



# FDO4DE

FAIRe Digitale Objekte für Deutschland:  
Universaler Datenstandard  
und hierarchische Informationsinfrastruktur

# PILOTPROJEKT



Universität  
Bremen

Intersektorale Initiative  
zur nachhaltigen Erneuerung  
der digitalen Infrastruktur

## Grunddaten

<b>Projekttitel</b>	FAIRe Digitale Objekte für Deutschland: Universaler Datenstandard und hierarchische Informationsinfrastruktur
<b>Projektkronym</b>	FDO4DE: Pilot
<b>Projektbeginn</b>	1. Juli 2024
<b>Laufzeit in Monaten</b>	12
<b>Gesamtfördersumme in €</b>	1.780.000

Nr.	Beteiligte Partner	Startup
1.	Coherence	<input checked="" type="checkbox"/> in Planung
2.	D2030 Deutschland neu denken e.V.	<input type="checkbox"/>
3.	Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI) GmbH	<input type="checkbox"/>
4.	FIZ Karlsruhe - Leibniz Zentrum für Informationsinfrastruktur	<input type="checkbox"/>
5.	Frankfurt School of Finance and Management	<input type="checkbox"/>
6.	Hochschule Bremen	<input type="checkbox"/>
7.	Hochschule Hannover	<input type="checkbox"/>
8.	IndexNow	<input checked="" type="checkbox"/> in Planung
9.	KIT - Karlsruher Institut für Technologie (Exzellenzcluster 3DMM2O)	<input type="checkbox"/>
10.	Knowledge Pixels AG, Schweiz	<input checked="" type="checkbox"/>
11.	nextlearning e.V.	<input type="checkbox"/>
12.	Pathologie Hamburg-West GmbH	<input type="checkbox"/>
13.	QBITFLOW GmbH	<input checked="" type="checkbox"/>
14.	Sense	<input checked="" type="checkbox"/> in Planung
15.	Transparency International Deutschland e.V.	<input type="checkbox"/>
16.	Technische Universität Dresden	<input type="checkbox"/>
17.	Universal Data Interface	<input checked="" type="checkbox"/> in Gründung
18.	Universität Augsburg	<input type="checkbox"/>
19.	Universität Bremen	<input type="checkbox"/>
20.	Universität der Bundeswehr München	<input type="checkbox"/>
21.	Universität Leipzig	<input type="checkbox"/>
22.	YardStick Robotics GmbH	<input checked="" type="checkbox"/>
23.	YottaSen Forschungs-GmbH	<input checked="" type="checkbox"/>

# Einleitung

## Executive Summary

### Synopsis

Um zugleich einen universellen Datenstandard und eine universelle Informationsinfrastruktur zu realisieren, sind FAIR Digital Objects (FDOs) der effizienteste Weg. Sie können das Fundament eines universellen Datenraumes bilden, der es erlaubt, ressourcenschonend große Informationsmengen zu sammeln und zusammenzuführen. Hieraus lässt sich nachhaltig neues Wissen generieren und kuratieren.

### Mangelnde FAIRness der Datenlandschaft

Die weltweite Datenmenge wächst seit Jahren exponentiell. Vor diesem Hintergrund verbessern sich denklogisch auch die Ausgangsbedingungen, um mit Hilfe von Datenanalysen neues bzw. profunderes Wissen zu generieren. Um dieses enorme Potenzial auch praktisch auszuschöpfen, müssten Daten und Metadaten sowie die ihnen zugrunde liegende Dateninfrastruktur aber den FAIR-Prinzipien entsprechen.<sup>1</sup> Als Eckpfeiler für das Design einer effizienten digitalen Infrastruktur zielen sie auf technische Aspekte der Offenheit. Ob und unter welchen Nutzungsbedingungen Akteure Zugriff erhalten, determinieren sie hingegen nicht.

### Mangelnde Auffindbarkeit (Findability)

Damit Menschen und Computer konkrete Daten leicht auffinden können, müssen diese – auf syntaktischer Ebene – identifizierbar und referenzierbar sein. Dies kann ein Persistent Identifier (PID) gewährleisten. In verschiedenen Disziplinen sind entsprechende PIDs bereits etabliert. So lassen sich etwa Bücher und andere Schriftstücke über den DOI (Digital Object Identifier) eindeutig identifizieren. Allerdings fehlt es derzeit noch an einem geeigneten Mechanismus, um PIDs bereichsübergreifend zu generieren und zuzuordnen. Nutzer suchen bestimmte Daten regelmäßig nicht um ihrer selbst willen, sondern wegen ihres semantischen Inhalts, mit dem sie ein konkretes Erkenntnisinteresse befriedigen wollen. Bei dieser Suche helfen Metadaten – also gewissermaßen Daten über Daten – die Form und Inhalt der Daten sowie den jeweiligen Erhebungskontext beschreiben. Von größtmöglichem Nutzen sind entsprechende Metadaten dann, wenn sie disziplinübergreifend einer einheitlichen Semantik folgen.

### Mangelnde Zugänglichkeit (Accessibility)

Die FAIR-Grundsätze verfolgen das Ziel, diejenigen technischen und kommunikativen Zugangshürden zu beseitigen, aufgrund derer ein vereinbarter Zugang zu Daten unnötig ressourcenaufwändig und langwierig ist. Daher sehen sie eine Authentifizierung und Autorisierung mittels standardisierter Kommunikationsprotokolle vor. Diese Protokolle müssen kostenlos zur Verfügung stehen und universell implementierbar sein. Der Zugang zu den Metadaten sollte überdies dauerhaft und unabhängig vom Zugang zu den Daten selbst gewährleistet werden, um PIDs auf Metadaten zu ermöglichen. Allerdings ist insoweit die Datenschutzgrundverordnung (DSGVO) zu beachten, welche verlangt, dass personenbezogene Daten löscherbar sind.

