
DATA GOVERNANCE UND DATA STEWARDSHIP IN DER FRAUNHOFER-GESELLSCHAFT

17.11.2020 11. DINI/nestor-Workshop:

Data Stewardship im Forschungsdatenmanagement - Was ist das? Rollen, Aufgabenprofile, Einsatzgebiete

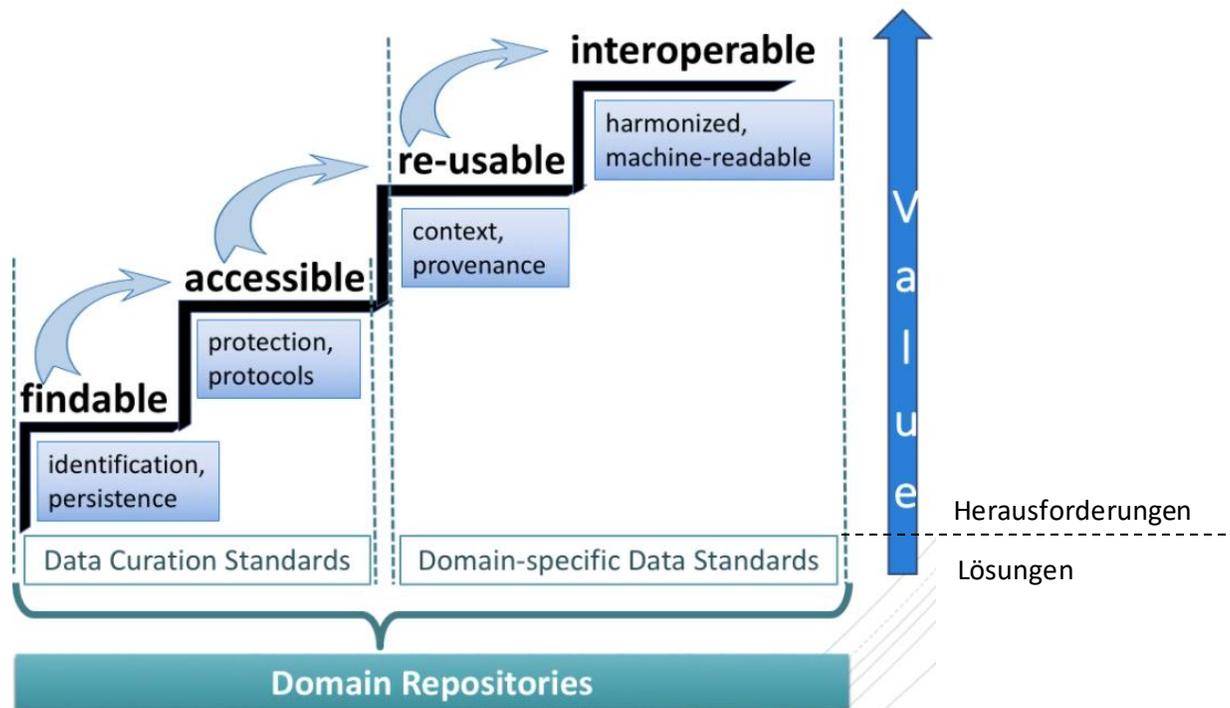


Andrea Wuchner, Fraunhofer IRB

Johannes Sautter, Fraunhofer IAO

Wie können wir sukzessive den Wert der Daten erhöhen?

Datenperspektive



Nach Tim Ahern, Ted Haberman, Kerstin Lehnert: The "I&R" in FAIR: Who develops, approves, and governs domain-specific standards in the Earth, Space, & Environmental Sciences?, 2019 ESIP Winter Meeting

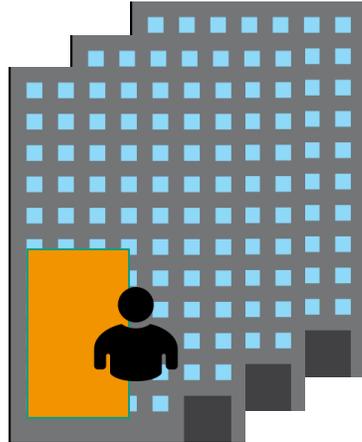
Organisationsperspektive

Herausforderungen

- **Operative Exzellenz**
 - Operative Effizienz
 - Austausch und Kollaboration
 - Datenverfügbarkeit
 - Operative Integration und Interoperabilität
- **Rechtliche Exzellenz**
 - Operative rechtliche Anforderungen
 - Kenntnis der Vorschriften
- **Exzellenz in der Prozessqualität**
 - Klare und zuverlässige Verantwortlichkeiten, Prozesse und Richtlinien
 - Datentransparenz und Überprüfbarkeit
 - Zentraler Datenzugriff
- **Datenqualität**
 - Intrinsisch
 - Zugänglichkeits-betr.
 - Kontextuell
 - Repräsentativ

Quelle: Sautter, J., Litauer, R., Fischer, R., Klages, T., Wuchner, A., Müller, E., Schaj, G., Dobrokhotova, E., Drews, P. and Riess, S.: Beyond Data Quality: Data Excellence Challenges from an Enterprise, Research and City Perspective. In Proceedings of the 7th International Conference on Data Science, Technology and Applications (DATA 2018), pages 245-252, ISBN: 978-989-758-318-6

Datenverantwortung wird an den Fraunhofer-Instituten übernommen



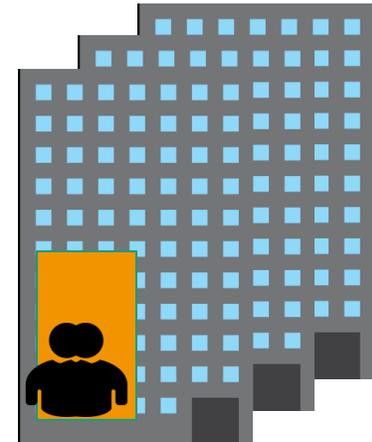
Aufgaben Fachinformationsmanager

- Versorgung mit Fachliteratur
- Unterstützung wissenschaftliches Publizieren
- Beratung im Forschungsdatenmanagement



