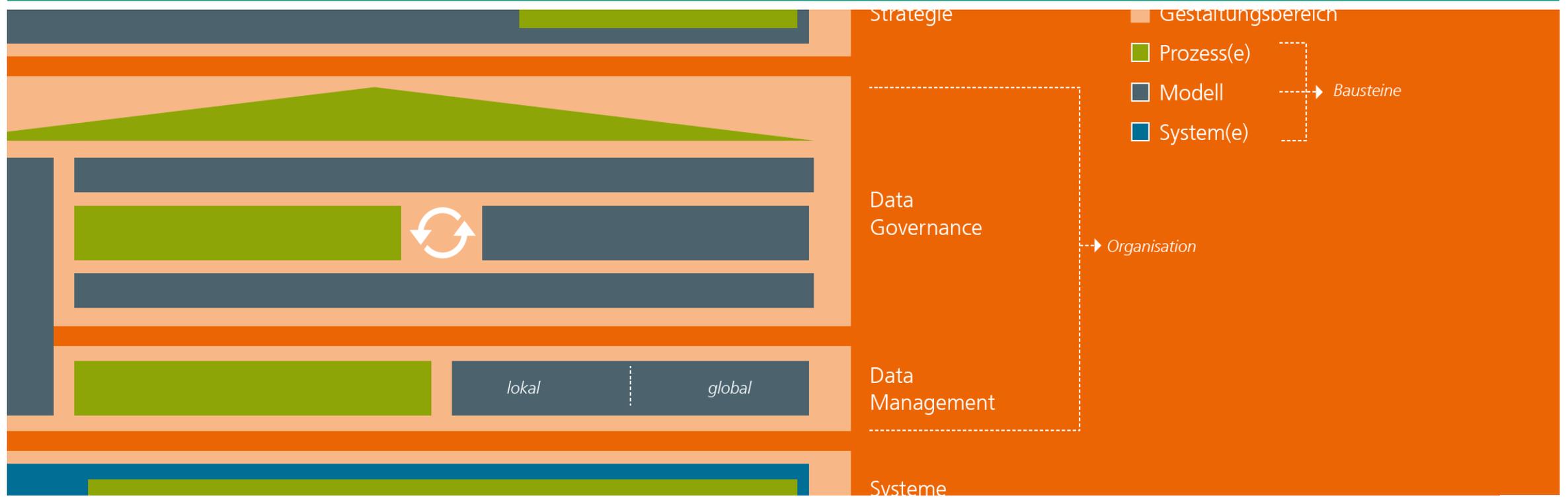


DATA GOVERNANCE FÜR FORSCHUNGSDATEN DER ANGEWANDTEN STADTFORSCHUNG

Johannes Sautter, Erica Boudjio, Rebecca Litauer, Fraunhofer IAO, Stuttgart
Andrea Wuchner, M.A., Fraunhofer IRB, Stuttgart

Universität Siegen, 04. April 2019,
DINI/nestor-Workshop „Strukturen
entwickeln: Organisation und
Governance für lokale FDM-Services“



IAO und IAT im Profil

Fraunhofer-Institut für Arbeitswirtschaft und Organisation IAO
und Institut für Arbeitswissenschaft und Technologiemanagement der Universität Stuttgart



- **Gründungsjahr:** IAO – 1981
IAT – 1991
- **Institutsleitung:** Prof. Dr.-Ing. Prof. e. h. Wilhelm Bauer (geschäftsführend)

Univ.-Prof. Dr.-Ing. Oliver Riedel

Univ.-Prof. Dr.-Ing. Dr.-Ing. E.h. Dr. h.c. Dieter Spath
- **Finanzvolumen:** 41,4 Mio €* , davon 34,2% im Auftrag der Wirtschaft
- **Personal:** 618 Mitarbeiterinnen und Mitarbeiter*

* Daten von 2017, inkl. IAT der Universität Stuttgart

www.iao.fraunhofer.de
www.iat.uni-stuttgart.de

Datenmanagement braucht Data Governance

“Management is the decisions you make. Governance is the structure for making them.”¹

BMBF-Projekt HEFE

01.05.2017 – 31.05.2019

Fraunhofer IAO, IRB



Ohne Governance fehlt es dem Datenmanagement an strategischer Richtung und einheitlicher Organisationsstruktur.

Benötigt werden daher im FDM:

- Ein möglichst **minimales** disziplinspezifisches Datenmodell
- Prozesse zum Datenlebenszyklus, die Richtlinien mittels **Plausibilisierung und Reviewinstanzen** umsetzen (CRUD²)
- Angebote, die den Forschern die Arbeit erleichtern, die **Datenmodell und Prozesse eingebaut** haben.

- Prozesskennzahlen zum Monitoring der Prozesse
- Datenqualitätsregeln, die den Datenbestand gegen Datenmodell und Richtlinien prüfen (Kurationskriterien)
- Eine Organisation die Richtlinien, Datenmodell und Prozesse **aktuell hält**



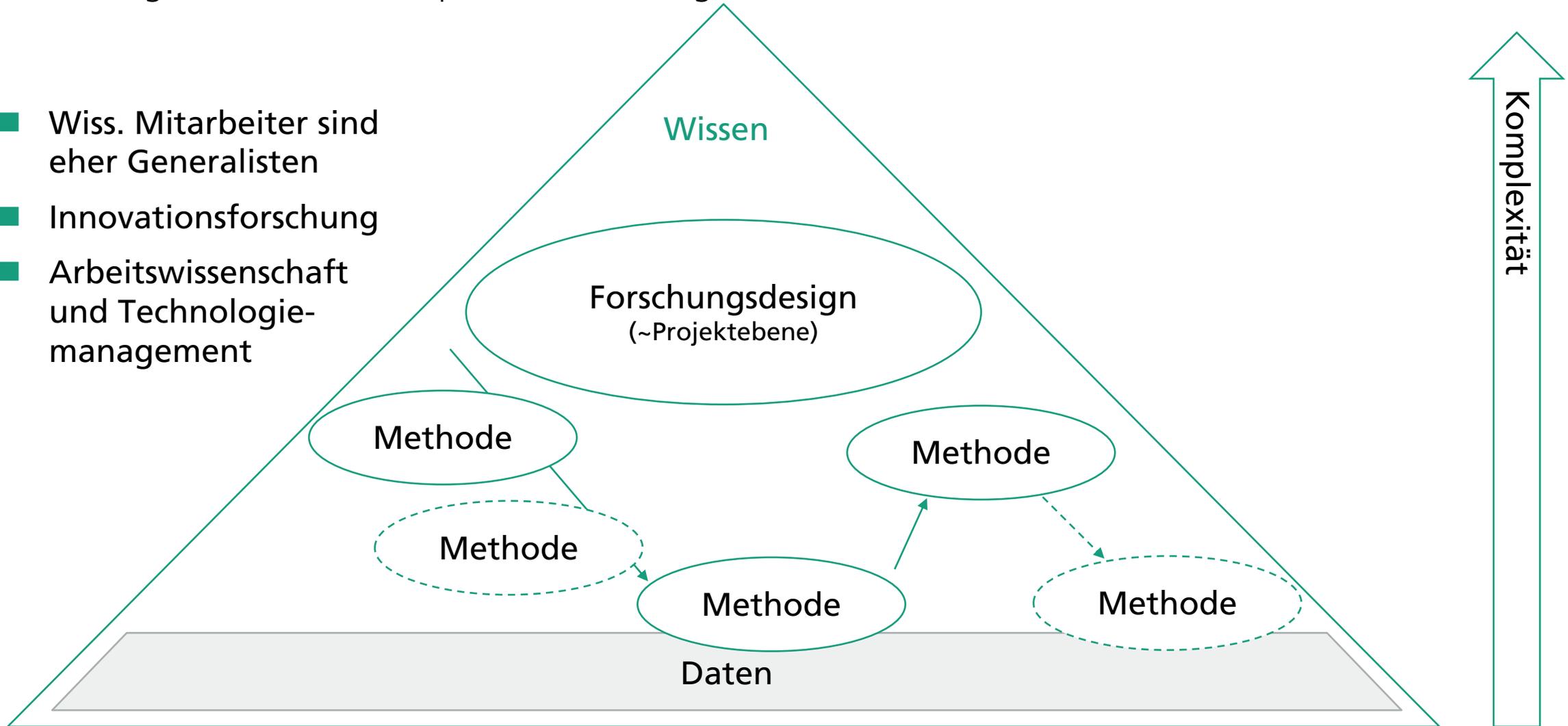
Out-of-Scope für HEFE, aber notwendig für die Zukunft