---

<sup>1</sup>Die gemeinsam vom BMDV, BMWK und dem BMI erarbeitete Datenstrategie der Bundesregierung "Fortschritt durch Datennutzung" benennt die FAIR-Prinzipien Findable, Accessible, Interoperable, Reusable als grundlegend notwendige Dateneigenschaften (Wilkinson 2015). Diese kuratiert die GO FAIR Foundation, mit der FDO4DE zusammenarbeitet, um maschinenlesbare FAIR Implementation Protocols (FIPs) zu erstellen. Hiervon profitiert die Entwicklung von KI ganz allgemein.

## Mangelnde Interoperabilität (Interoperability)

Der praktische Befund auf dem Feld der Datenwirtschaft ist eindeutig: Eine hinreichende Interoperabilität der verschiedenen Formate und logischen Strukturen, in denen unsere Daten vorliegen bzw. gespeichert werden, fehlt.<sup>2</sup> Dementsprechend ressourcenaufwändig ist es, Daten aus verschiedenen Quellen zusammenzuführen. Das betrifft nicht nur private Datenverarbeitungen; mangelnde Dateninteroperabilität zieht sich auch wie ein roter Faden durch die staatliche (Daten-)Infrastruktur. So ist es beispielsweise nicht möglich, die Datenbanken der Medizininformatikinitiative (MII) und der nationalen Forschungsdateninfrastruktur (NFDI) übergreifend zu durchsuchen. Daten müssen aber in der Regel mit anderen Daten integriert werden, um eine vergleichende Betrachtung gemeinsamer Bezüge barrierefrei zu ermöglichen. Das setzt wiederum voraus, dass Anwendungen oder Arbeitsabläufe zur Analyse, Speicherung und Verarbeitung interagieren. Eine vor diesem Hintergrund essenzielle Interoperabilität<sup>3</sup> ist erreicht, wenn:

- ☑ eine formale, zugängliche, gemeinsame und breit anwendbare Sprache zur Wissensdarstellung zum Einsatz kommt,
- ☑ verwendete Vokabulare den FAIR-Grundsätzen entsprechen,
- ☑ (Meta)daten qualifizierte Verweise auf andere (Meta)daten enthalten und
- ☑ über statistische Konzepte aufeinander beziehbar sind.

## Mangelnde Weiterverwertbarkeit (Reusability)

Metadaten werden im Optimalfall mit einer Vielzahl von relevanten Attributen so genau und vollständig erfasst, dass es möglich wäre, die Daten erneut zu erheben, oder z.B. eine konsistente Verifizierung eines Herstellungsverfahrens mit ähnlichen Parametern (Attributen) durchzuführen. Die Replizierbarkeit und Validität von Forschungsergebnissen zu bestimmen, setzt nicht nur "open data", sondern auch "open practices" und "open tools" voraus (Voelkel & Freese 2022)<sup>4</sup>. Konsistent bedeutet, dass zum Beispiel eine zu verifizierende Messreihe mit höherer Auflösung oder einem erweiterten Messbereich mit einem Ergebnis durchgeführt werden kann, das die ursprünglich erhaltenen Daten passend inter- oder extrapoliert.

Nur ein vollständiger Nachweis der Passgenauigkeit im jeweiligen Kontext sichert die Weiterverwertbarkeit. Diese ist gegeben, wenn Metadaten:

- ☑ mit klaren und zugänglichen Datennutzungslizenzen versehen und
- ☑ mit einer detaillierten Provenienz (also Informationen darüber, woher das Datenmaterial stammt bzw. mit welchen Methoden es gewonnen wurde) verbunden sind,
- ☑ die für den jeweiligen Bereich relevanten Standards erfassen und
- ☑ durch sie parametrisierte, im Fach übliche Vorgehensweisen vollständig erfüllt werden.

Die Weiterverwertbarkeit hängt insbesondere auch mit der Zertifizierung der Qualität von Daten sowie der Vertrauensstufe (Trust Level) von Datentreuhändern und Softwareherstellern, Softwaresystemen und Rechnerarchitekturen zusammen, die erst in den Kinderschuhen steckt.

<sup>2</sup>Dies hat auch der Europäische Gesetzgeber erkannt. Er etabliert daher in den Art. 33 ff. Data Act konkrete Interoperabilitätsanforderungen.

<sup>3</sup>Interoperabilität bezieht sich hier nur auf Daten. Das European Interoperability Framework (EIF) definiert den Begriff der Interoperabilität im Gegensatz hierzu allgemein und fasst ihn weiter

<sup>4</sup>Voelkel, Jan G. & Jeremy Freese. 2022. Open Computational Social Science. In Uwe Engel, Anabel Quan-Haase, Sunny Xun Liu & Lars Lyberg (Eds.), Handbook of Computational Social Science, Volume 1: Theory, Case Studies and Ethics, 119-130. London/New York: Routledge

Viele Bereiche der Wissenschaft und des Datenmanagements leiden seit Jahren unter einer handfesten Reproduzierbarkeitskrise (Ioannidis, 2005). Diesem Problem kann mit dem Konzept der Preproducibility (Stark, 2018) begegnet werden: Die Mitteilung eines wissenschaftlichen Resultats erfordert die Aufzählung, Aufzeichnung und Mitteilung derjenigen Dinge, die nicht ohne weiteres weggelassen werden können, um das Ergebnis zu reproduzieren. Ein Experiment oder eine Analyse ist vorproduzierbar, wenn sie ausreichend detailliert beschrieben wurden, damit andere sie durchführen können. Workflows – unter besonderer Berücksichtigung der Interdisziplinarität und Datenheterogenität in der Erforschung komplexer Systeme – sollten daher auf den Anforderungen der Preproducibility basieren, die auf einem konsistenten Ökosystem von Prozessen aufbaut.