Aufgaben zentrale Dienstleister

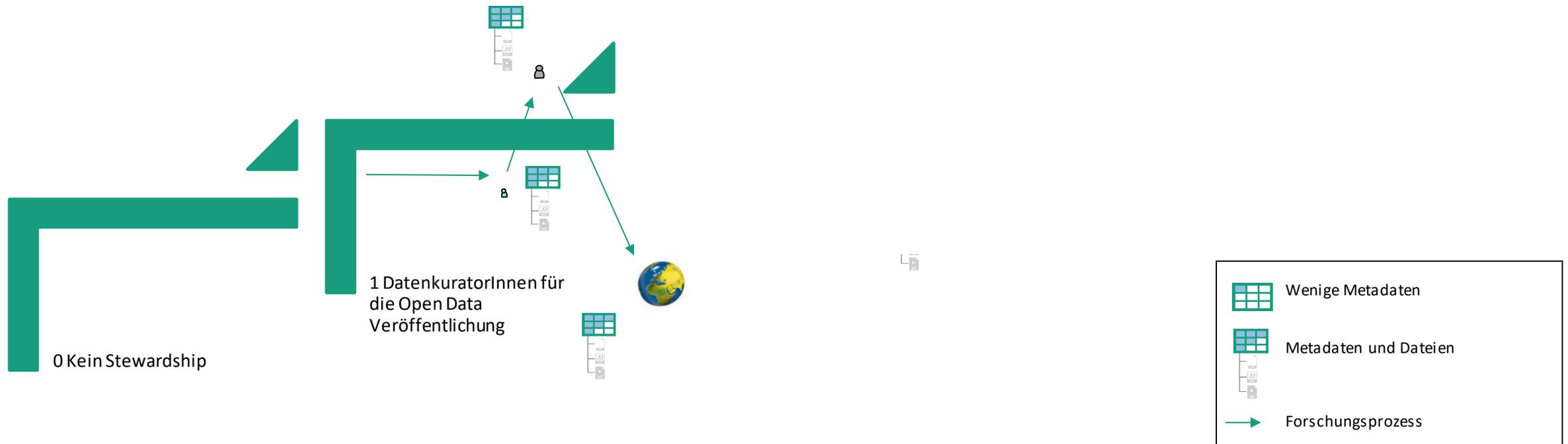
- Support
- Betrieb Forschungsdatenrepositorium
- Zentrale Materialien, Regelwerk



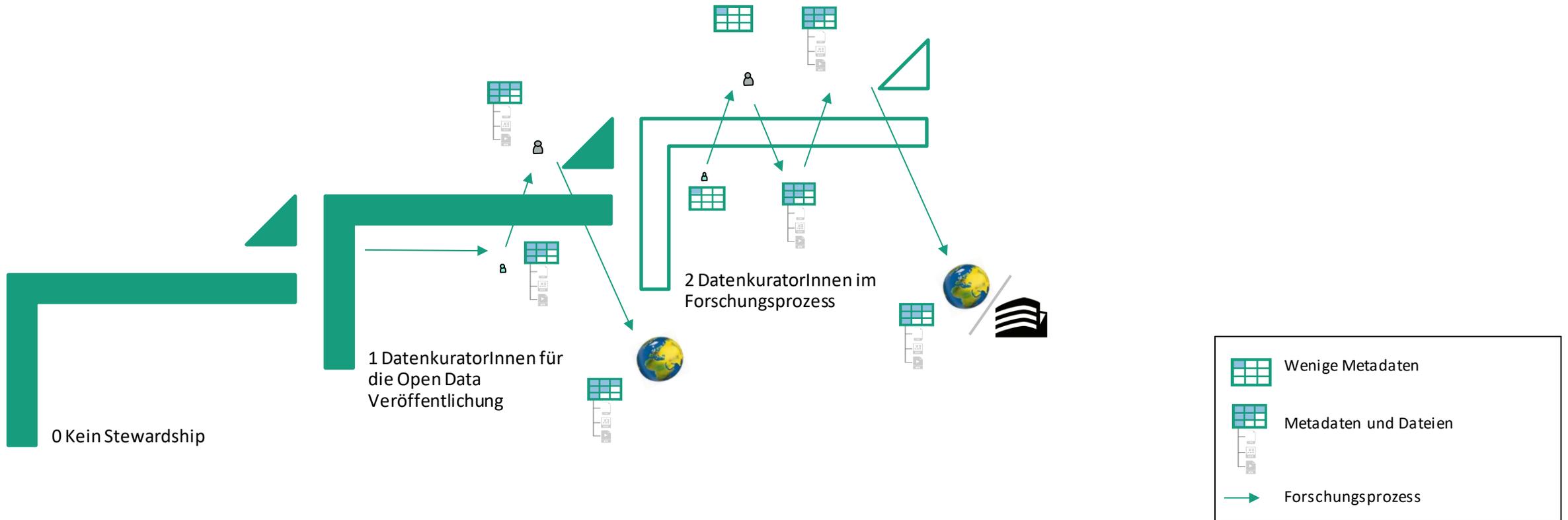
Aufgaben Datenkurator

- Überprüfung der Metadaten
- Pflege der Sammlung
- Unterstützung bei Eingabe

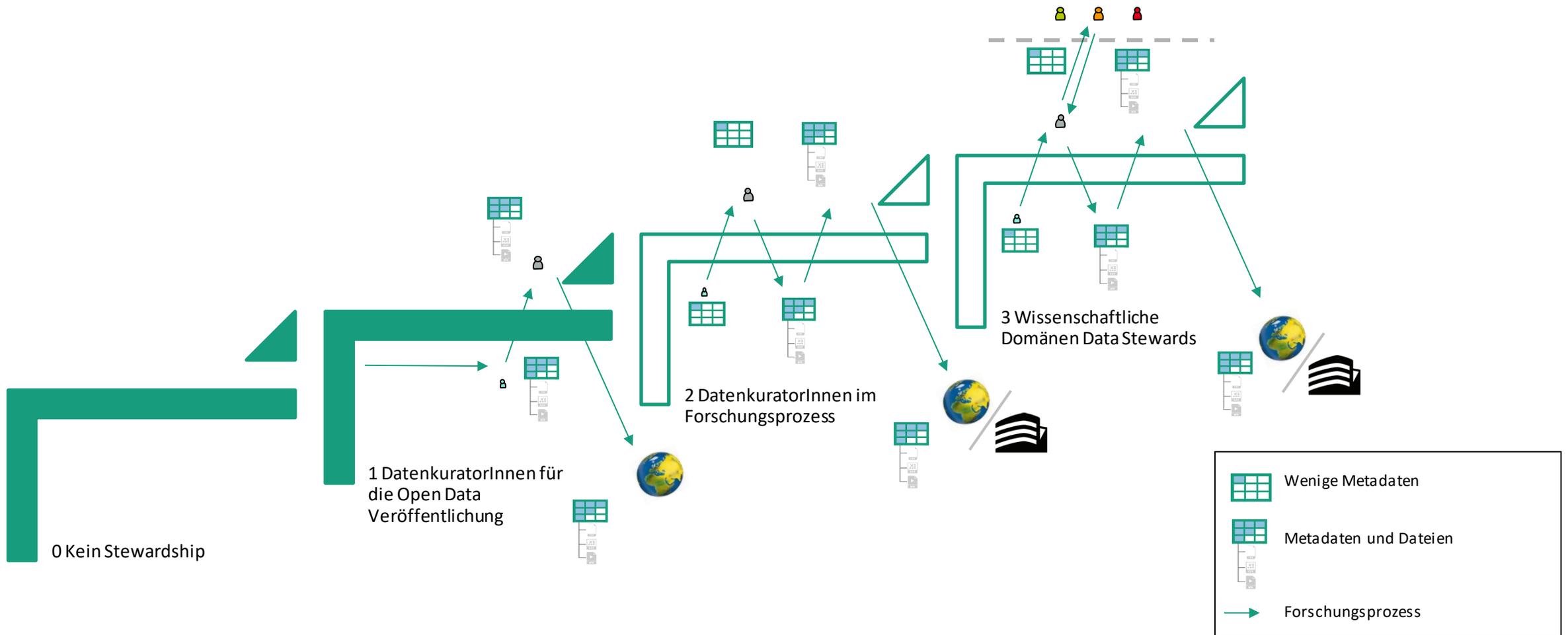
Status quo: Data Stewardship bei Fraunhofer



