay.com/de/konferenzstaats-und-regierungschefs-156068>

EUDAT

- <https://www.eudat.eu/>

Deutsches Forschungsnetz (DFN)

- <https://www.dfn.de/dienstleistungen/>

Versionierungstool, z.B. git

- <https://git-scm.com/>



Kollaboration

Leitfragen:

- Mit welchen **externen Partnern/Dienstleistern** soll kooperiert werden?
- **Wie** erfolgt der **Datenaustausch**?
- Welche **Sicherheitsstufe** wird benötigt?
- Welche **Menge** an Daten muss mit welcher **Häufigkeit** transferiert werden?
- Welche **technische Infrastruktur** ist für den **Datenaustausch** erforderlich?

Backup

3 ... 2 ... 1 ... Backup!

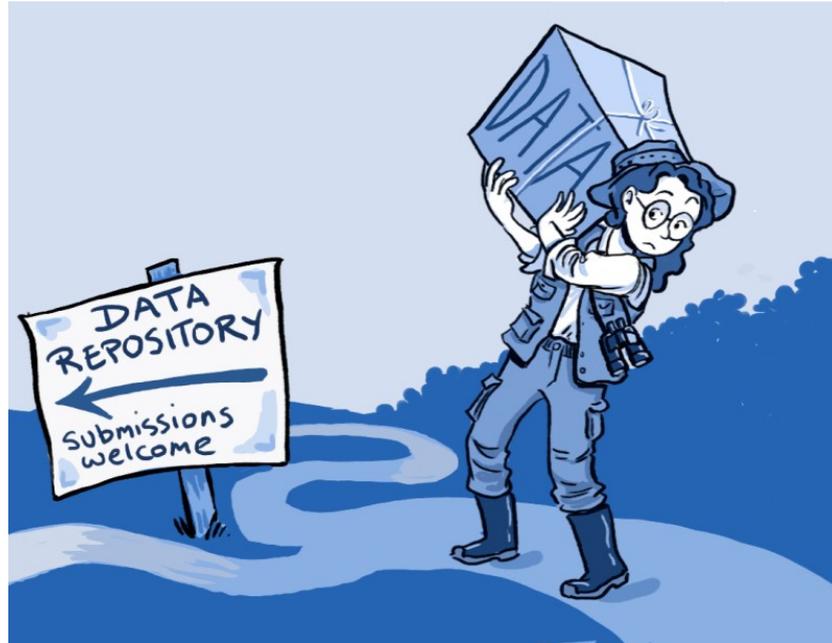
- Mindestens 3 Kopien einer Datei
- Auf mindestens 2 verschiedenen Medien mit mindestens 1 Medium, dass sich an einem anderen Ort befindet
- Zugriff nach Möglichkeit auf spezialisierte Anbieter (z.B. Rechenzentren)



Backup

Leitfragen:

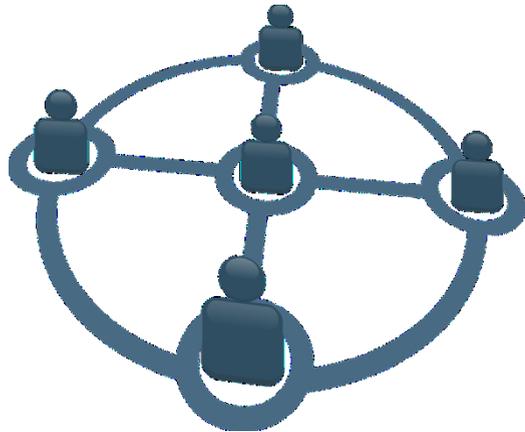
- Wie werden **Backups** der Forschungsdaten erstellt?
- **Wie oft** werden Backups durchgeführt?
- **Wer** ist für die Datenspeicherung und –sicherung **verantwortlich**?