Die Weiterverwertbarkeit setzt zudem voraus, dass entsprechende Datensätze verfügbar sind. Die reichhaltigen Daten des statistischen Verbundes stehen aber aufgrund von (noch) nicht vorhandenen Rechtsgrundlagen<sup>5</sup> für die Weiterverwendung nicht oder nur sehr eingeschränkt zur Verfügung und können daher nicht genutzt werden. Auch fehlt die Rechtsgrundlage für eine aktive Zusammenarbeit der statistischen Landesämter und des Statistischen Bundesamtes (destatis) mit Forschenden.

## **Chancen und Herausforderungen des Wandels**

### **Permanente Kuratierung des Digitalen Wissens für PIDs**

Auf dem Weg, die FAIR-Prinzipien zu implementieren, ist es nicht zielführend, ausschließlich die Art und Weise der Datenspeicherung zu verändern. Wie vom Data Act der EU vorgesehen, benötigen wir vielmehr einen geeigneten Mechanismus, um unser Wissen technologieunabhängig zu wahren und an zukünftige Generationen weitergeben zu können. Gelingt es nicht, einen solchen zu finden, würde das bedeuten, dass wir unseren Datenschatz, den wir derzeit misstrauisch und eifersüchtig in Silos hüten, langfristig verlieren werden. Dieses Szenario hat Vinton Cerf bereits 2015 als “Digital Dark Age” bezeichnet.

Die FAIR-Prinzipien auf alle bestehenden Daten anzuwenden, verschlänge zwar immense Ressourcen. Zugleich steht aber auch fest: Je länger Staat und Gesellschaft diese wichtige Aufgabe aufschieben, desto teurer wird uns dieses Zögern in Zukunft zu stehen kommen. Es besteht sogar die reale Gefahr, dass der Aufwand, um die bestehenden Daten angemessen permanent zu kuratieren, infolge des eingangs erwähnten exponentiellen Wachstums so stark ansteigt, dass er unbezahlbar bzw. grundsätzlich nicht mehr leistbar ist. In diesem Fall käme es unweigerlich zu einem unwiederbringlichen Verlust von Informationen und folglich auch dem entsprechenden Wissen.

Einen solchen umfassenden Mechanismus einzuführen, entpuppt sich als schwierige Aufgabe. Diese erfordert eine transdisziplinäre Agenda, und ist zudem ohne Mitwirkung der Zivilgesellschaft (als Träger des Bürgerwillens) nicht möglich. Die Digitale Transformation sollte daher transdisziplinär, staatlich transsektoral sowie in Zusammenarbeit aller Ressorts der Bundesregierung erfolgen, um die Funktionalität und damit den ökonomischen Wert von Daten zu steigern. Weil infrastrukturelle Aufbauprojekte stets langfristige Aufgaben sind, muss der Prozess sich über mehrere Legislaturperioden erstrecken und ist daher überparteilich anzulegen.

---

<sup>5</sup>Vgl. § 12a EGovG.

## Harmonisierung von Archivierung und Transfer

Die klassische, tradierte Trennung zwischen Grundlagenforschung, angewandter Forschung sowie industrieller Forschung und Entwicklung existiert in der KI-getriebenen Digitalen Transformation weitgehend nicht mehr. Diese Entwicklungsstadien überlappen sich häufig oder finden sogar gleichzeitig statt. Dies erfordert und treibt den Transfer von Know How aus den Wissenschaften in die Wirtschaft.<sup>6</sup>

Es bedarf einer Kombination aus Agilität und Erfahrung, um unsere Gesellschaft weiterzuentwickeln. Technologische Innovationen müssen auf gewachsenen Erfahrungswerten aufbauen. Die maximale Zeitspanne, innerhalb derer sich Archivare und Sprung-Innovatoren austauschen sollten und für eine optimal-regulative, fiskalische und föderale Governance auch einigen müssen, ergibt sich aus jährlichen Haushalten sowie Legislaturperioden.

## Transformation durch Objektifizierung

### Instanziierung des FDO-Frameworks durch Nanopublikationen

Unsere Lösung der geschilderten Anforderungen und Probleme der Digitalen Transformation besteht darin, FAIR Digital Objects (FDOs) in der von Bonino 2022 vorgeschlagenen Form zu nutzen. Um ihre Instanziierung<sup>7</sup> zu ermöglichen, bedarf es sog. Nanopublikationen<sup>8</sup> als Systemarchitektur: Sie ermöglichen es, Wissen und dessen Kontext in kleinen Portionen auf präzise, zuverlässige und schnelle Art zu kommunizieren.<sup>9</sup> FDOs erlauben zum einen den universellen Austausch beliebiger Datenpakete, ergeben zum anderen aber auch eine mathematisch unterlegte<sup>10</sup> Kategorisierung in einer Informationsinfrastruktur, die durch Kombination einfacher Dateneinheiten entsteht. Dies sind Elementarzellen, aus denen sich komplexe Objekte zusammensetzen lassen. Insbesondere sind zusammengesetzte FDOs wieder FDOs und lassen sich daher auch komplex und semantisch aggregieren.<sup>11</sup> Dies führt dann letztlich zu einem universell interoperablen Datenraum. Dieser ist wegen seines Gesamtzusammenhangs<sup>12</sup> auch vorteilhaft für KI-Systeme, die damit Zugang zu allen möglichen und relevanten Trainingsdaten erhalten.