¹ CIO Magazine ² Create, Read Update and Delete

Methoden-Nutzung an Fraunhofer IAO/IAT der Universität Stuttgart

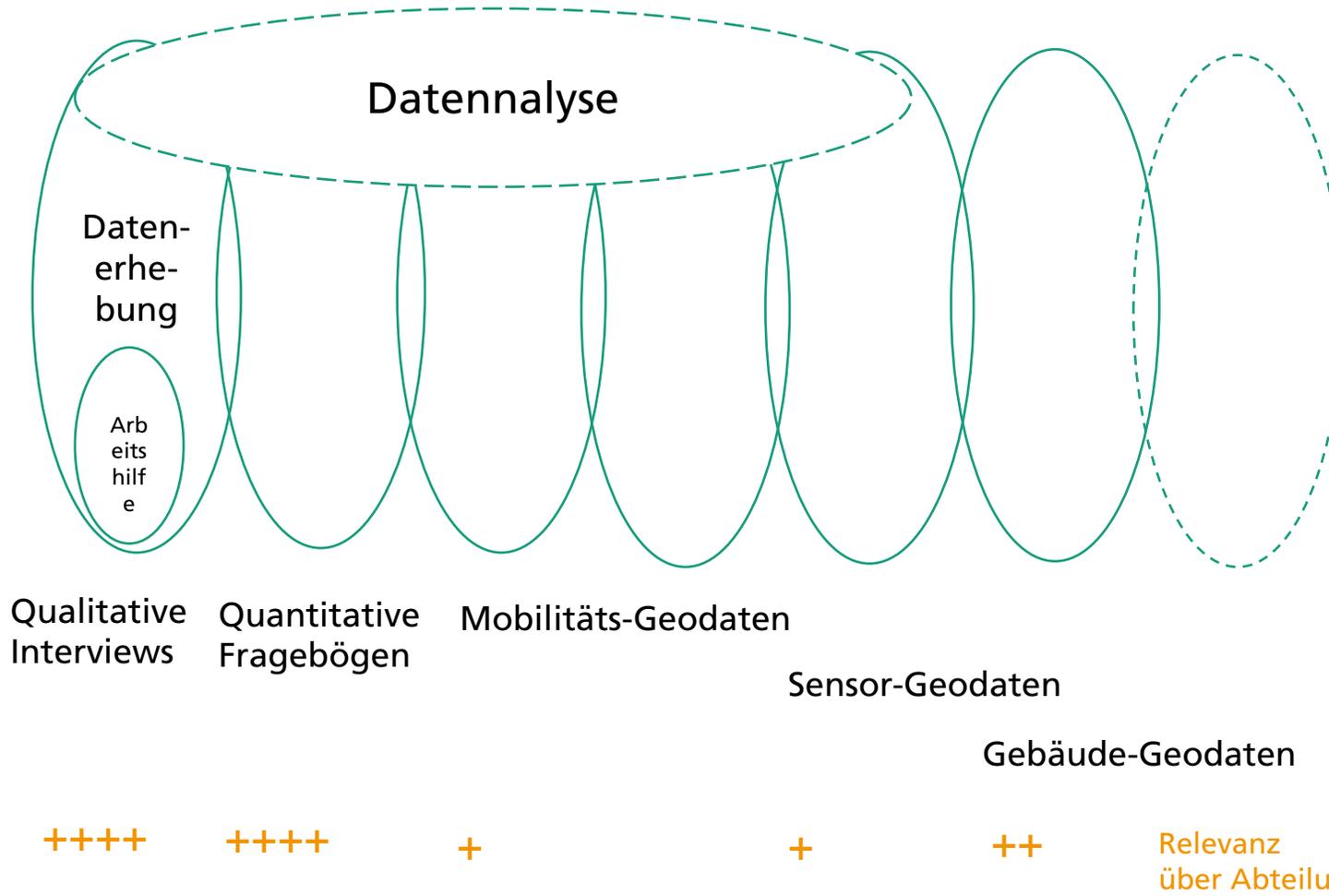
→ heterogene und interdisziplinäre Forschungsdaten

- Wiss. Mitarbeiter sind eher Generalisten
- Innovationsforschung
- Arbeitswissenschaft und Technologie-management



Domänen

Scope Domänenarbeitsgruppen und Arbeitshilfen/Checklisten für Methoden



■ Mögliche Scopes

■ Methoden

■ **Datenerhebung** ← bisher

■ Datenanalyse

■ Fachthema

(~ Logik des Organigramms)

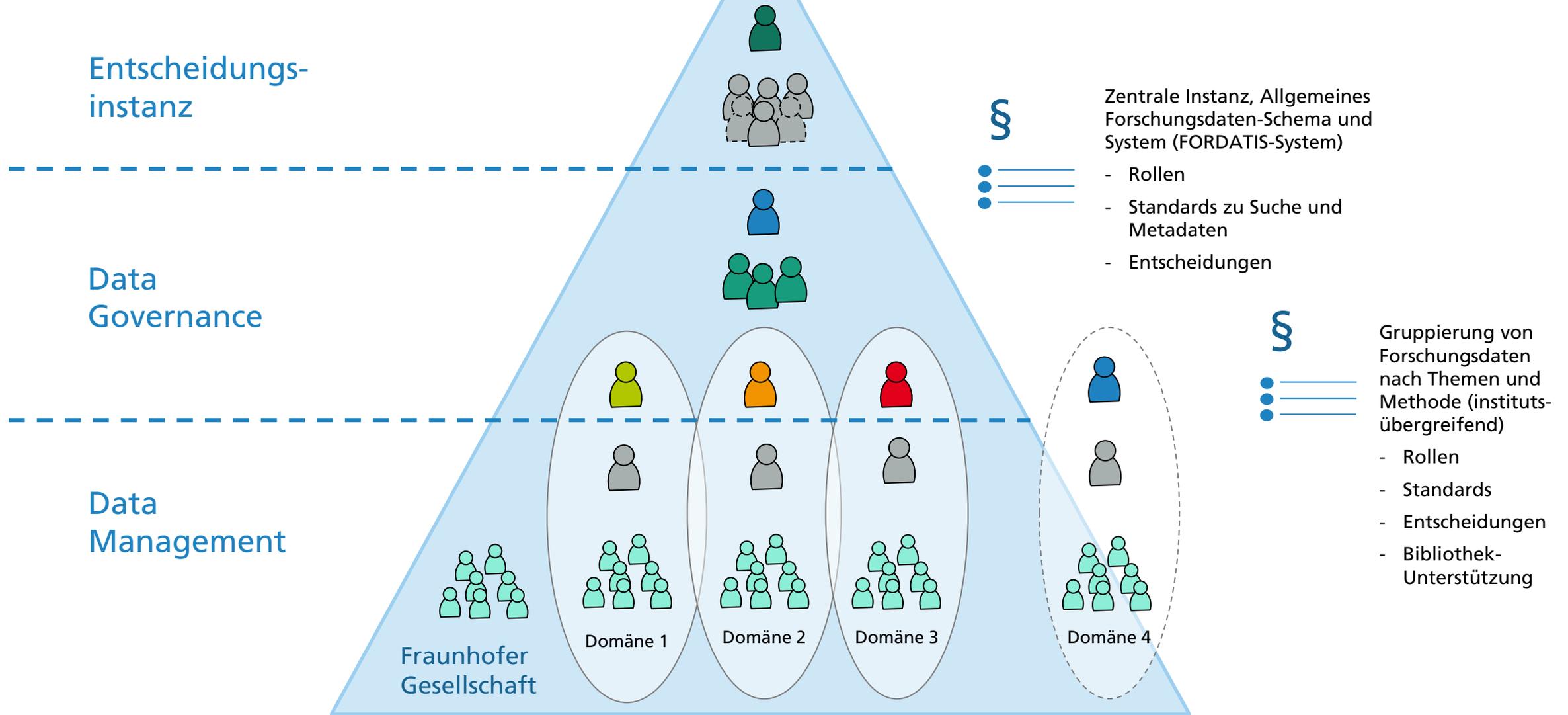
■ Unterschiedlicher Scope

■ Domänen-Arbeitsgruppe

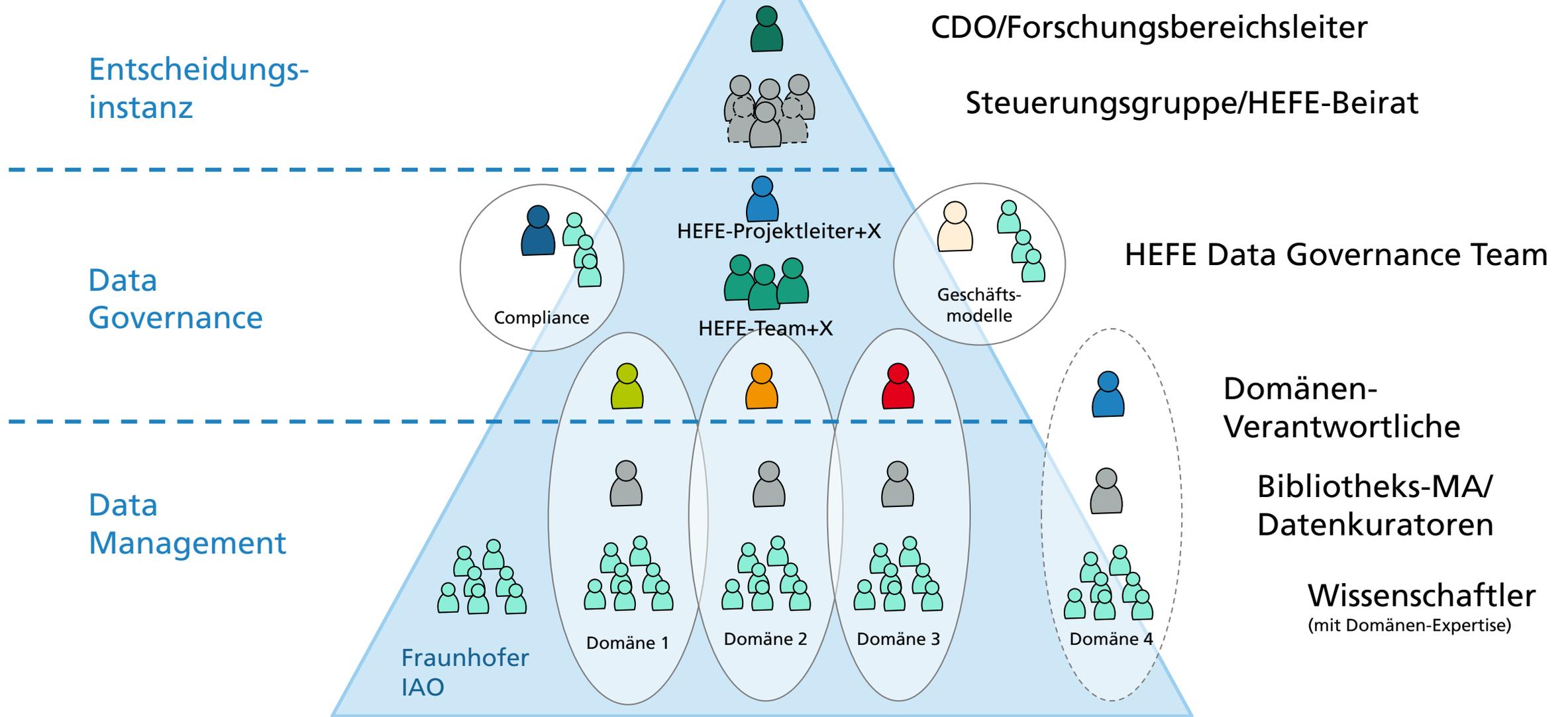
■ Methoden der Domäne

(zugeh. Checklisten/Arbeitshilfen)