Perspektiven: Data Stewardship bei Fraunhofer



Perspektiven: Data Stewardship bei Fraunhofer



- I Einführung
- II Abfolge der Daten
- 1 Forschungsdesign**
- 2 Suche
- 3 Vorbereitung der Erhebung
- 4 Datenerhebung
- 5 Datenaufbereitung
- 6 Datenauswertung
- 7 Dokumentation und Abschluss
- III Literatur

1 Forschungsdesign

Schritt	Aktivität
1	Präzisierung des Untersuchungsgegenstandes und Generierung einer Forschungsfrage
2	Die Voraussetzungen für qualitatives Vorgehen stimmen mit meinem Erkenntnisinteresse überein
3	Bestimmung der Fallauswahl → siehe Tipps und Hinweise 1 #Fallauswahl

Alle Schritte erledigt? Dann folgende Metadaten ausfüllen:

Metadaten der Ablaufphase

In dieser Ablaufphase (DA1) müssen bestimmte Metadaten eingegeben und aktualisiert werden. In der Metadaten-Excel-Tabelle müssen daher die folgenden Tabellenreiter bearbeitet werden:

- Projektmitarbeiter
- Forschungsdatensatz
- Forschungsmethode

Welche Metadaten zum ersten Mal eingegeben oder aktualisiert werden müssen, lässt sich der farblichen Markierung und Beschriftung direkt im jeweiligen Tabellenreiter der Metadaten-Excel-Tabelle entnehmen.

• **Tipps und Hinweise 1**

2 Suche

Das vorliegende Kapital soll Forschende dahingehend sensibilisieren, vor der Erhebung eine Suche nach möglicherweise bereits vorhandenen Daten im Sinne einer Literatursuche durchzuführen. Diese Suche kann beispielsweise auf das Erkenntnisinteresse oder auf die Zusammensetzung der Stichprobe ausgerichtet sein. Das Ziel ist es redundante Erhebungen zu vermeiden und eine Weiterverwendung vorhandener Daten im eigenen Vorhaben. Informationen zu Internetseiten, auf denen Inhalte wie z.B. Datensätze gefunden werden können, sind zu finden unter: http://publikationssupport.humboldt-berlin.de/index.php/jahresmitteilung_bei_der_suche

Alle Schritte erledigt? Dann folgende Metadaten ausfüllen:

Metadaten der Ablaufphase

In dieser Ablaufphase (DA2) müssen bestimmte Metadaten eingegeben und aktualisiert werden. In der Metadaten-Excel-Tabelle müssen daher die folgenden Tabellenreiter bearbeitet werden:

- Forschungsdatensatz

Welche Metadaten zum ersten Mal eingegeben oder aktualisiert werden müssen, lässt sich der farblichen Markierung und Beschriftung direkt im jeweiligen Tabellenreiter der Metadaten-Excel-Tabelle entnehmen.

3 Vorbereitung der Erhebung

Schritt	Aktivität
4	Erstellung eines Leitfadens
4.1	Nachfragen zu Unterspeziken und »Nächste-Techniken« formulieren (zur gezielten Lenkung des Gesprächs)
5	Datenschutz und Rechtsicherheit → siehe Tipps und Hinweise 2 & 4
5.1	Prüfen ob eigenes Datenschutzkonzept nötig ist oder Standard-Informationen reichen
6	Erster Kontakt zu möglichen Interviewpartner:innen (telefonischer Mail)
6.1	Offizielle Einladung an Interviewpartner:innen versenden
7	Erstellung einer ersten Codebuch-Version

Alle Schritte erledigt? Dann folgende Metadaten ausfüllen:

Metadaten der Ablaufphase

In dieser Ablaufphase (DA3) müssen bestimmte Metadaten eingegeben und aktualisiert werden. In der Metadaten-Excel-Tabelle müssen daher die folgenden Tabellenreiter bearbeitet werden:

- Forschungsdatensatz
- Detail

Welche Metadaten zum ersten Mal eingegeben oder aktualisiert werden müssen, lässt sich der farblichen Markierung und Beschriftung direkt im jeweiligen Tabellenreiter der Metadaten-Excel-Tabelle entnehmen.

Daten der Ablaufphase

Datent	Vorlage	Pflicht?	Sensibilitätstufe
Interviewleitfaden		ja	normal
Einverständniserklärung	Einwilligungserklärung.docx	ja	normal
Datenschutzinformation	Datenschutzinformation_interview.docx	ja	normal

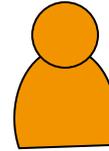
• **Tipps und Hinweise 3**

4 Datenerhebung

Schritt	Aktivität
8	Ort und Termine mit Interviewpartner:innen abstimmen und festlegen (evtl. Räumlichkeiten buchen)
8.1	Interview-Liste erstellen (anonymisiert, Entscheidung nur Projektleiter:in + Mitarbeiter:innen)

Data Stewards für die Domäne

im HEFE-Pilot



■ Domäne = Logische Gruppierung bestimmter Forschungsdaten nach deren Erhebung

(bspw. Fragebogen, Mobilitäts-Sensor-Daten mit Zeitbezug oder Gebäudemodell)