---

<sup>6</sup>Insoweit ergeben sich jedoch erheblich divergierende Zeitspannen. Dem Transfer aus der State-of-the-Art KI mit einer minimalen Innovationszeit von ca 10 Wochen stehen die Archivierung von bleibenden Kulturgütern und die Dokumentation von langfristigen gesellschaftlichen Prozessen mit Zykluszeiten auf einer Skala von 10 Jahren gegenüber. Bei der Periode der Generationsfolge ergeben sich 15 Jahre für den Zeitgeist (Millennials, Digital Natives, Generation Alpha) und 30 Jahre für biologische Populationen.

<sup>7</sup>Instanziierungen bestimmen die Grundeinheiten des dimensionslosen Rahmens. Die relevanten Einheiten von Raum und Zeit eines Datenraumes, sowie weitere emergente Eigenschaften, sind vom Kontext abhängig. So hat ein Teilchendetektor am CERN einen anderen Anspruch an zeitliche Präzision und Auflösung als eine industrielle Fertigungsstraße oder ein einfaches Buchregister einer Stadtbibliothek.

<sup>8</sup>Erik Schultes, Barbara Magagna\*, Tobias Kuhn, Marek Suchánek, Luiz Bonino da Silva Santos, Barend Mons, The Comparative Anatomy of Nanopublications and FAIR Digital Objects, Research Ideas and Outcomes 8: e94150 2022, doi: 10.3897/rio.8.e94150

<sup>9</sup>In jüngerer Zeit sind Nanopublikationen mit ihren semantischen Fähigkeiten und ihrer Universalität auch als Implementierung von FAIR Digital Objects in den Fokus geraten. Konkret können Nanopublikationen die Metadaten Records von FDOs abbilden und so als zentraler Grundstein die praktische Einführung von FDOs ermöglichen.

<sup>10</sup>FDOs sind mathematische Kategorien, repräsentieren also die Eigenschaften von Objekten.

<sup>11</sup>Eine semantische Aggregation ist eine Zusammenstellung, die die aggregierten Teile im Kontext in Beziehung setzt. Ein universeller globaler Daten- und Kommunikationsstandard ergibt sich durch die Darstell- und Kommunizierbarkeit jedes Datenpaketes als FDO.

<sup>12</sup>Gesamtzusammenhang bedeutet, dass alle Teile des Raumes kontinuierlich miteinander verbunden sind.

## Open Data und die Rolle der Zivilgesellschaft

Die FAIR-Grundsätze sind um das Prinzip "Open Data" zu ergänzen: Dieses verlangt eine möglichst uneingeschränkte Lizenz zur Weiterverwendung von Daten. Hiervon soll ein möglichst großer Personenkreis profitieren, auch grenzüberschreitend sowie zu kommerziellen Nutzungszwecken. Der Open-Data-Ansatz erfasst dabei insbesondere solche Daten, die im Wesentlichen durch Steuergelder finanziert sind; zunehmend aber auch Daten, die private Stellen (etwa privatwirtschaftliche Akteure) erhoben haben, wenn an den Datensätzen ein hinreichendes öffentliches Interesse besteht. Einer Weiterverwendung dürfen allerdings keine gravierenden Interessen der Unternehmen entgegenstehen.

### Eine universelle Bewertung von Information liefert transdisziplinäre Maßzahlen

Komplex aggregierte FDOs ermöglichen es, transdisziplinäre Maßzahlen der verschiedenen Unterräume des universellen Datenraumes zu erstellen. Anhand dieser Maßzahlen gelingt eine umfassende Darstellung und Bewertung der im Fokus stehenden Dateneigenschaften inklusive ihrer semantischen Inhalte.

Für den Piloten haben wir exemplarisch paradigmatische Projekte ausgewählt, die überwiegend Start-ups im Transfer zwischen Wissenschaft, Technologie, Wirtschaft und Gesellschaft umsetzen. Aus transdisziplinär zusammengeführten Mehrfachbewertungen von Daten ergibt sich insoweit eine universelle Multi-Indexierung. In sehr unterschiedlichen Fachgebieten lassen sich nach dem gleichen universellen Prinzip transdisziplinäre Maßzahlen<sup>13</sup> gewinnen. Diese Maßzahlen bewerten die fachspezifische Information quantitativ und demonstrieren gleichzeitig damit den praktischen Nutzen der FDOs.