HEFE-Konzept: Data Governance in Forschungsdomänen



Data Governance in Forschungsdomänen – Pilot-Projektorganisation



Leitfragen

1. Welche Organisations-, Kooperations- und Service-Strukturen sind sinnvoll? Was sind Best Practices?

- Möglichst viel möglichst zentral lösen (Synergiepotentiale heben)
- Sinnvolle Grenzen ausloten ist Aufgabe der Data Governance
- Wenige spezifische Metadatenfelder bei Domänen

2. Welche Kriterien haben Einfluss auf die Ausrichtung des Angebots?

- **Grad des Spezialistentums**
 - Methodenmanagement zusätzlich zu Datenmanagement
 - Ähnliche Methoden in

untersch. Abteilungen
→ koordinierendes
Domänengremium

3. Wie gelingt die Aktivierung und Einbindung der unterschiedlichen Akteure und Stakeholder?

- Führungskräfte
 - Ausrichtung des FDM an der Institutsstrategie
 - **Daten als Führungsinstrument (Dashboards sind möglich)**
- Mitarbeiter
 - Interne Kommunikation: „Institutsfrühstück für Ausgeschlafene“ & „Café Express“
 - **Jahreszielvereinbarungen**

- „Eigennutz“ durch nützliche Intranetvorlagen (Usability und Dienstleistungs-Tests der Angebote)

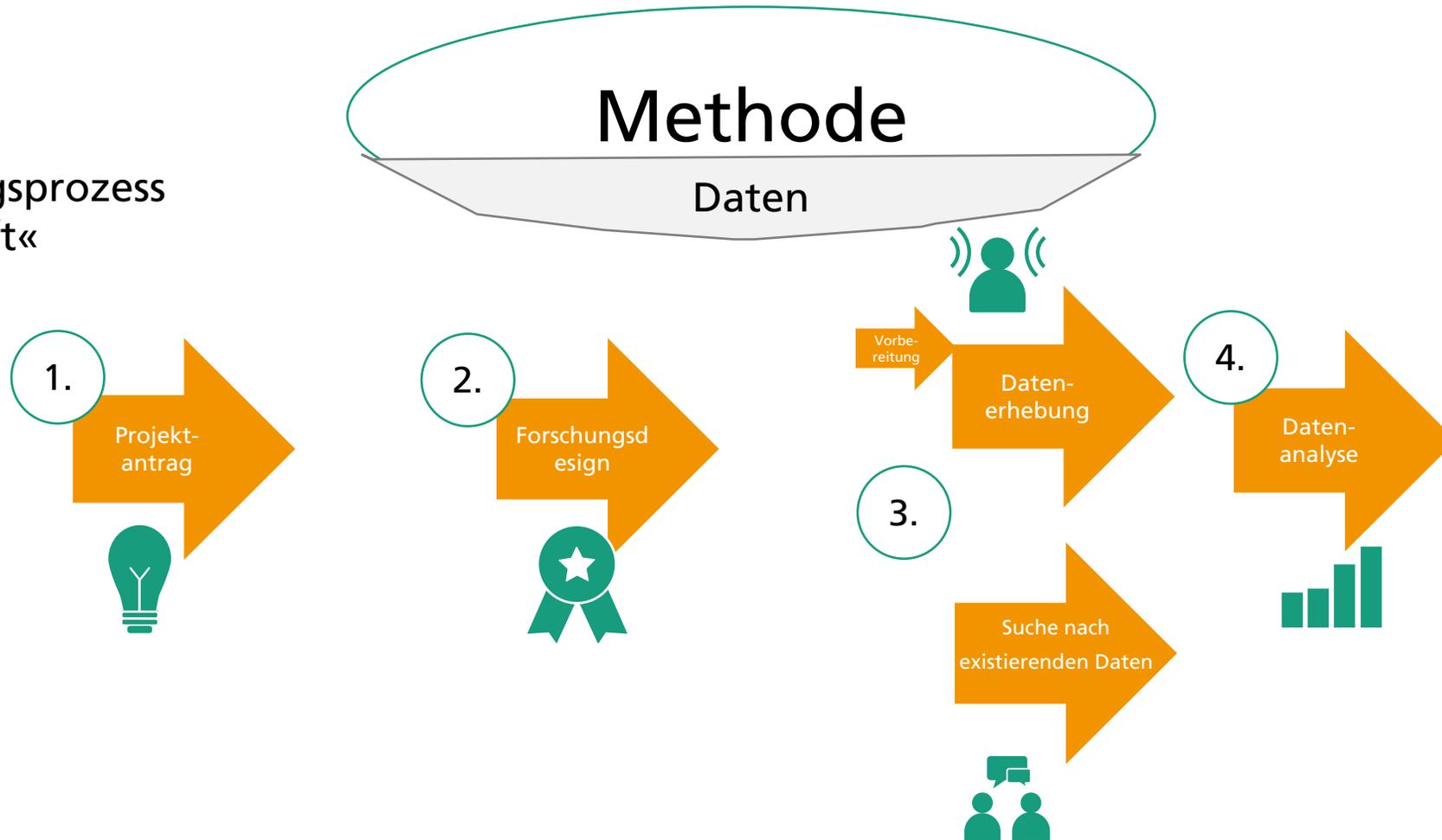
- Fördergeber, Praxispartner: ?

4. Welche Strategien gibt es bezüglich der Nachhaltigkeit von Organisations- und Governance-Formen für FDM-Angebote?

- **Attraktive Rollen als Domänen-Verantwortliche**
- interne Services zu DMPs Beratung zu Methodik, Onlinefragebogen aufsetzen, etc.
 - **Gegen €/PM**
 - Auf Basis Service Level Agreement (SLA)

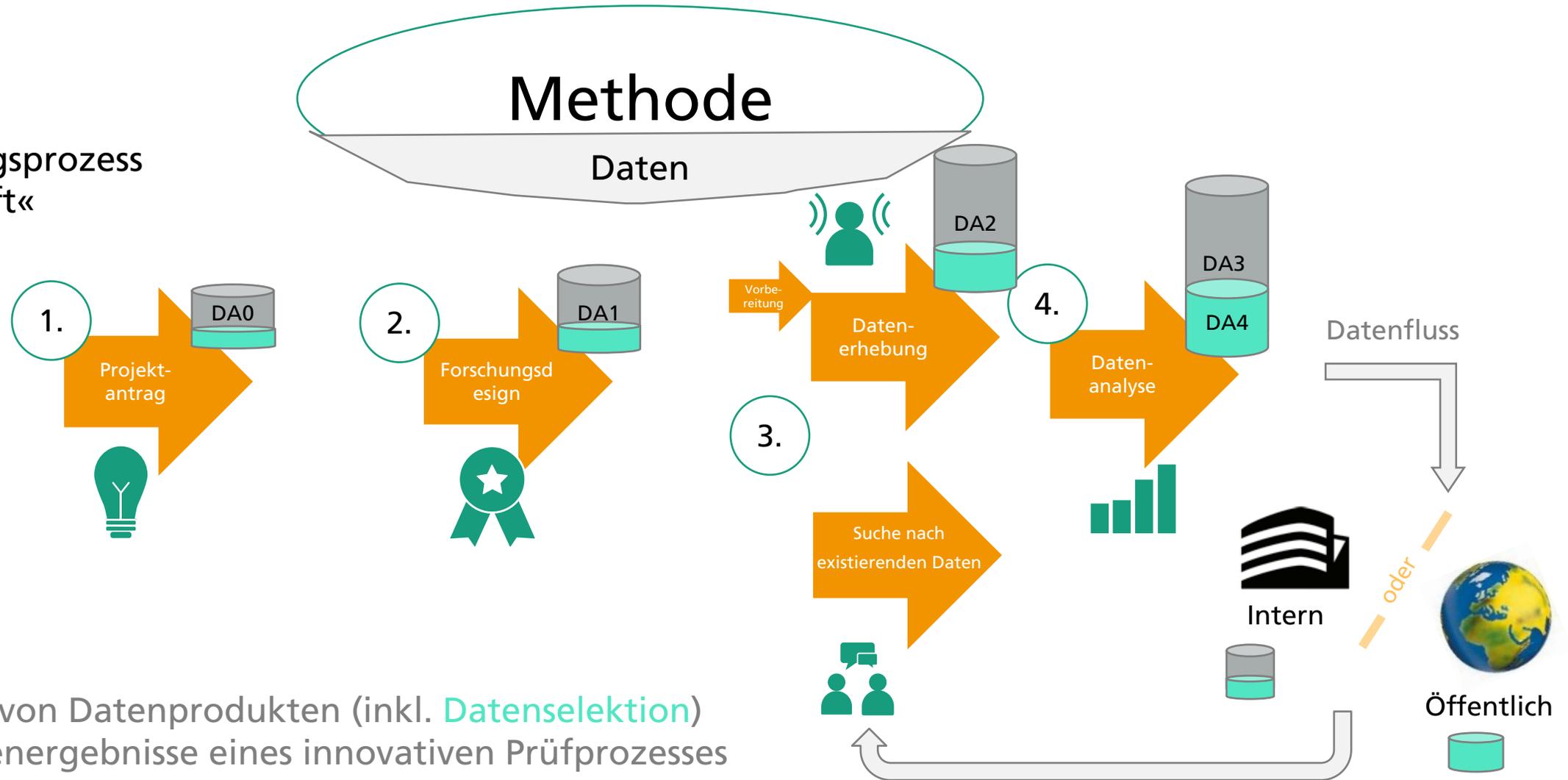
Schritte der Durchführung für eine Datenerhebungsmethode