■ Data Steward ist erfahrene(r) WissenschaftlerIn in der Domäne

■ Regelmäßig tagendes Domänengremium

■ Austausch mit zuständigen DatenkuratorInnen

■ Fachliche Datenverantwortung

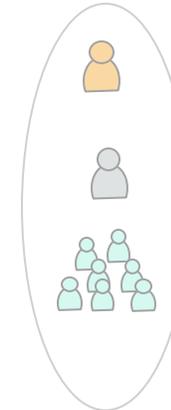
■ Moderationskompetenz

■ Erstellung und Pflege von Checklisten

■ Empfehlungen zu Schritten + Arbeitshilfen

■ Regeln für Metadaten (Pflichtfelder)

■ Regeln zu (Pflicht-) Dateien und Dateistruktur



- I Vorwort
- II Abfolge der Daten
- 1 Forschungsdesign**
- 2 Datensuche und -akquisition
- 3 Vorbereitung der Datenerhebung
- 4 Datenerhebung oder -übergabe
- 5 Datenaufbereitung
- 6 Datenauswertung
- 7 Veröffentlichung
- III Literatur

1 Forschungsdesign

Schritt	Aktivität
1	Forschungsfrage präzisieren
2	Grobkonzept erstellen
3	Finkonzept erstellen

Metadaten der Ablaufphase

In dieser Ablaufphase (DA1) müssen bestimmte Metadaten eingegeben und aktualisiert werden. In der Metadaten-Excel-Tabelle müssen daher die folgenden Tabellenreiter bearbeitet werden:

- Projektmitarbeiter
- Forschungsdatensatz
- Forschungsmethode
- Detail
- Software

Daten der Ablaufphase

Datent	Erhebtes Datentformat	Langzeitformat	Vorlage	Pflicht? (Ja/Nein)	Sensibilitätstufe
Fragebogenkonzept, ggf. Übersetzungen	DOC, XLS, PDF	PDF		ja	
ggf. Vertrag mit Dienstleister	PDF	PDF		ja	

• **Tipps und Hinweise 1**

2 Datensuche und -akquisition

Das vorliegende Kapital soll Forschende dahingehend sensibilisieren, vor der Erhebung eine Suche nach möglicherweise bereits vorhandenen Daten im Sinne einer Literatursuche durchzuführen. Diese Suche kann beispielsweise auf das Erkenntnisinteresse oder auf die Zusammensetzung der Stichprobe ausgerichtet sein. Das Ziel ist es redundante Erhebungen zu vermeiden und eine Weiterverwendung vorhandener Daten im eigenen Vorhaben. Informationen zu Internetseiten, auf denen Inhalte wie z.B. Datensätze gefunden werden können, sind zu finden unter: publikationssupport.humboldt-berlin.de

Metadaten der Ablaufphase

In dieser Ablaufphase (DA2) müssen bestimmte Metadaten eingegeben und aktualisiert werden. In der Metadaten-Excel-Tabelle müssen daher die folgenden Tabellenreiter bearbeitet werden:

- Forschungsdatensatz

Welche Metadaten zum ersten Mal eingegeben oder aktualisiert werden müssen, lässt sich der farblichen Markierung und Beschriftung direkt im jeweiligen Tabellenreiter der Metadaten-Excel-Tabelle entnehmen.

3 Vorbereitung der Datenerhebung

Schritt	Aktivität
4	Operationalisierung der Befragung
5	Umsetzung der Befragung
6	Datenschutz
7	Pretest
8	Konkreten Befragungszeitraum definieren

Daten der Ablaufphase

Datent	Erhebtes Datentformat	Langzeitformat	Vorlage	Pflicht? (Ja/Nein)	Sensibilitätstufe
ggf. Einwilligungserklärung	PDF	PDF		ja	ja
ggf. Befragungsstruktur	LIS, XLSL	XLSL		nein	

• **Tipps und Hinweise 3**

4 Datenerhebung oder -übergabe

Schritt	Aktivität
9	Einladungen erstellen
10	Rücklaufmanagement
11	Befragung abschließen

Daten der Ablaufphase

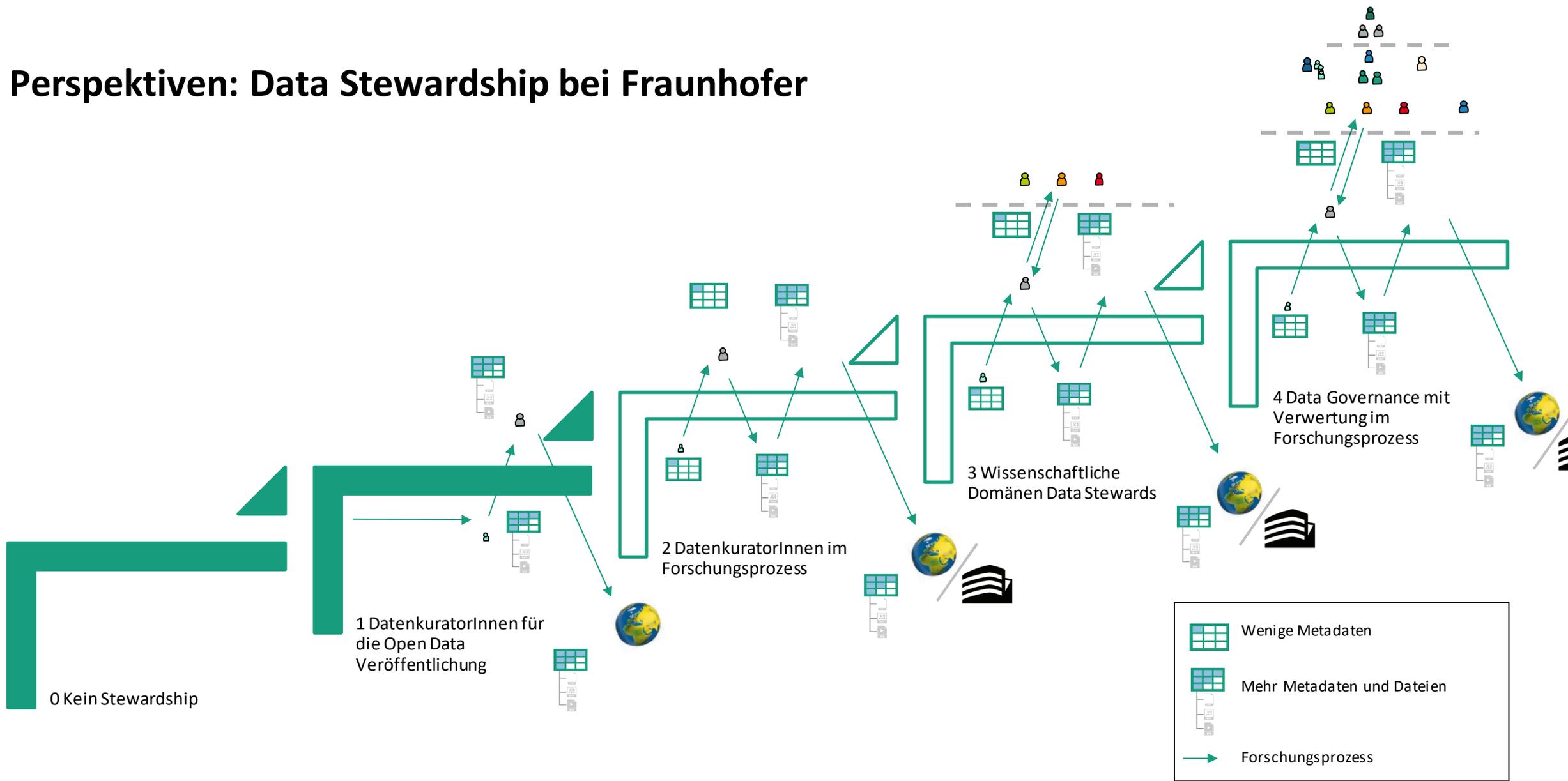
Datent	Erhebtes Datentformat	Langzeitformat	Vorlage	Pflicht? (Ja/Nein)	Sensibilitätstufe
Ggf. Rohdatensatz	LISA, CSV, XLS, SPSS	CSV		nein	