---

*Datenräume sind hierarchisch gegliedert. Diese Hierarchie hat ihren Ursprung in einer Sequenz des von Anderson bereits 1972 eingeführten Prinzips des "More is Different". Auf jeder Organisationsebene eines komplexen Systems entsteht durch Emergenz aus den Teilen der tieferen Ebene das neue Ganze der höheren Ebene. Typischerweise gliedert sich das Ganze oberhalb einer für jede Ebene charakteristischen räumlichen Granularität der Materie heterarchisch in neue Teile mit bestimmten charakteristischen Eigenschaften. Diese neuen Eigenschaften existieren erst auf der neuen, emergenten Ebene,<sup>14</sup> siehe Sektion Informationstheoretischer Hintergrund. Jeder dieser Organisationsstufen ordnet sich eine Wissenschaft zu, die die effektiven Objekte dieser Ebene bestmöglich beschreibt. FDOs tragen die Eigenschaften der realen Objekte. Sie sind die Eigenschaften der Objekte, nicht die Objekte selbst. Mathematisch betrachtet sind sie daher Kategorien von Eigenschaften. Die Bündelung und Bewertung dieser Eigenschaften konstituiert transdisziplinäre Multi-Indizes<sup>15</sup> oder einfach Maßzahlen.*

---

<sup>13</sup>Maßzahlen sind Metriken, die Attribute von Daten quantitativ bewerten, siehe Informationstheoretische Grundlagen. Die vielfältige Anwendbarkeit von FDOs soll in stark unterschiedlichen Datenräumen gezeigt werden.

<sup>14</sup>Die Eigenschaften von Wasser, nass und zähflüssig zu sein, ergeben sich erst im kondensierten Zustand von Wasser. Einzelne Wassermoleküle benetzen nicht und besitzen keine Viskosität. Die neuen Teile, die die neuen Eigenschaften tragen, sind physikalisch betrachtet die Quasiteilchen der nächsthöheren Ebene. Ein aus dem Schulunterricht bekanntes Beispiel dieser Hierarchie ist die Sequenz von Elementarteilchen, Kernen, Atomen, Molekülen, Makromolekülen, Organellen, Zellen, Geweben, Organen und schließlich autonomen Organismen, die wiederum Gruppen, Gemeinschaften und Staaten bilden.

<sup>15</sup>Diese Multi Indizes sind Metriken des jeweils zugeordneten Datenraumes. Dies begründet den Begriff semantische Metrik.

## Paradigmatische Beispiele transdisziplinärer Maßzahlen in der Praxis

- personalisierte Therapien zur Behandlung von Brustkrebs aus genetischen, molekularen, histologischen und biophysikalischen Indikatoren (**Lebenswissenschaften**)
- hierarchisch geordnete kontinuierliche Statusmeldungen des Systemzustandes von autonomen Robotern (**Ingenieurwissenschaften**)
- individualisierte Inflationsraten aus dem Preisindex varianter Warenkörbe (**Wirtschaftswissenschaften & Verwaltungswissenschaften**)
- Grad der Erfüllung eines automatisiert semantisch bewerteten Attributs, z.B. eines der adaptierten Ziele für nachhaltige Entwicklung der Vereinten Nationen (**Zivilgesellschaft in der Nachhaltigkeit, Open Data for Open Governance**).

## Abgrenzung von anderen Lösungsansätzen und Fokussierung

Neben den FDOs<sup>16</sup> gibt es noch weitere<sup>17</sup> valide Lösungsansätze, um eine semantische Interoperabilität in einem universellen Datenraum herzustellen. Unser Konsortium FDO4DE wird mit der Go FAIR Foundation als Kontraktor zur Nutzung von FAIR Implementation Profiles den FAIR Gedanken in Deutschland

- i) innerhalb einiger ausgewählter Datenräume mit FDOs in die Praxis umsetzen (**Mission**) und so Daten einer verstärkten Nutzung zuführen sowie
- ii) konzeptionell für alle Sektoren des Staates Daten ökonomisch optimal entwickeln (**Vision**).

Es gibt zwei sich ergänzende FDO-Konsortien. Wohingegen sich FDO One weitgehend auf den Teilaspekt eines universellen Datenstandards beschränkt, fokussiert sich FDO4DE bereits im Rahmen dieses Pilotprojekts verstärkt auf den Aufbau und die Nutzung einer universellen Informationsinfrastruktur, ausgerichtet auf die Mission KI und deutlich darüber hinaus.

Es zeigt sich, dass FDOs keinesfalls nur als Datenstandard wirken, sondern Daten auch kategorisieren. Ihre Bedeutung reicht sogar noch weit darüber hinaus: Sie tragen fundamental zum systematischen Verständnis unserer Welt bei. Es sind:

- i) die FAIR Prinzipien,
- ii) die Kategoriestructur der FDOs,
- iii) die inhärente attributive Bewertungsfunktion der Digitalen Objekte,
- iv) das wissenschaftliche Erklärungspotential,
- v) der Primat des elementar Einfachen und
- vi) die sich darin gründende Ordnungsmacht für komplexe Daten und damit für unser Wissen,

welche die FDOs zusammengenommen auszeichnen und als beste Wahl für die fundamentalen Bausteine der Digitalen Transformation nahelegen. FDO4DE baut damit letztendlich die unerlässliche Infrastruktur für eine universelle Kommunikation sowie für ein universelles Verständnis von Information.

## Projektstruktur der Transdisziplinären Maßzahlen: Rolle der Start-ups

Das Ziel dieses zentralen Arbeitspaketes ist es, PID-markierte Informationen aus verschiedenen Objekten variabler Disziplinen zu transdisziplinären Maßzahlen im Rahmen der Multi-Indexierung

<sup>16</sup>Diese kuratiert das international organisierte FDO Forum in Zusammenarbeit mit der GO FAIR Foundation mit Sitz in Leiden (Niederlanden).

<sup>17</sup>Linked Data, Signposting und RO-CRATE

zusammenzuführen. Indem sie den gemeinnützigen Verein IndexNow e.V. gründen, benennen die Gründungsmitglieder die Multi-Indexierung als wesentliches funktionales Bindeglied der Sektoren des Staates.