»Forschungsprozess der Zukunft«



Schritte der Durchführung für eine Datenerhebungsmethode

»Forschungsprozess der Zukunft«



Definition von Datenprodukten (inkl. **Datenselektion**) als Zwischenergebnisse eines innovativen Prüfprozesses

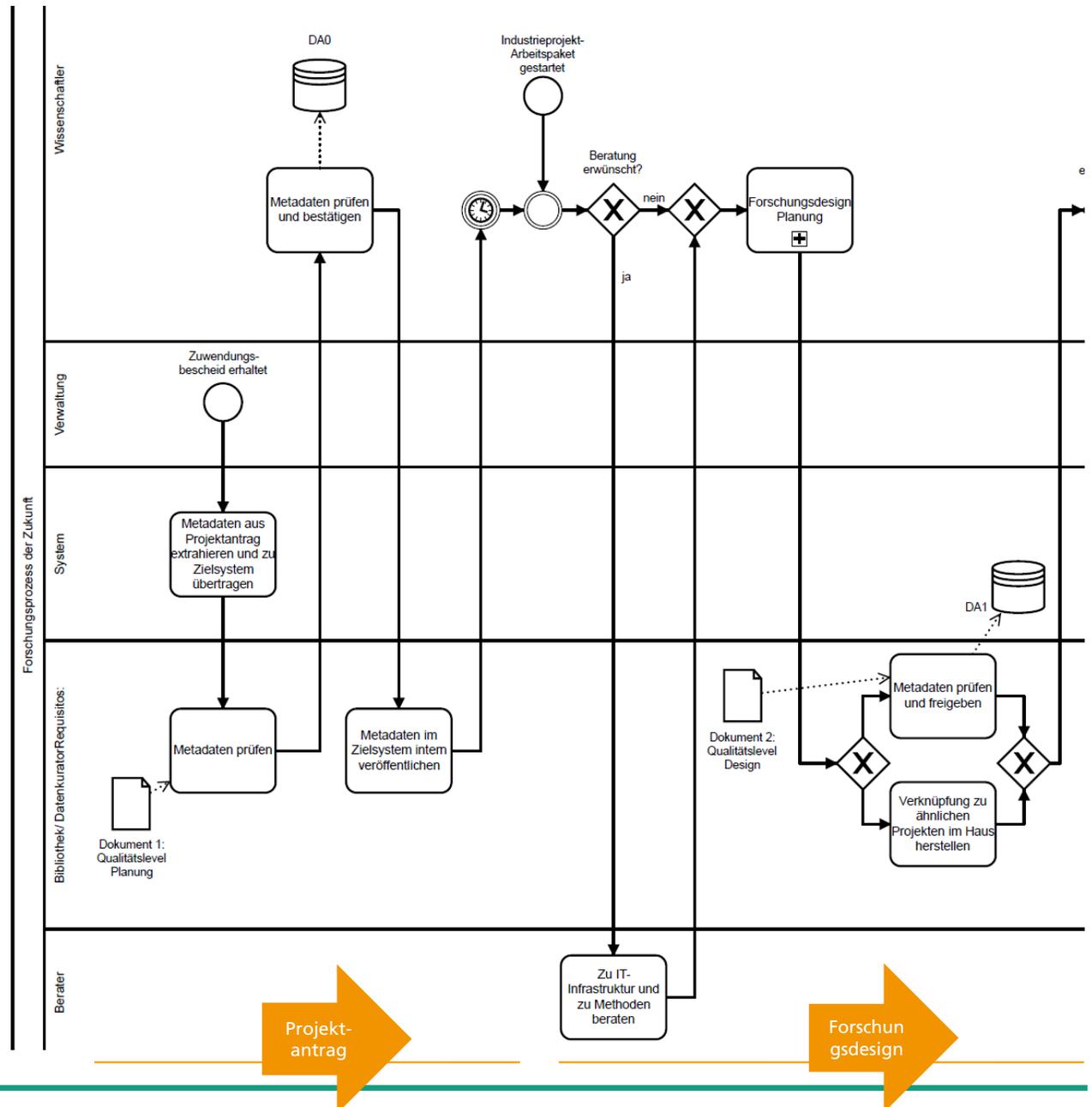
Definition von Datenprodukten

als Zwischenergebnisse eines innovativen Prüfprozesses

(Ausschnitt des Prozesses in BPMN¹)

- Erfassung der genehmigten Projektanträge
 - Interne Plattform für alle mit geplanten Datenerhebungsaktivitäten
 - → Aspekt des Wissensmanagements
- Prüfung durch Datenkuratoren (FIMs²)
 - Regelwerke mit 5 Qualitätsstufen/-spezifika geplant:
 1. Planung
 2. Forschungsdesign
 3. Datenerhebung
 4. Archivierung
 5. Veröffentlichung (existiert, FORDATIS)

¹Business Process Modeling Notation ²FachinformationsmanagerInnen



2. Forschungsdesign – am Beispiel „Qualitatives Leitfadengespräch“



HEFE Intranet-Seite
<http://s.fhg.de/GeG>

Schritt	Aktivität
1	Präzisierung des Forschungsgegenstandes und Generierung einer Forschungsfrage
2	Stimmen die Voraussetzungen für qualitatives Vorgehen mit meinem Erkenntnisinteresse überein
3	Bestimmung der Fallauswahl → Infobox

Alles erledigt? >>> DA1 ausfüllen

Dokumente	Klasse im Datenmodell	Metadaten	Metadaten Wert
-	Projekt	Projekttitel	
		Abkürzung	
		Thema	
		Projektzeitraum	
		Beschreibung	
		Projektmitarbeiter	
	Institut	Institut	
	Fördergeber	Fördergeber	
	Datensatz	Untersuchungseinheit	
		Forschungsfrage	
		Forschungsgegenstand	
		Erhebungsmethode	
		Domäne	Interview

[Datenablagepunkt 1]

Infobox: Fallauswahl (Sampling)

- (1) Vorab Definition: Zugrundeliegendes **Erkenntnisinteresse** über **bestimmte Personengruppen** oder Vermutung gewisser Mechanismen vorhanden, zielgerichtete Antwort auf konkrete Forschungsfrage
- (2) Offene Herangehensweise: Vage Vorstellung von Untersuchungsgebiet, Fokus und **Konkretisierung geschieht explorativ** aus Untersuchung, Ausschaltung möglicher Fehlwahrnehmungen durch Vorgaben

Samplingregel: Möglichst **enge Beschreibung** der Fälle anhand gewisser **Merkmale** (z.B. Alter, Position im Unternehmen oder Berufserfahrung) und gleichzeitig **breite Variation innerhalb** der Gruppe (maximale Unterschiede, „Bildung von Gegenhorizonten“)

Die Fallauswahl determiniert die **Reichweite und Verallgemeinbarkeit** der Erkenntnisse → Suche nach typischen Mustern zur Beantwortung der Forschungsfrage

IT-Systeme

Fachsystem

für Erhebung und Analyse im »System Stadt«

■ Geodatensystem

- arcGIS System
- Etablierte COTS¹
Geoinformationslösung



Intern

■ Metadatensystem

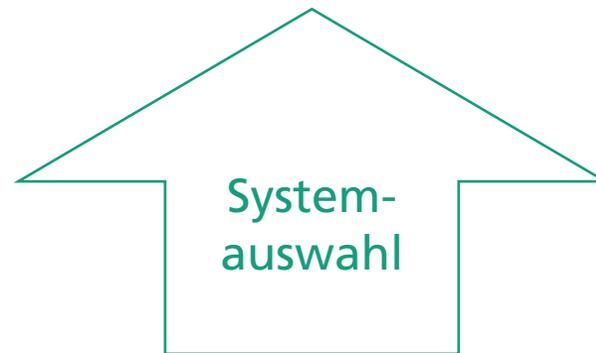
- „Single Version of Truth“
- „heiße Forschungsdaten“



Öffentlich

■ Forschungsdaten-Repository

- System FORDATIS
- „Schaufenster“/„Kalte Forschungsdaten“
- Ab ~Juni 2019
- Basierend auf DSpace Technologie