• **Tipps und Hinweise 4**

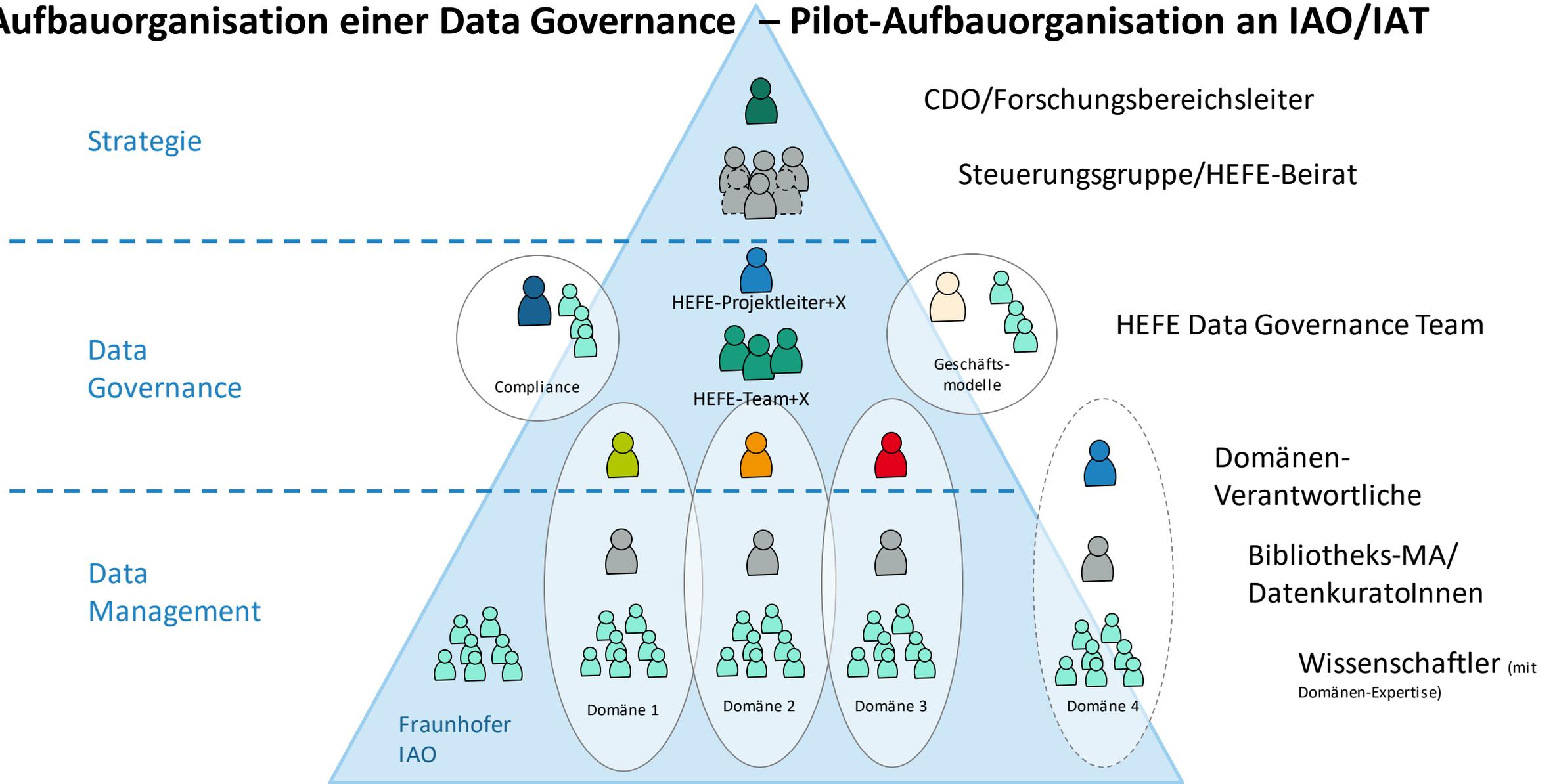
Checkliste zur Planung, Durchführung und Auswertung qualitativer Leitfadengespräche

Checkliste zur Planung, Durchführung und Auswertung von (Online-) Befragungen

Perspektiven: Data Stewardship bei Fraunhofer



Aufbauorganisation einer Data Governance – Pilot-Aufbauorganisation an IAO/IAT



Der Data Governance-Ansatz wird weiterverfolgt

Weitere Aktivitäten

- Data Governance im Projekt »Krisenmanagement und Resilienz – Corona« (KResCo)
 - Data Curation Profiles für 3 Domänen in 5 Nachnutzungen und Erhebungs-Aps geplant?
 - Erster Test des iterativen »Forschungsprozesses der Zukunft«
- Ganzheitlicher Data Governance-Ansatz innerhalb der Fraunhofer-Gesellschaft
 - Fraunhofer Datenraum für interne Datenpublikation und –suche
 - Business Analytics-Perspektiven auf Projekt- und Forschungsdaten

Herzlichen Dank für Ihre Aufmerksamkeit!

Fragen?