Datenpublikation

Datenaustausch

Mit Projektpartnern



Daten sind veränderlich

Abb. 8: <https://pixabay.com/de/konferenz-staats-und-regierungschefs-156068>

(Offener) Datenaustausch



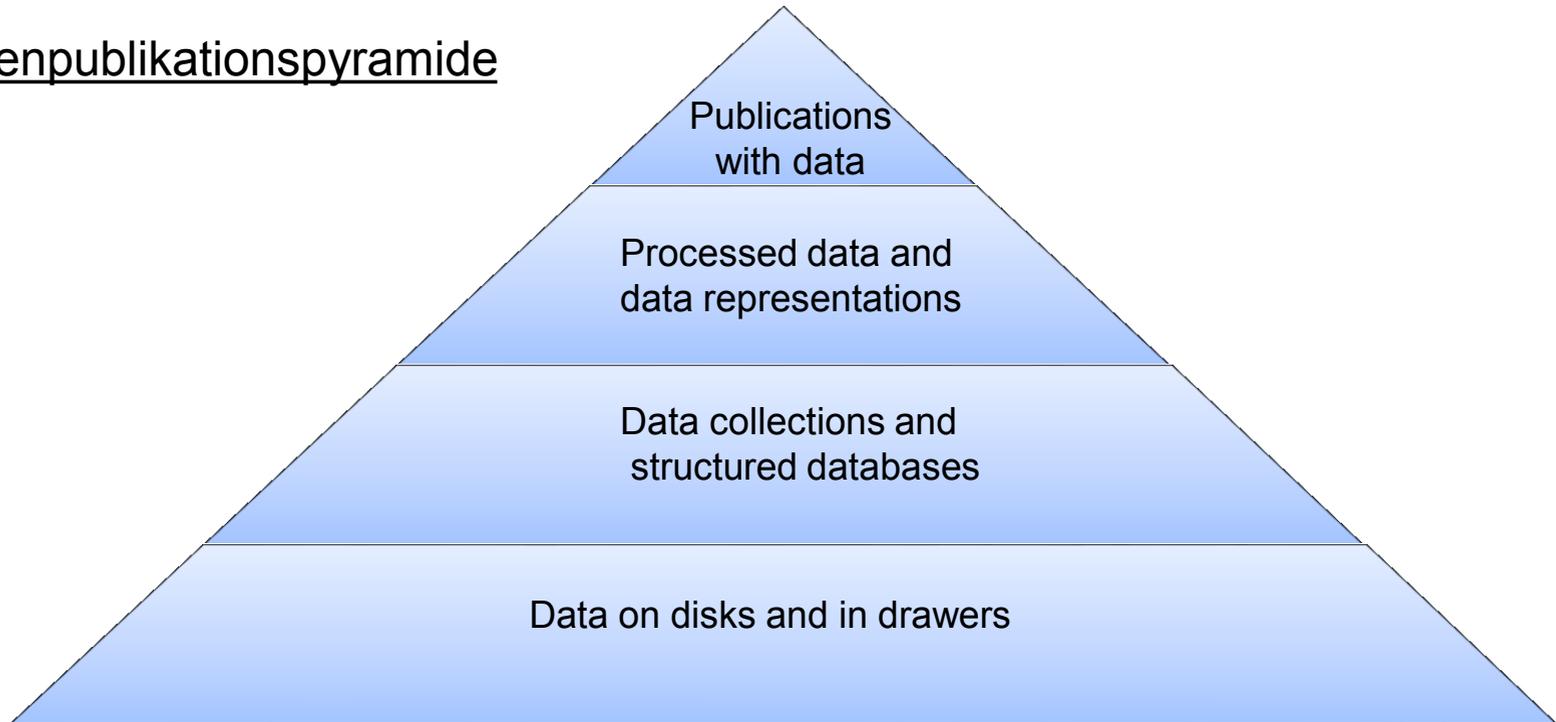
Abb. 9: <https://pixabay.com/de/suche-lupen-dokument-lupe-pr%C3%BCfen-97587/>



Abb. 10: <https://pixabay.com/de/erde-globus-menschen-silhouetten-1195668>

Daten sind persistent, auffindbar,
offen lizenziert und zitierbar

Datenpublikationspyramide



Die Datenpublikationspyramide: Veranschaulicht die Mengenverhältnisse an Manifestationen von Daten im Forschungsprozess.