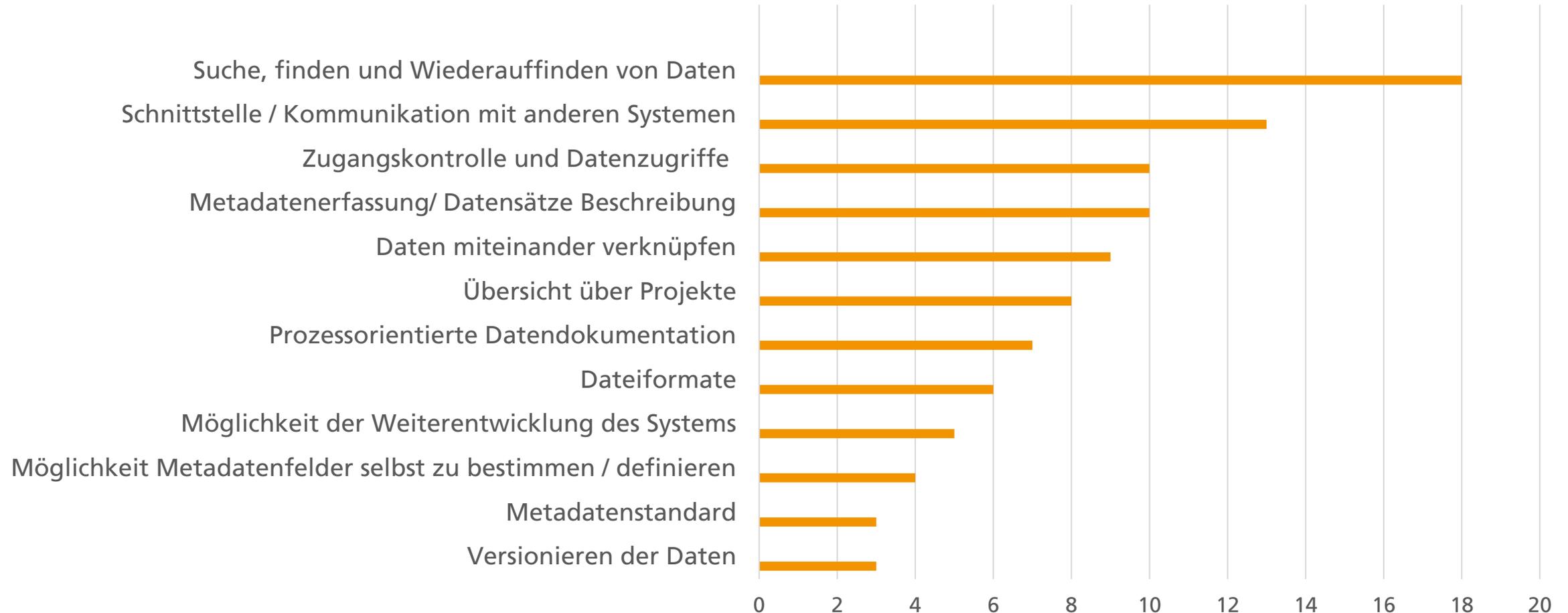


¹Commercial Off the Shelf – proprietäre Standardsoftware

Anforderungen an ein Metadatenystem

Priorisierung aus einem Workshop mit Führungskräften und Wissenschaftler

8 Bewertungspunkte pro Person, Doppelbewertung erlaubt, 12 Bewertende, davon 5 Führungskräfte

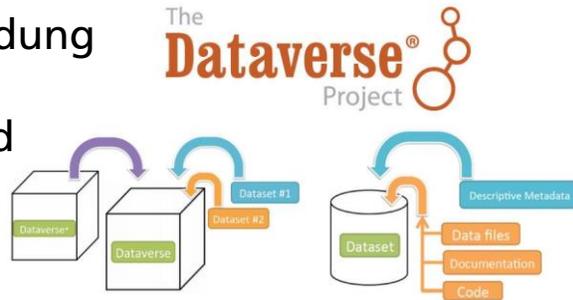


PROJEKT HEFE

POTENTIELLE METADATENSYSTEME

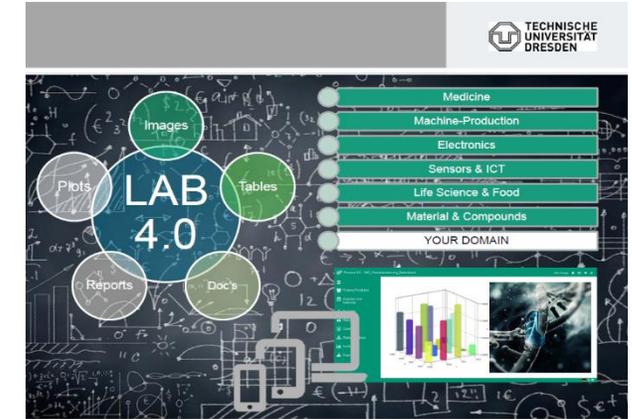
Dataverse

Open-Source-Webanwendung zum Teilen, Speichern, Zitieren, Untersuchen und Analysieren von Forschungsdaten.



Lab 4.0

Datenbanksystem für Forschungsdaten im Fraunhofer Datenraum



DSpace

Open-Source Software zur Erfassung, Speicherung und Weiterverbreitung von digitalen Ressourcen



Colectica Designer

DDI-Metadateneditor und Fragebogen-Designer, zur Dokumentation von Datensätzen und Datenerfassung. Nutzung zusammen mit Colectica Repository und Colectica Portal

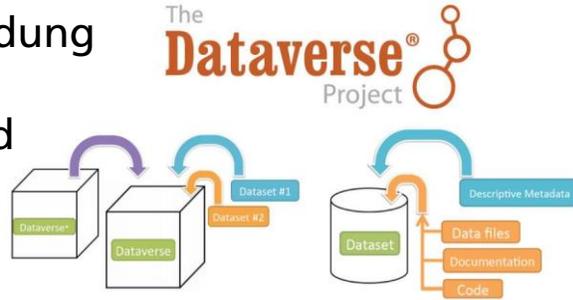


PROJEKT HEFE

POTENTIELLE METADATENSYSTEME

Dataverse

Open-Source-Webanwendung zum Teilen, Speichern, Zitieren, Untersuchen und Analysieren von Forschungsdaten.



Lab 4.0

Datenbanksystem für Forschungsdaten im Fraunhofer Datenraum



DSpace

Open-Source Software zur Erfassung, Speicherung und Weiterverbreitung von digitalen Ressourcen



Colectica Designer

DDI-Metadateneditor und Fragebogen-Designer, zur Dokumentation von Datensätzen und Datenerfassung. Nutzung zusammen mit Colectica Repository und Colectica Portal

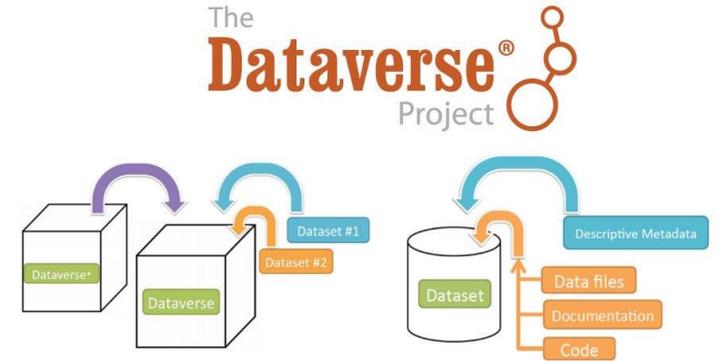


Dataverse

Beurteilung auf Basis der Anforderungen mittels Test und Dokumentation

■ Vorteile

- Unterstützt die Nutzung von Disziplinspezifischen Metadatenstandards
- Möglichkeit Pflichtfelder selbst zu bestimmen
- Möglichkeit ausgefüllte Metadatenfelder als Template für weitere Datensätze zu nutzen
- Möglichkeit eigene Suchmaske zu konfigurieren und Datensätze nach bestimmten Kriterien zu filtern
- Klassifikation der Daten nach der Dewey-Dezimalklassifikation
- Ermöglicht das Suchen nach Metadaten
- Unterstützt das Hochladen von Daten aus unterschiedlichen Formate
- Export von Daten in unterschiedliche Formate
- Hat eine Rollenbasierte Zugriffssteuerung
- Ermöglicht Datensätze miteinander zu verknüpfen
- Bietet Schnittstellen und Protokollen an, für die Kommunikation mit anderen Systemen