Index Now wird als Träger und Mittler der fachlichen Informationen eng mit dem KIT Start-up Universal Data Interface zusammenarbeiten. Diese Kooperation bezweckt, ausgewählte Pilotprojekte zur Bereitstellung von Indizes relevanter sozio-kultureller, ökologischer und ökonomischer Maßzahlen zu realisieren und mit FIZ Karlsruhe und Knowledge Pixels zu repositionieren. Das Start-up Coherence übernimmt in Zusammenarbeit mit dem Datenraum Komplexe Systeme sowie mit weiteren Akteuren sowohl die systematische Koordination als auch die universelle Struktur der Indexerstellung. Es leistet durch ein konzeptionelles Design des Gesamtprozesses eine transdisziplinäre Vermittlung und Definition von Begrifflichkeiten im Kontext unter den Disziplinen, zum Beispiel als Grundlage emanzipatorischer Infrastrukturen.<sup>18</sup>

Mit der themenfeldübergreifenden Strukturierung der Indexierung öffnet sich diese für eine Kommerzialisierung der Zuarbeit für staatliche Verwaltungen und Behörden sowie für die Beratung von Wirtschaft und Industrie. Wir werden dieser erwarteten Entwicklung durch die Gründung von Index Now Commercial Rechnung tragen.

Multi-Indizierte öffentliche Datensätze (Open Data) bilden die Grundlage sowohl einer wissensbasierten offenen Regierungsarbeit (Open Government) als auch einer informierten Gesellschaft. Im Hauptprojekt ist es unser Ziel, FDOs im Allgemeinen und im Speziellen als Träger von Open Data in den 5. Nationalen Aktionsplan im Rahmen der Teilnahme Deutschlands an der Open Government Partnership (OGP) für 2025-2027 einzubringen.

## **Visionen des Gesamtprojekts**

Der Pilot soll folgende Aufgaben für das Gesamtprojekt konzeptionell vorbereiten:

- Etablierung der Multi-Indexierung als universalen Ordnungsrahmen des Datenmanagements für alle Disziplinen und Anwendung in allen staatlichen Sektoren
- Proklamierung und Umsetzung der FDOs als ideale Träger Offener Daten sowie als Grundlage einer informierten, transparenten Politik und des Regierungshandelns
- Datenrechtliche Analyse des Zusammenspiels von bürgerlichem, öffentlichem und europäischem Recht mit technologischen Erfordernissen und umgekehrt
- Bereitstellung zivilgesellschaftlicher und infrastruktureller Expertise bei der Gründung des Dateninstituts
- Kombination von klassischen FDOs mit Quantentechnologie in den Quantum FDOs, um Deutschland weltweit an die Spitze der Digitalen Transformation zu katapultieren

Weitere Ziele:

- Feedback einer sozio-technologisch-epistemischen Analyse der Wirkungen von Infrastruktur auf deren Konzeption
- Feedback einer makroökonomischen Analyse der Wirkung von Standardisierung auf deren Gestaltung

---

<sup>18</sup>Rack, Oliver. *Öpen Government und die Teilnahme Deutschlands an der Open Government Partnership* Forschungsjournal Soziale Bewegungen, vol. 36, no. 3, 2023, pp. 370-380. <https://doi.org/10.1515/fjsb-2023-0033>; Übersetzung ins Englische, siehe <https://www.oliverrack.eu/ogpbw/ogp-germany>

- Entwicklung von xAI, d.h. verständlicher KI, in allgemeinen komplexen Systemen und Umsetzung von Paradigmen in digitaler Medizin und Robotik
- Aufzeigen der Notwendigkeit verstärkter öffentlicher Diskussion zur Ethik von KI und Durchführung derselben im Think Tank
- Verdeutlichung des universellen Anspruchs der FDOs als Ordnungs- und Strukturelemente eines integrierten Datenraumes; zunächst beginnend im föderalen Deutschland, dann in Europa und letztlich auch auf internationaler Ebene in Interaktion mit der OECD und den Vereinten Nationen

## **Mission von FDO4DE**

Die Ergebnisse dieses Pilotprojekts dienen als Fundament für ein sich anschließendes Hauptprojekt FDO4DE. Dieses soll das Ziel verfolgen, die Anwendung von FDOs einerseits zu verstärken und andererseits auch in weiteren Disziplinen und Sektoren voranzubringen. Dies setzt voraus, die Pilotprojekte mit verstärkter Personalausstattung fortzuführen und um zusätzliche relevante Dimensionen der Datenräume zu erweitern: Diesen multisektoralen Ansatz mit trans- und disziplinären Mehrwerten werden wir umfassend demonstrieren.

Es gilt, die Digitale Transformation in

- i) technologischer (klassisch wie quantenmechanisch)
- ii) rechtlicher,
- iii) soziologischer,
- iv) makroökonomischer und
- v) systemischer Beziehung

zu beleuchten bzw. zu entwickeln. Diese absolut notwendige Entwicklung kann nur in einem entsprechend ausgestatteten Hauptprojekt angegangen werden. Unser Ansatz ist deutschlandweit der erste, der eine derartig umfassende Philosophie verfolgt. FDO4DE nimmt daher eine Pionierrolle ein und kann seine Vorhaben transdisziplinär in die Tat umsetzen:

Ganz konkret werden wir z.B. die Möglichkeiten der Digitalen Medizin erheblich ausweiten: Zu den individuellen Krebstherapien, die sich tatsächlich erst in einem Hauptprojekt realisieren lassen, werden wir einen herstellerunabhängigen Bildstandard in der Kernspintomographie hinzunehmen. In der Robotik erweitern wir die Betrachtung von struktureller auf kognitive Robotik. Kognitive Roboter leisten anspruchsvolle Aufgaben, die eine kognitive Leistung erfordern. Ein körperbetonter Roboter hingegen kann z.B. gut in wechselndem schwierigem Gelände laufen. Wir werden komplexe Systeme semantisch erschließen, interoperabel quantifizieren und damit zu xAI beitragen. Quanten-FDOs werden Quantentechnologien in die sich entwickelnde Informationsinfrastruktur einbinden. Eine datenschutzrechtliche Analyse sollte zudem um eine umfassende öffentlich-rechtliche Prüfung erweitert werden. Hinzukommen muss auch eine Betrachtung rechtsphilosophischer und ethischer Aspekte, um überhaupt erst eine umfassende Bewertung zu ermöglichen. Außerdem werden wir eine verstärkte Öffentlichkeitsarbeit leisten müssen, um die Bürger umfassend über FDOs zu informieren und eine breite Akzeptanz in der Bevölkerung herzustellen.

## Relevanz und Bedeutung

- FDOs errichten die Informationsinfrastruktur des 21. Jahrhunderts.
- Deutschland braucht innovationstreibende FDOs weil zB die Niederlande und die USA bereits FDOs entwickeln. Es gilt, im internationalen Vergleich den Anschluss nicht zu verlieren.
- FAIR Digital Objects für Deutschland hat im Hauptprojekt die erforderliche Expertise und transdisziplinäre Struktur, um FDOs in Zusammenarbeit mit der FITKO und der KoSIT im Rahmen des IT-Planungsrates des Bundes und der Länder in die Digitale Transformation einzubringen.
- FDO4DE wird soziologische, rechtliche sowie technologische, makroökonomische und systemische Aspekte integrieren und umfassend beleuchten.
- Kanonische Arbeitsabläufe von Forschung und Produktion sind abstrakt betrachtet ähnlich zu Verwaltungsvorgängen und können dadurch in der Verwaltungsdigitalisierung wirken und ebenso umgesetzt werden.

## Arbeitspakete

Das Pilotprojekt besteht aus vier Kernarbeitspaketen, die sich gegenseitig komplementieren. Die Arbeitspakete *AP1 Systemarchitektur* und *AP2 Technologische Werkzeuge* liefern die für FDOs notwendige Infrastruktur. Ihre Zielgruppe sind die User, die in dem Arbeitspaket *AP3 Datenräume* verankert sind. Das Arbeitspaket *AP4 Gesellschaftliche Werkzeuge* behandelt die semantische Analyse, die Multi-indexierung (charakteristisch für alle Datenräume), die Integration von Technologie und Recht, die makroökonomische Bewertung, die strategische Vorausschau, sowie die Kommunikation und Interaktion der Wissenschaft mit dem Bürger in einem Think Tank. Im Folgenden beschreiben wir die Aufgaben unserer Start-ups in einer effizienten Konsortialstruktur. Die Start-ups der einzelnen Datenräume arbeiten fachlich jeweils direkt mit ihren zugeordneten Projektpartnern zusammen.