Kontakt

Competence Center
Research Services & Open
Science
Fraunhofer-
Informationszentrum Raum
und Bau IRB

Andrea Wuchner
Fraunhofer IRB
Andrea.Wuchner@irb.fraunhofer.de
Tel: +49 (0)711 / 970-2714



&

Urban Data and Resilience
Fraunhofer-Institut für
Arbeitswirtschaft und
Organisation IAO

Johannes Sautter
Fraunhofer IAO
Johannes.Sautter@iao.fraunhofer.de
Tel: +49 (0)711 / 970-2387



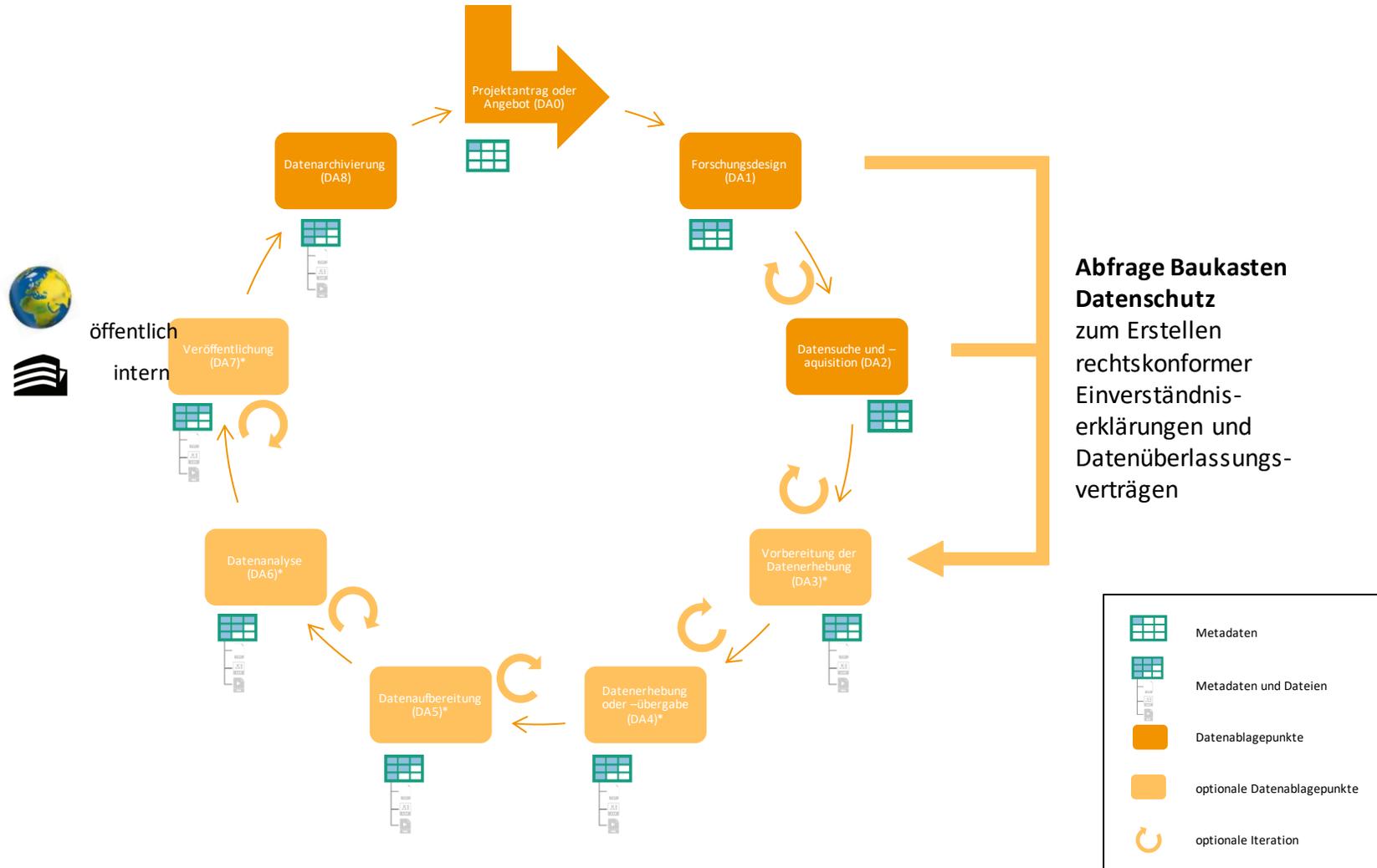
Nobelstraße 12
D-70569 Stuttgart

<http://s.fhg.de/datenexzellenz-forschung>

BACKUP-FOLIEN

»Forschungsprozess der Zukunft«

Definition von Datenprodukten als Zwischenergebnisse eines innovativen Prüfprozesses

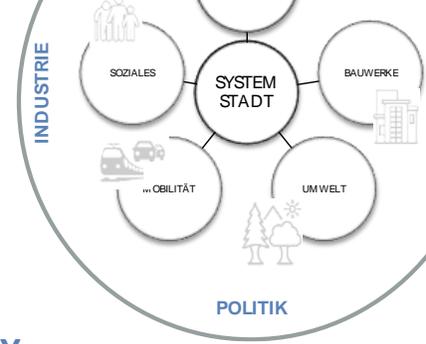


- Vision »Erfassung der genehmigten Projektanträge, halbautomatische Nutzung als Metadaten«
- Interne Plattform für geplante Datenerhebungsaktivitäten
- Prüfung durch Datenkuratoren (FIM) Regelwerke mit 5 Qualitätsstufen/-spezifika geplant:
 - Planung
 - Forschungsdesign
 - Datenerhebung
 - Veröffentlichung
 - Archivierung

Rollen und Konzepte NFDI4City



Data Owner
(Führungskraft)