Nutzen von wissenschaftlichen Publikationen

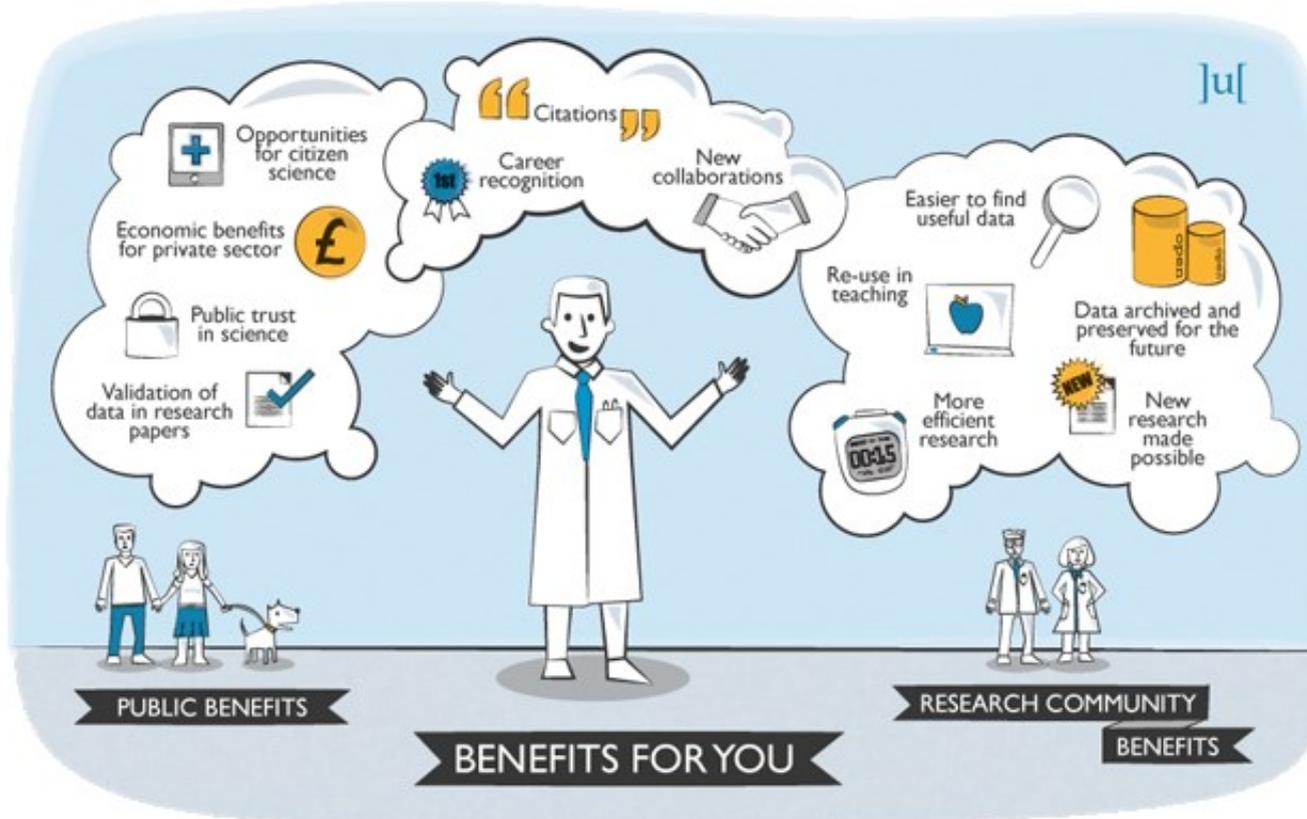


Abb. 11: <http://openarchaeologydata.metajnl.com/about>

Open Science

- Open Access: Offener Zugang zu Publikationen
- Open Science: Übergeordnetes Paradigma
- Freier Zugang nicht nur zu Publikationen, sondern auch zu weiteren Ergebnissen wissenschaftlicher Arbeit