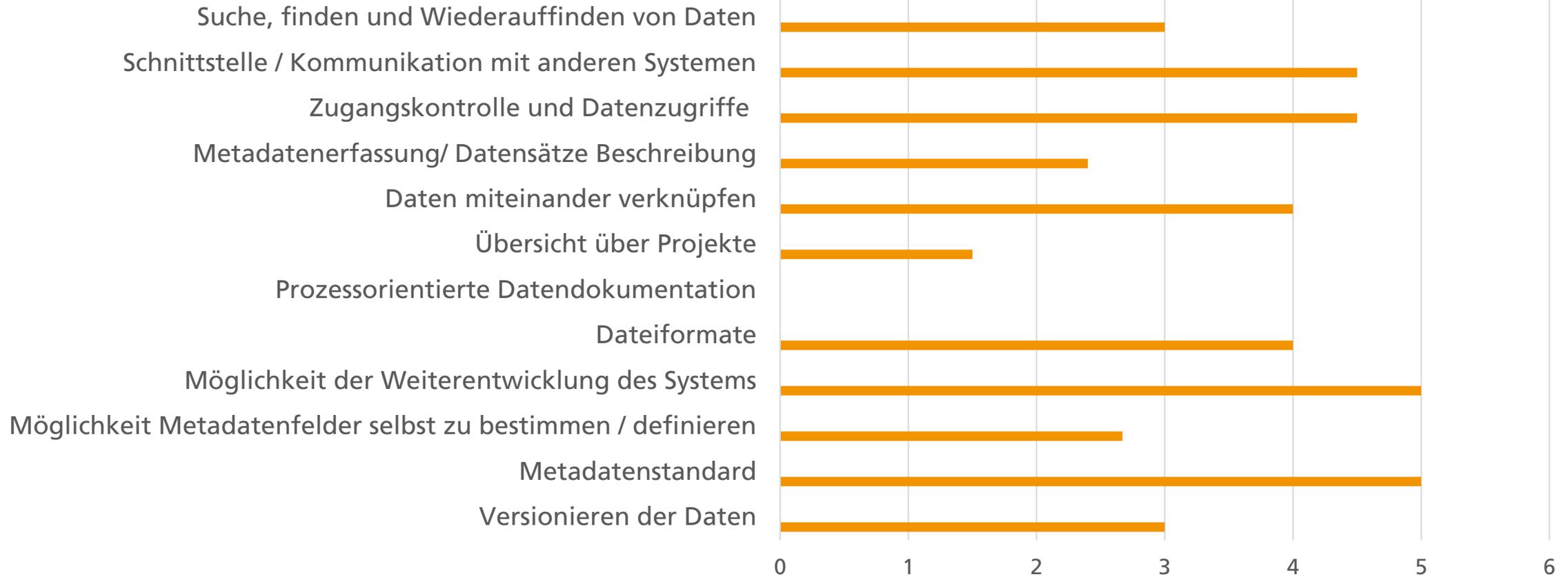
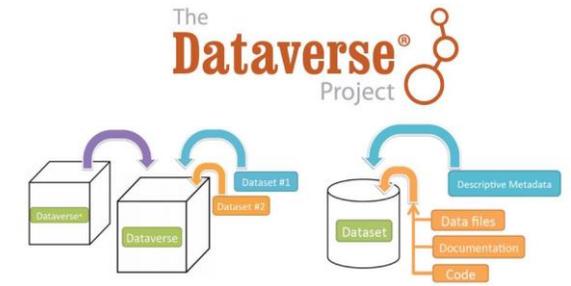


■ Nachteile

- Keine Unterstützung von Prozessorientierter Datenablage
- Keine automatisches Ausfüllen von Metadatenfeldern
- keine Projektübersicht
- Benutzerdefinierte Metadatenfelder werden nicht unterstützt
- Volltextsuche auf Daten nicht möglich

Dataverse – Bewertung

Basierend auf den Anforderungen des Fraunhofer IAO an das System
(Maximal 5 Bewertungspunkte pro Anforderung)



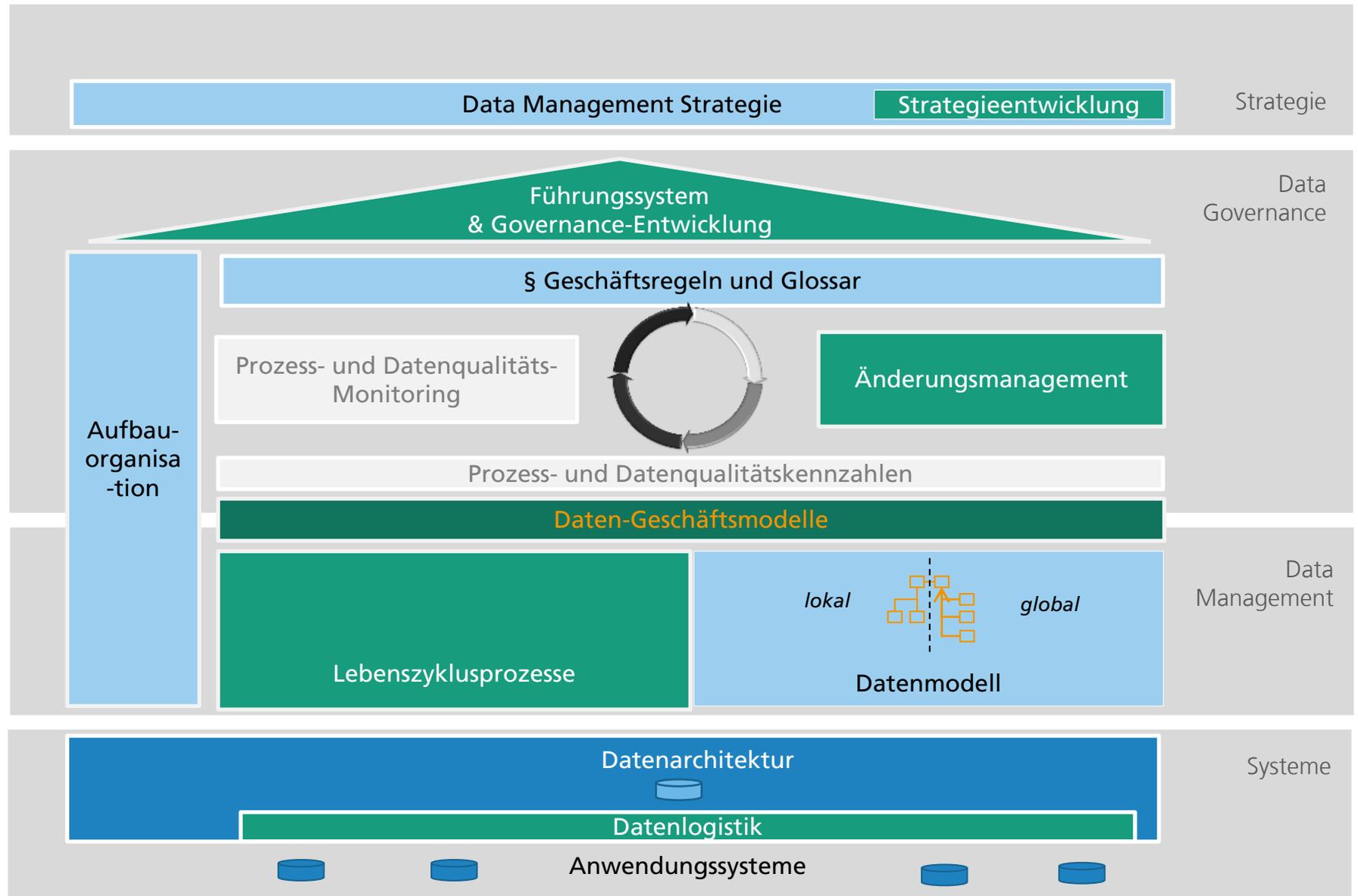
Data Governance für Forschungsdaten

HEFE-Ziele

- Strategische Verankerung
- Nachhaltiger Datenlebenszyklus
- Langfristiges Steuern (noch nicht in Hefe)

Bausteine

- Gestaltungsbereich
- Out-of-Scope für HEFE, aber notwendig für die Zukunft
- Prozess(e)
- Modell
- System(e)



Herzlichen Dank für Ihre Aufmerksamkeit!

Fragen, Anmerkungen, Kontakt?

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



Erica Boudjio Dongmeza
Fraunhofer IAO
Erica-Paradis.Boudjio-
Dongmeza@iao.fraunhofer.de



&

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

Rebecca Litauer
Fraunhofer IAO
Rebecca.Litauer@iao.fraunhofer.de



Nobelstraße 12, D-70569 Stuttgart

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



BACKUP-FOLIEN

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
- Metadaten (Inhaltsbeschreibende Metadaten).

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

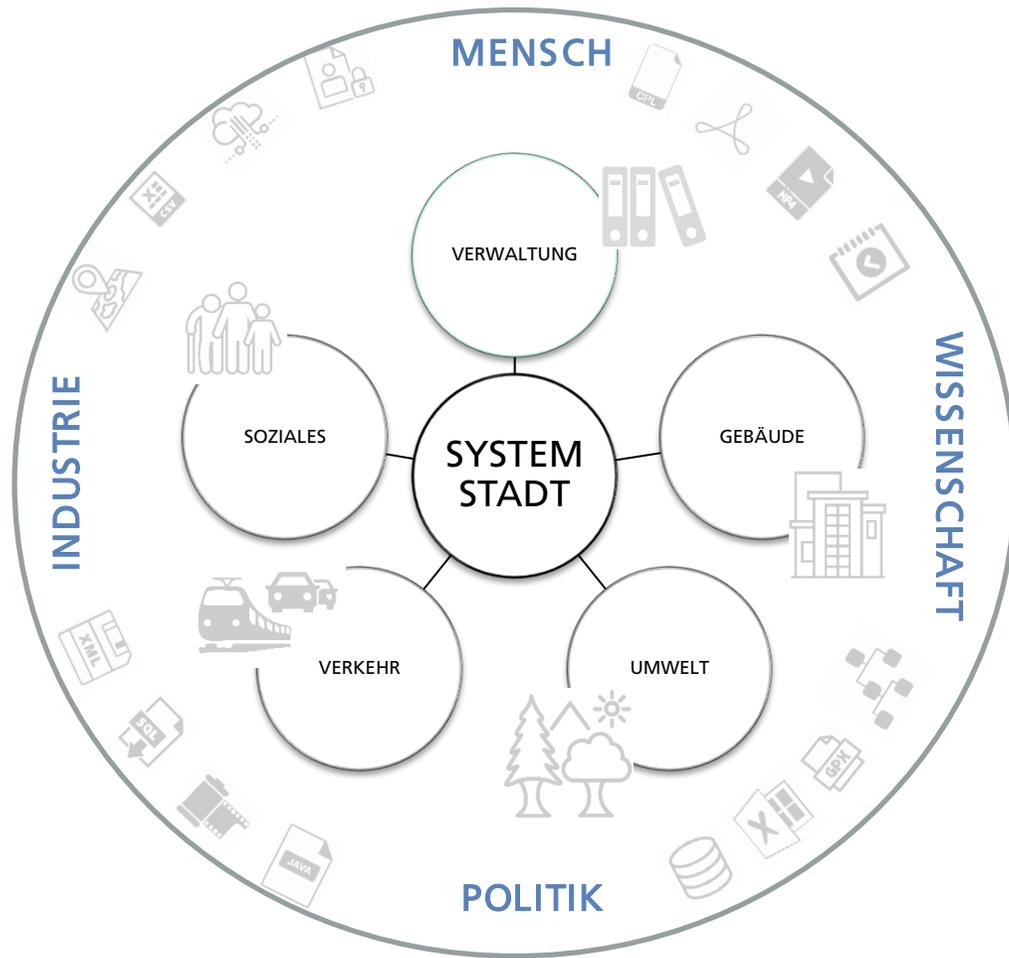
Daten, die Informationen über andere Daten enthalten. Folgende Arten von Metadaten unterscheiden wir in HEFE:

■ **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.