NFDI4CITY

Soziales

Umwelt

Bauwerke

Verwaltung

Mobilität

Sektor

Data Steward

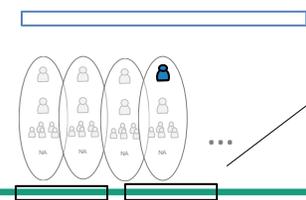
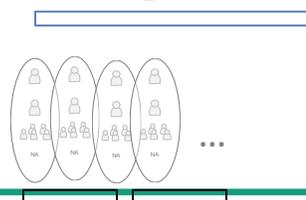
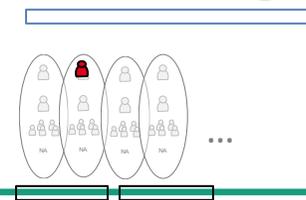
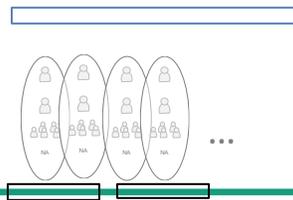
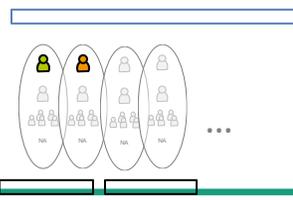
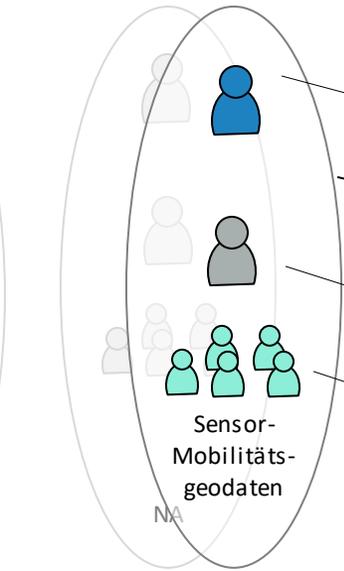
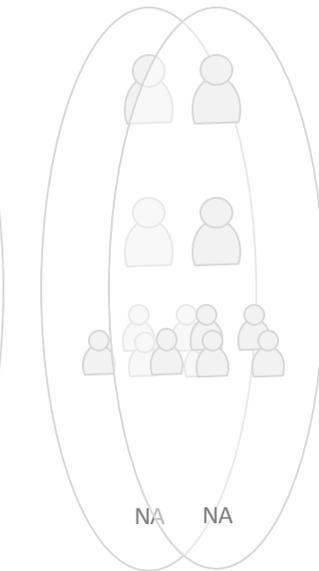
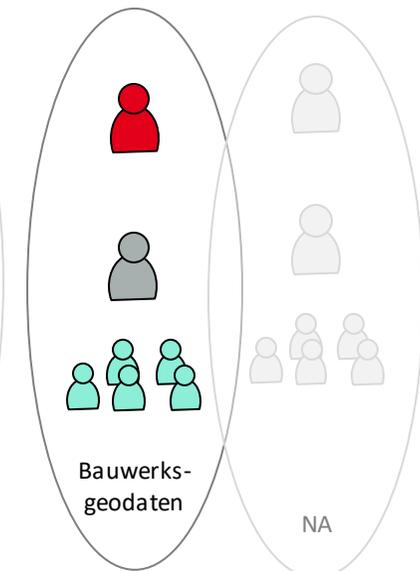
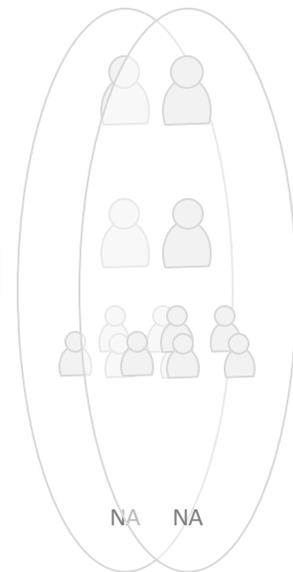
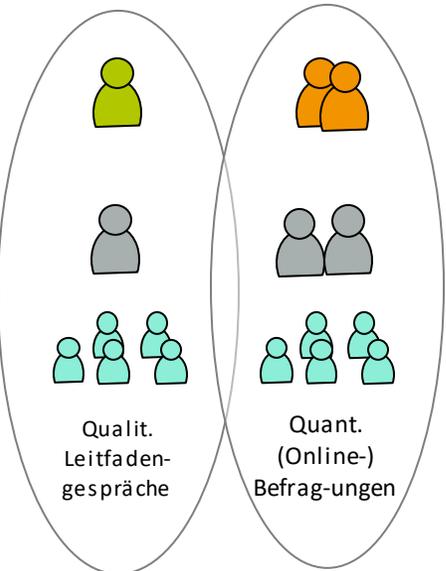
Domäne

Data Curators/
Fachinformations-
mangerInnen

Forschende mit
Domänenexpertise

Verteilung der Domänen und deren Rollen
auf die Forschungs-organisationen

(Data Responsibles für alle Domänen pro
Forschungsorganisation)



Glossar

■ Dateien

Die eigentlichen Daten werden in der Regel als Dateien verarbeitet und abgelegt. *Einem Datensatz sind eine oder mehrere Dateien zugeordnet.* Diese können nach folgenden Datenarten klassifiziert werden:

- Rohdaten
- Bereinigte Daten
- Aggregierte Daten (z.B. Variablenreport)
- Inhaltsbeschreibende Metadaten (z.B. Interviewleitfaden, Anschreiben, Codebuch)

Außerdem haben Sie einen bestimmten Dateityp (bspw. CSV, PDF oder RTF).

■ Datensatz

Ein Datensatz ist eine *Gruppe von inhaltlich zusammenhängenden Datenfeldern (bib. Metadaten) und Dateien*, die im Rahmen eines Forschungsvorgehens entstanden sind. Im Rahmen eines Arbeitspaketes eines Forschungsprojektes können bspw. ein oder zwei Datensätze entstehen (bspw. Befragung IAO/IAT-Mitarbeiter zu Anforderungen an Forschungsdatenmanagement: Je ein Datensatz zu den Zielgruppen wiss. Mitarbeiter und Dienstleistungsteams).

Glossar

■ Domäne

Logische Gruppierung bestimmter Forschungsdaten, die mit einer bestimmten Methode erhoben wurden (z.B. Fragebögen) oder in einem bestimmten Datenformat gespeichert sind (bspw. Sensor-Geodaten mit Zeitbezug).