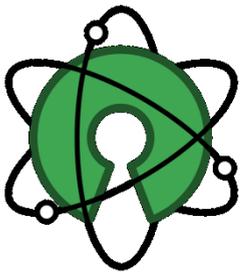


Abb. 12: <https://pixabay.com/de/offene-wissenschaft-wissenschaft-735787/>



Abb. 13: <http://okfn.de/opendata>

Publikation und Auffindbarkeit:

Wo veröffentlichen? – Beispiele:

- Fachspezifische Datenarchive und Datenzentren
 - Dryad
 - Pangaea
- Übergreifende Angebote
 - Zenodo
- Datensupplemente in Fachzeitschriften,
 - Nature
- Eigenständige Datenzeitschriften
 - GigaScience
 - Journal of Chemical and Engineering Data
- Institutionelle Repositorien

Zitierbarkeit:

Persistente Adressierung

- Dauerhafte Bezeichnung von Dateien
- Z.B. Digital Object Identifier (DOI):
 - „The DOI System provides a technical and social infrastructure for the registration and the use of persistent interoperable identifiers“ (DOI Foundation 2016)

Verzeichnis an Repositorien für Forschungsdaten





Datenpublikation

Leitfragen:

- Auf welche Weise sollen die Daten **zur Verfügung gestellt** werden?
- Welche **Zugänglichkeiten** sollen die Daten haben?
- **Wer** darf die Daten nutzen unter welchen **Lizenzen**?
- Welche **Einschränkungen im Zugriff** und der **Nachnutzung** müssen gelten?



Datenarchivierung

Warum Archivierung?

- Reproduzierbare vs. nicht-reproduzierbare Daten
- Schätzungen zufolge gehen bis zu 90% der im Forschungsprozess erzeugten Daten in einem relativ kurzen Zeitraum verloren. (Büttner et al. 2011)
- „[...] data is the currency of science, even if publications are still the currency of tenure. To be able to exchange data, communicate it, mine it, reuse it, and review it is essential to scientific productivity, collaboration, and to discovery itself.“ (Gold 2007)
- Projektende als kritischer Punkt: Speicherung und langfristige Interpretierbarkeit der Daten