Angewandte Forschung

HEFE-Pilotbereich: Analyse-, Planungs- und Echtzeitdaten im System Stadt



- Definition Angewandte Forschung:
 - Konkrete methodische Problemlösungen
 - Direktaufträge als zentrales Ziel
 - Enge Kooperation mit Praxisorganisationen
 - Nutzung der Ergebnisse von Grundlagenforschung mehrerer Disziplinen
 - Stand der Forschung und Stand der Praxis relevant¹

¹Sautter, Johannes: Stand der Praxis, anwendungsorientierte Konzepte und Expertensprache als Schlüsselaspekte guter angewandter Sicherheitsforschung, in Gabel, Friedrich; Eckert, Judith; Drews, Patrick; Hieronymus, Jan (Hrsg.):

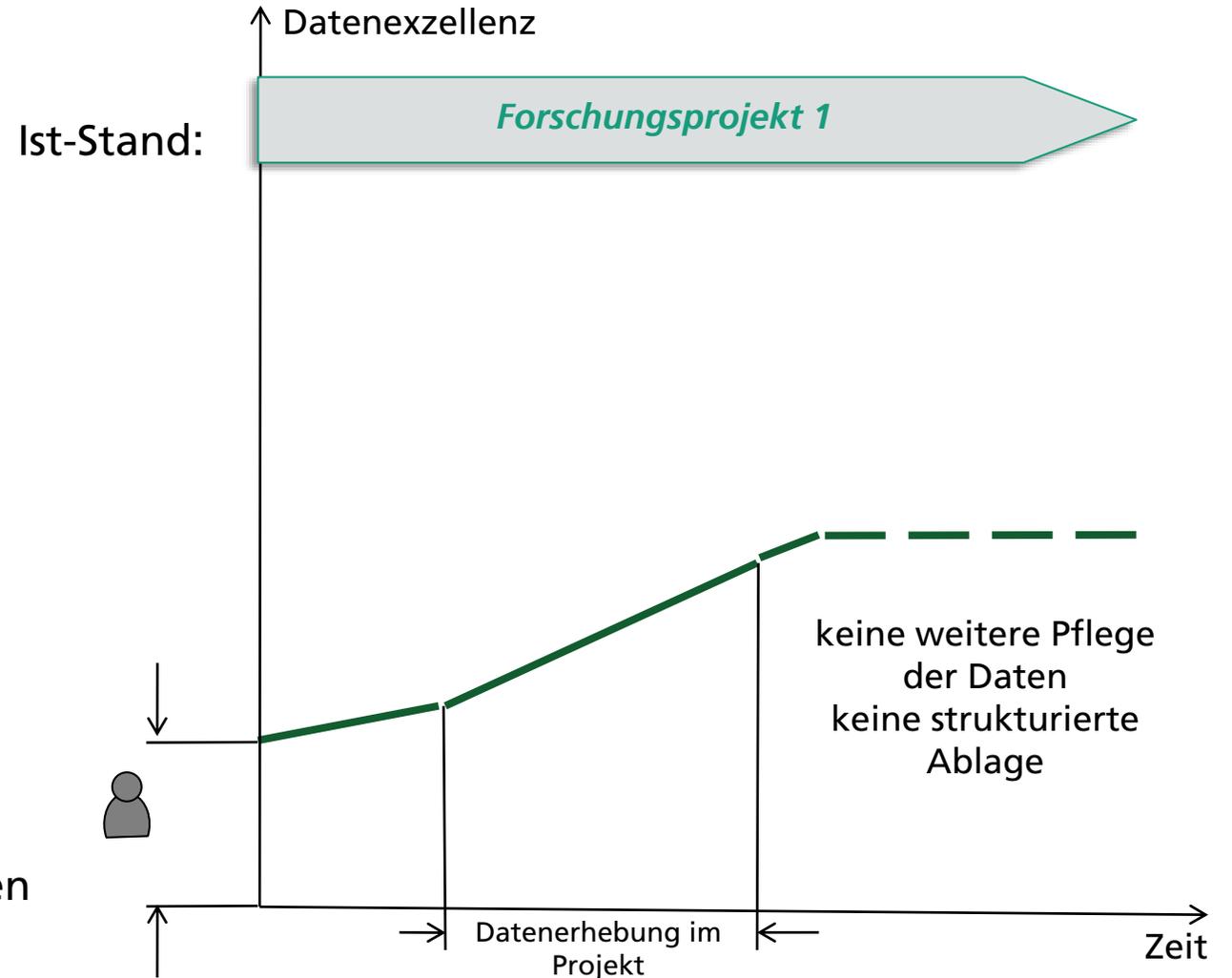
Was ist gute zivile Sicherheitsforschung? : Zweiter Workshop des Graduierten-Netzwerks "Zivile Sicherheit" zum Thema: "Was ist gute zivile Sicherheitsforschung?" vom 25./26. November 2016 in Stuttgart; Ergebnisbericht

Stuttgart, 2017, pp.46-48.

Projekt HEFE: Heterogene Forschungsdaten im Stadtkontext: Governance für Analyse-, Planungs- und Echtzeitdaten

Projektziele:

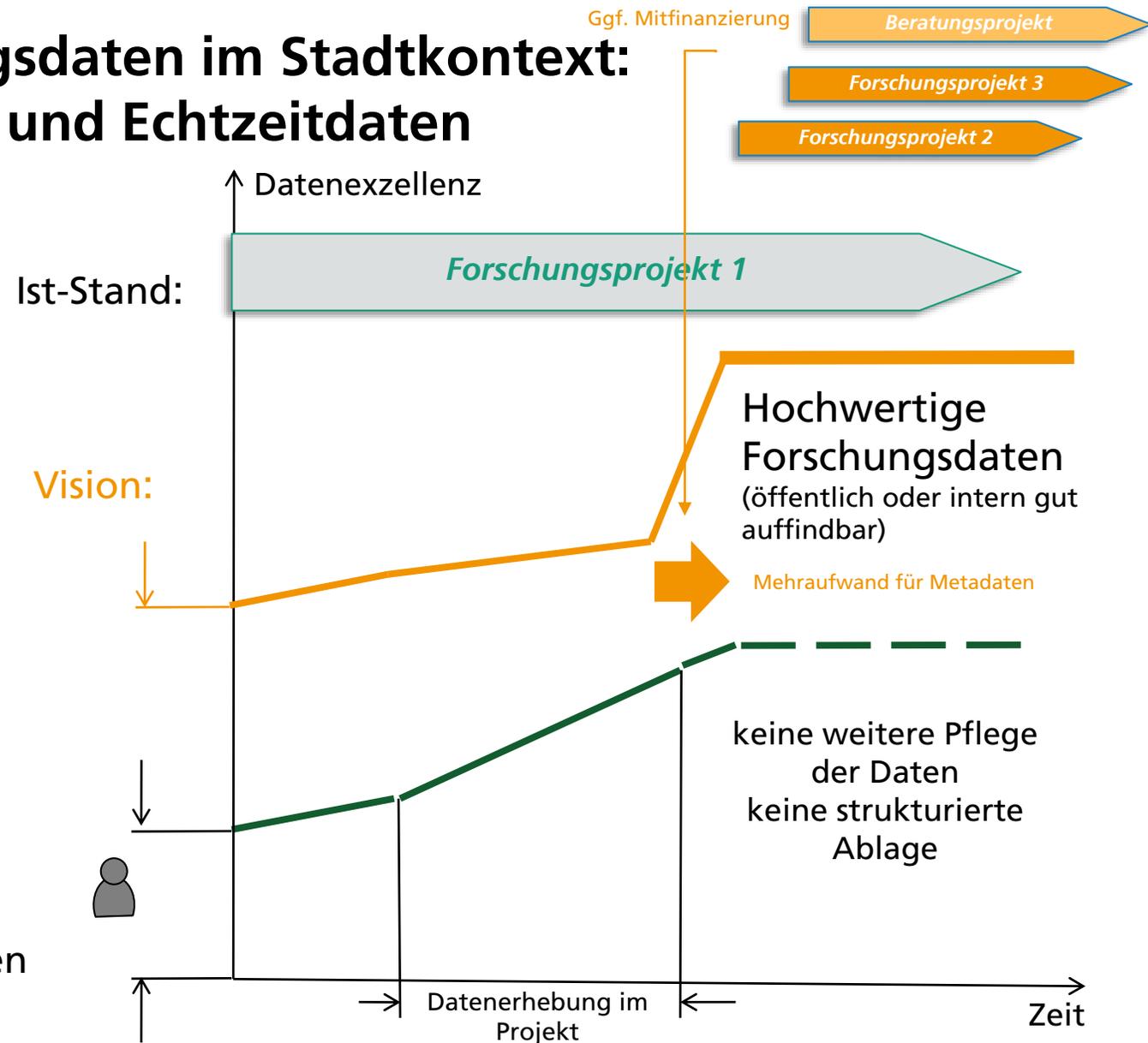
- Strategie und Organisation für Forschungsdatenmanagement bei Fraunhofer
- Data Governance für Forschungsdaten
 - Rollen und Rechtenkonzept
 - Lebenszyklusprozesse
 - Datenmodell
- Standard-Bausteine für den Forschungsprozess
 - Vorlagen
 - Checklisten
 - Datenprüfung
- Geschäftsmodelle für städtische Forschungsdaten