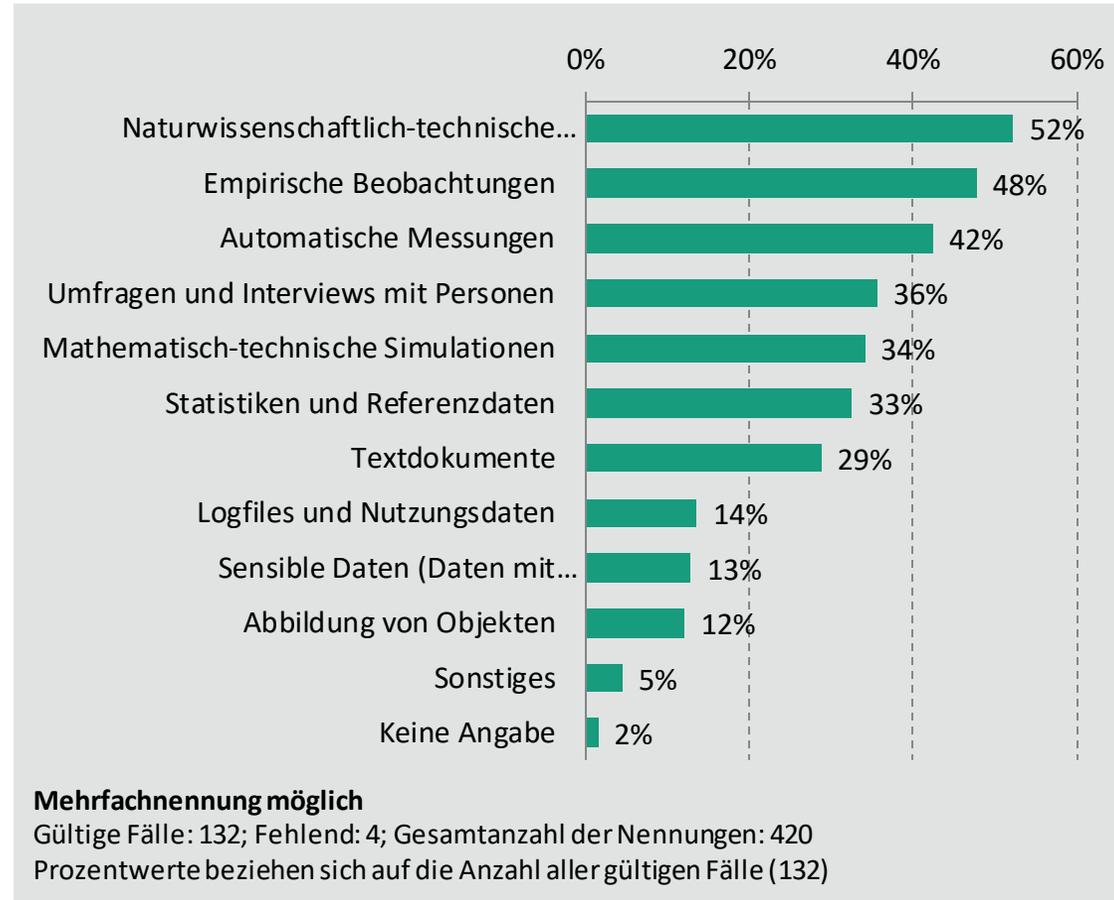
■ Metadaten

Daten, die Informationen über andere Daten enthalten. Folgende Arten von Metadaten unterscheiden wir in HEFE:

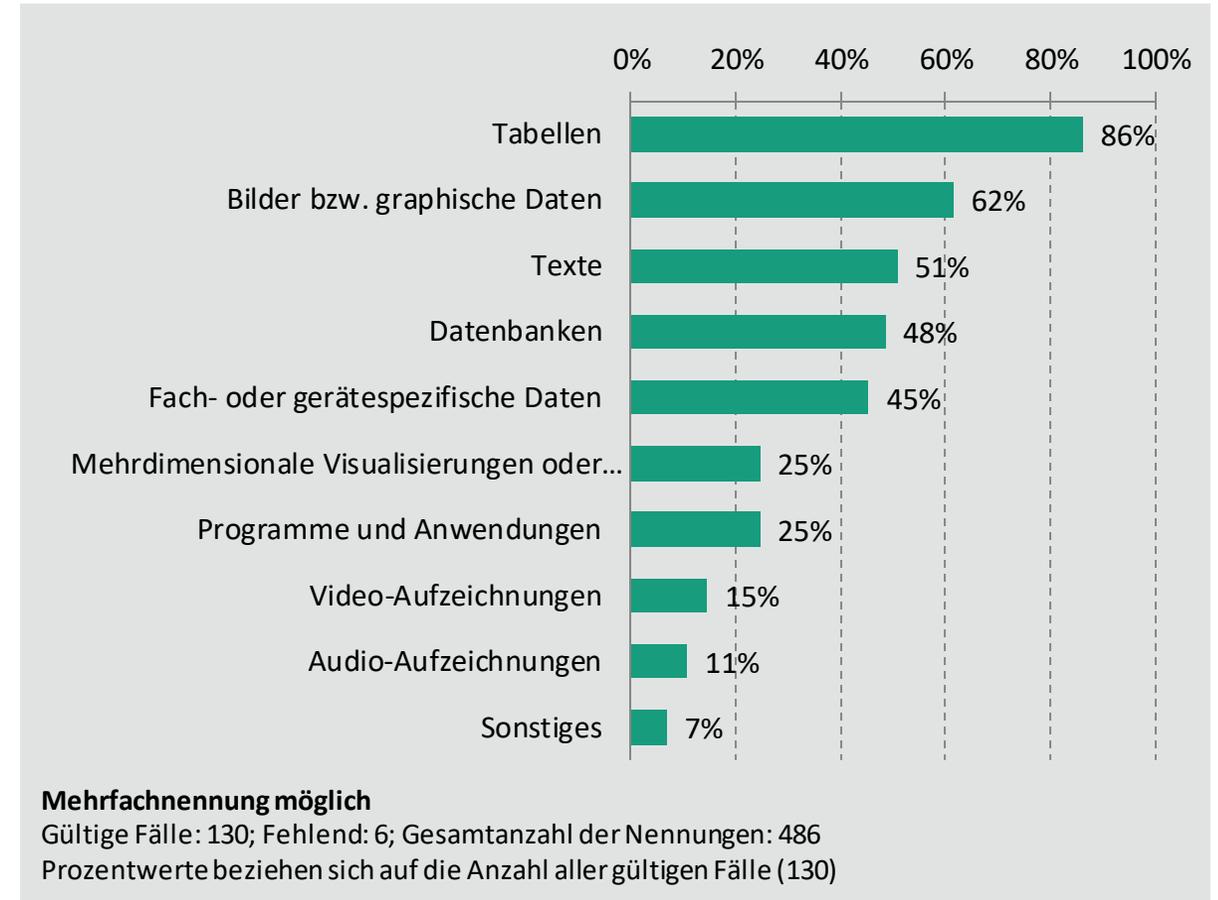
- **Bibliographische Metadaten:** Daten zum Zweck der Identifikation (bspw. Titel, Abstract/Zusammenfassung, Autor und Schlagwörter) sowie Methodische Texte, die auf Ebene des Datensatzes festgehalten werden. Informationen zu Entstehung, Erhebungszeitraum und Beschaffenheit der Daten.
- **Inhaltsbeschreibende Metadaten:** Merkmale, Struktur und Versionierung strukturierter Daten (bspw. Spalten einer Tabelle), die auf Ebene der Dateien entweder beschrieben werden (bspw. Dateityp, Rohdatum vs. Aufbereitete Daten) oder in Dateien enthalten sind.

Forschungsdaten bei Fraunhofer¹

Quellen



Typen



Basis: Wissenschaftler/in, die Forschungsdaten erheben (136)

¹Quelle: Umfrage im Projekt Fordatis,
Andrea Wuchner, Fraunhofer IRB 2017