Aspekte der Archivierung

Erhaltung der
Datensubstanz

- Bitstream Preservation

Erhaltung der
Funktionalität

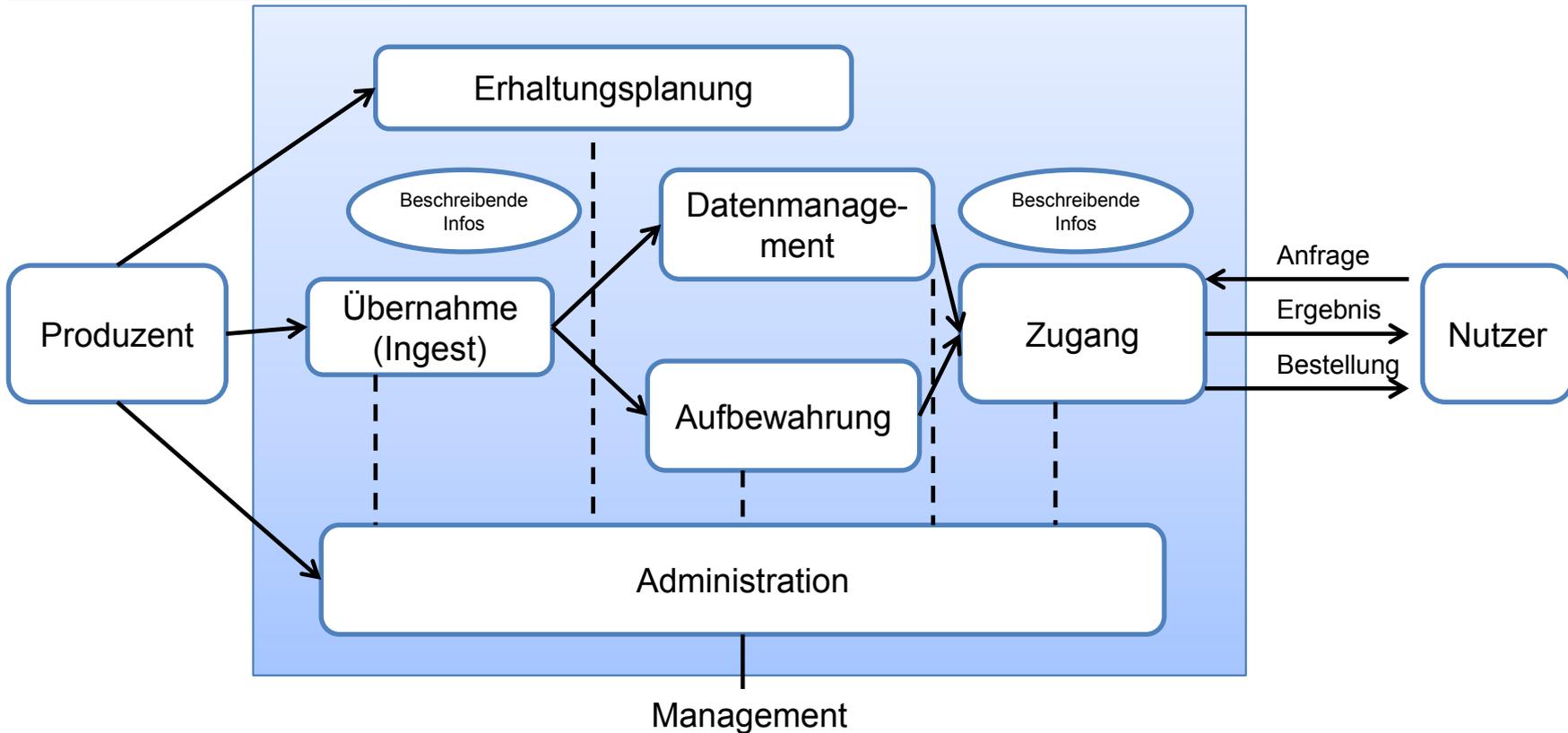
- Migration
- Emulation

Erhaltung der
Benutzbarkeit

- Documentation
- Metadata

Technische Umsetzung:

OAIS Referenzmodell



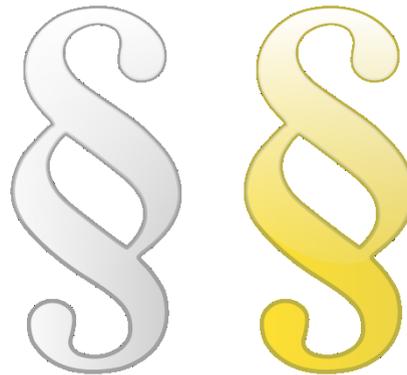
→ Metadatenpflege ab Datenerhebung



Datenarchivierung

Leitfragen:

- Mit welcher **Dauer** sollen Daten aufbewahrt werden? (z.B. bis zum Ende des Projekts, bis 10 Jahre nach Projektende, unbefristet etc.)
- **Welche Daten** müssen ganz oder teilweise aufbewahrt werden?
- Zu welchem **Zweck** sollen die Daten archiviert werden? (z.B. Nachnutzbarkeit, Dokumentation, Nachweis der guten wissenschaftlichen Praxis)
- Welches **Datenformat** ist für die Archivierung geeignet?
- Gibt es eine institutionelle **Archivlösung**?



Policies & Rechtliche Aspekte

Richtlinien der DFG

- „[...] Wenn aus Projektmitteln systematisch Forschungsdaten oder Informationen gewonnen werden, die für die Nachnutzung durch andere Wissenschaftlerinnen und Wissenschaftler geeignet sind, legen Sie bitte dar, ob und auf welche Weise diese für andere zur Verfügung gestellt werden. Bitte berücksichtigen Sie dabei auch - sofern vorhanden - die in Ihrer Fachdisziplin existierenden Standards und die Angebote existierender Datenrepositorien oder Archive. [...]“ (DFG 2014: 5)
- „Die für die Nachnutzung der Forschungsdaten anfallenden projektspezifischen Kosten können Sie im Rahmen des Projekts beantragen.“ (DFG 2014: 5)
- DFG-Leitfaden für die Antragstellung
- DFG: Anregungen und Best-Practice-Beispiele