Projekt HEFE: Heterogene Forschungsdaten im Stadtkontext: Governance für Analyse-, Planungs- und Echtzeitdaten

Projektziele:

- Strategie und Organisation für Forschungsdatenmanagement bei Fraunhofer
- Data Governance für Forschungsdaten
 - Rollen und Rechtekonzept
 - Lebenszyklusprozesse
 - Datenmodell
- Standard-Bausteine für den Forschungsprozess
 - Vorlagen
 - Checklisten
 - Datenprüfung
- Geschäftsmodelle für städtische Forschungsdaten



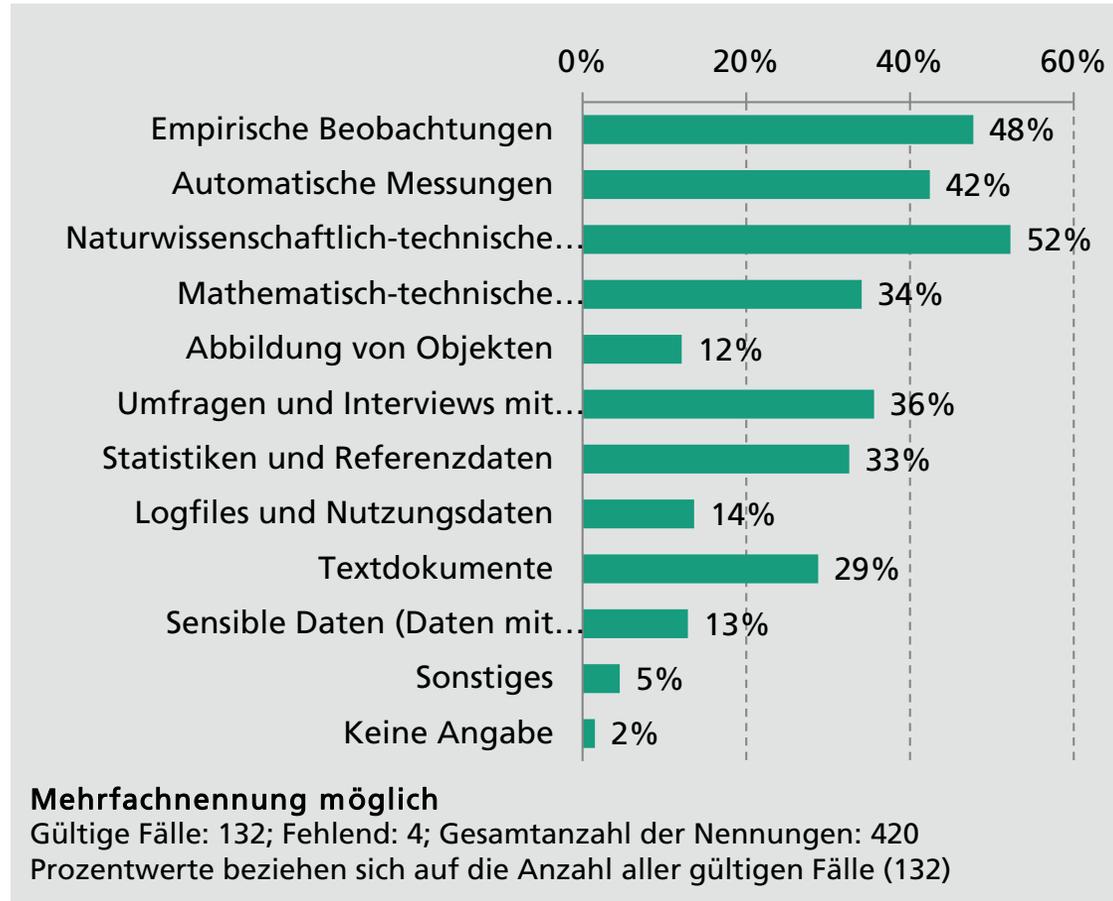
Identifizierte Datenexzellenz-Herausforderungen am IAT/IAO+X

Empirie, Abgeleitet

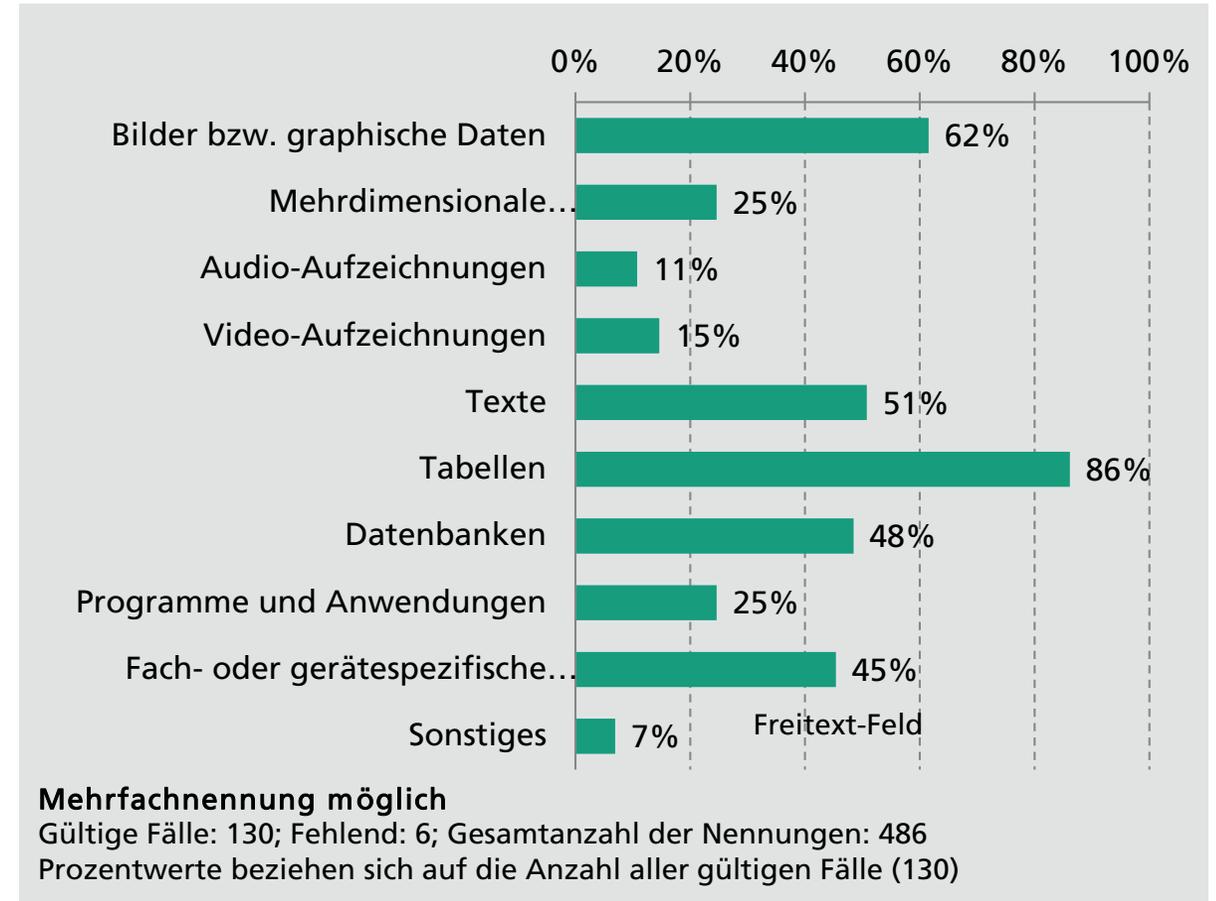
Main Challenges	Operational Excellence Challenges	Legal Challenges	Data Management Process Challenges	Data Quality Challenges
Internal competition	Not existing exchange post-usage hinders research excellence	–	Competitive thinking between scientists	Lack of accessibility for excellent research
No internal exchange	No efficient collaboration possible	–	Lack of communication between scientists within research field	Lack of representational data quality and accessibility data quality
Effort-benefit balance	Value for effort in metadata structure is not seen	–	Easy and low-effort processes for metadata standards are missing	Lack of representational and intrinsic data quality
Inattention on duties to funder	No provision of data to research community	Lack of knowledge on funder compliance	Not sufficient knowledge management	–
No process standards	Scientists collect their data in different ways	–	Lack of tools and processes guaranteeing more standardized collection of data	Insufficient standardization of data and meta data (representational data quality)
No data standards	There is no minimum standard for metadata and data storage	Ensure that institute/community standard complies to legal requirements	Lack of process for agreement of scientists and metadata standard coordination	Lack of representational consistency
Shortage of scientist development	Little methodological expertise	–	Provision of little methodological support in form of workshops/consultancy services	Lack of data quality in all dimensions

Forschungsdaten bei Fraunhofer¹

Quellen



Typen



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