Horizon 2020

- 8. Europäisches Forschungsrahmenprogramm
 - Laufzeit 2014-2020
 - Open Access Mandat
- Open Research Data Pilot (ORD pilot)
 - Guidelines
 - *“The ORD pilot aims to improve and maximise access to and re-use of research data generated by Horizon 2020 projects and takes into account the need to balance openness and protection of scientific information, commercialisation and Intellectual Property Rights (IPR), privacy concerns, security as well as data management and preservation questions.” (Europäische Kommission 2017: 3)*
- Erstellung eines Datenmanagementplans
 - Verpflichtend bei geförderten Projekten
 - Muss innerhalb der ersten sechs Monate des Projektes eingereicht werden
 - Muster bereitgestellt (Nutzung empfohlen aber nicht obligatorisch)



Policies & Rechtliche Aspekte



Policies

Leitfragen:

- Von welchen **Teilhabern** müssen Vorgaben berücksichtigt werden? (z.B. Universität, Projektträger, externe Partner, Urheber der Daten)
- Welche **Standards und Richtlinien** gibt es seitens der Teilhaber und der Community?
- Welche projektrelevanten **Gesetze** gibt es?
- Stellt zu nutzende **institutionelle Infrastruktur** zur Verwaltung und Speicherung von Daten eine vorgabenkonforme Dokumentation bereit?
- Gibt es **Anforderungen und Einschränkungen** für eine **Veröffentlichung** der Daten?

Datenschutz bei Forschungsdaten

Sensible Daten nach Bundesdatenschutzgesetz (BDSG) §3:

- Personenbezogene Daten, insbesondere Herkunft, politische Meinung, Religion, Gesundheit und Sexualleben

Weitere Gesetze:

- Gesetz über die Statistik für Bundeszwecke (Bundesstatistikgesetz - BStatG)
- SGB X Sozialgesetzbuch – Sozialverwaltungsverfahren und Sozialdatenschutz

Datenschutz bei Forschungsdaten

Umgang mit personenbezogenen Daten in der Forschung:

Anonymisieren

- Weitestgehende Aufhebung der Zuordnung personenbezogener Daten

Pseudonomisieren

- Ersetzung von Identifikationsmerkmalen

Aggregieren

- Aufstellung einer Statistik über einen ausreichend großen Personenkreis

Forschungsdaten und Urheberrecht

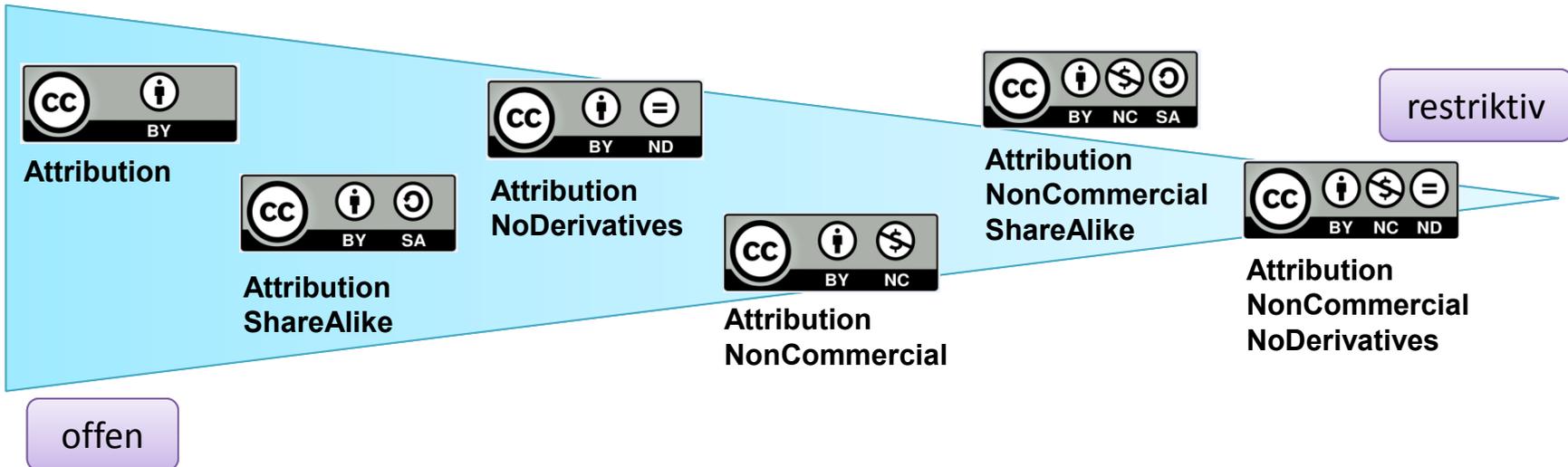
- Das Urheberrecht „wahrt die Persönlichkeitsrechte des Autors wie Erstveröffentlichung, Urheberbezeichnung und Rückruf des Werkes und sichert dessen Verwertungsrechte“ (siehe Forschungsdaten.org)
- Das Urheberrecht gilt für literarische, künstlerische und wissenschaftliche Werke, welche die Schutzvoraussetzungen erfüllen
- Fakten wie Messwerte erreichen in der Regel nicht die erforderliche Schöpfungshöhe
- Datenbanken können nach EU Richtlinie 96/09/EG, UrhG §§ 87a-e urheberrechtlich geschützt werden

Policies & rechtliche Aspekte

Hilfestellung bei der Auswahl unter:
<https://creativecommons.org/choose/>

Creative Commons Lizenzen

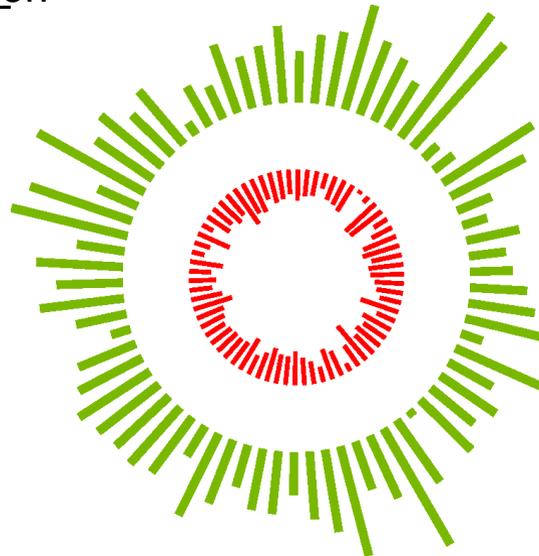
- Standardlizenzverträge
- Sowohl leicht verständliche Formulierungen als auch die jeweilige nationale rechtsentsprechende Form
- Modulare Struktur mit mehreren „Freiheitsgraden“



Policies & rechtliche Aspekte

Open Data Commons

- Lizenzen speziell für Datenbanken
- Inhalte einer Datenbank können gesondert lizenziert werden.
- Open Knowledge Foundation



OPEN KNOWLEDGE

Abb. 15: <http://assets.okfn.org/p/okfn/img/okfn-logo-portrait-l.png>

Policies & Rechtliche Aspekte



Rechtliche Aspekte

Leitfragen:

- Welche Daten sind **urheberrechtlich geschützt**?
- Welche Daten fallen unter den **Datenschutz**?
- Wie werden die **Rechte an geistigem Eigentum** für die Daten dokumentiert?
- Unter welchen **Lizenzen** können Daten nachgenutzt, bzw. für Dritte zur Verfügung gestellt werden?
- Wie können Vereinbarungen mit **Datenarchiven, bzw. -repositorien** vertraglich festgehalten werden?

Vielen Dank!

Weitere Anregungen



EUDAT



Open Aire



Forschungsdaten.org



Forschungsdaten.info