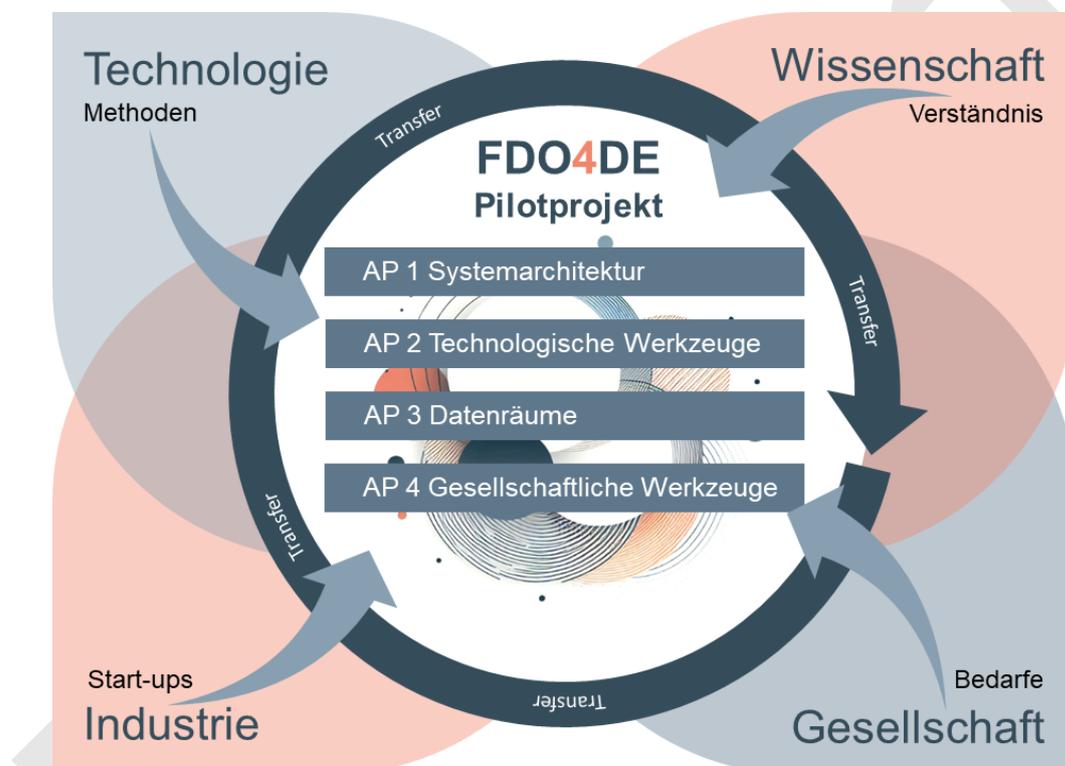


Abbildung 1: Struktur des FDO4DE Pilotprojekts 2024/2025

## AP<sub>P</sub> 1. Systemarchitektur

---

<b>Start-up</b>	Knowledge Pixels AG Tobias Kuhn, Philipp von Essen
<b>Mission</b>	Knowledge Pixels und FIZ Karlsruhe kombinieren ihre Expertise aus Nanopublikationen und Forschungsdateninfrastruktur — insbesondere Repositorien — und führen diese Technologien konzeptionell und praktisch zusammen für den Aufbau einer flexiblen Systemarchitektur für das Management von Information auf allen Detailebenen als FDOs. Im Zentrum der Betrachtung steht die Modellierung und Entwicklung von Datenpublikationen (mit DOIs, maschinenlesbaren Landingpages von Datensätzen, Metadaten, Binärdaten) als komplexe FDOs, deren Bestandteile durch Nanopublikationen (semantischen Aussagen über beliebige Entitäten) als elementare FDOs erzeugt werden.
<b>Partner</b>	FIZ Karlsruhe, Felix Bach
<b>Zielgruppe</b>	Use Cases AP 3. Datenräume

---

### Ziel des Arbeitspakets

- Konzeption einer flexiblen FDO Systemarchitektur basierend auf der Kombination von diesen zwei Hauptkomponenten:
  - Nanopublikationen als Implementierung von FDOs mit der Möglichkeit durch Komposition komplexe FDOs aus einfacheren zu erstellen.
  - Datenpublikationen in einem Datenrepositorium, die als komplexe FDOs modelliert sind, zusammengesetzt aus Elementen wie Versionen, Formate, Lizenzen, Bedingungen, Dokumentation, und Tabellen.
- Proof of Concept für eine verteilte und vernetzte Infrastruktur für FDOs auf Basis von Nanopublikationen und Datenpublikationen, die einen Umgang mit Daten völlig unterschiedlicher Datengranularität erlaubt.

### Aufgaben

- Konzepterstellung einer flexiblen, vernetzten FDO-Infrastruktur als Basis für FDO4DE.
- Proof-of-Concept-Implementierung dieses Konzepts als Testbett für FDO4DE auf Basis des Datenrepositoriums RADAR.

### Meilenstein(e)

- Konzeptentwurf für Nanopublikationen als FDO-Implementierung.
- Konzeptentwurf zur Modellierung von Datenpublikationen als FDOs/Nanopublikationen.
- Lauffähiges und testbares Proof-of-Concept-System, das FDOs/Nanopublikationen mit RADAR verbindet und die generelle Anwendbarkeit demonstriert.

## AP<sub>p</sub> 2. Technologische Werkzeuge

---

**Start-up** Universal Data Interface  
Frank Tristram

**Mission** Unsere Überzeugung ist, dass alles was Menschen händisch konvertieren können, von einem Computer vollautomatisch, oder mit nur wenig Anweisung konvertiert werden kann. Dementsprechend lesen wir maschinell unstrukturierte Dateien (z.B. pdf) "wie ein Mensch es würde" ein, um einfache Sinnstrukturen (Tabellen, Key-Value-Paare, Typen etc.) automatisch anzureichern und für Suchen vorzustrukturieren.

---

**Partner** Exzellenzcluster 3DMM2O, KIT  
U Siegen, Mario Agio

---

**Zielgruppe** Use Cases AP 3. Datenräume

---

### Ziel des Arbeitspakets

- Die Tools aus dem Exzellenzcluster 3DMM2O, die Interfaces zwischen Programmen und Formaten automatisch überbrücken, sollen in einer Ausgründung gebündelt, sowie im Projekt und am Markt verfügbar werden. Automatische Kontextualisierung, Extraktion und Kombination von Metadaten und Daten soll dadurch auch für IT-Laien machbar werden, die Daten aus verschiedenen Quellen nutzen wollen.

### Aufgaben

- Allgemeine, unstrukturierte Daten automatisch erfassen und in bekannten Strukturen zurückgeben.
- Mapping und Fusion von heterogenen Datenquellen konfigurierbar ermöglichen.

### Meilensteine

- Ein Onlinetool und eine API bereitstellen, um Datenstrukturen in unstrukturierten Dateien zu erkennen.
- Erste prototypische Konfigurationen der gewünschten Rückgabestrukturen für das Mapping der Projektpartner verfügbar.

## AP<sub>p</sub> 3. Datenräume

### AP<sub>p</sub> 3.1. Digitale Medizin

---

<b>Start-up</b>	YottaSen Forschungs-GmbH Nils Niendorf
<b>Mission</b>	Unserer Überzeugung, dass sich die "Information im Bild befindet", folgend bringen wir eine der größten Datensammlungen auf dem Gebiet Brustkrebs mit pathologischer, medizinischer und KI-Expertise zusammen um patientinnenindividuelle Therapielösungen zu entwickeln. Unsere zukünftigen Modelle sollen in der Lage sein Prognose, Prediktion und Stratifizierung für Brustkrebs-Patientinnen zu ermöglichen und damit den Outcome auf individueller Ebene zu verbessern.
<b>Partner</b>	Pathologie HH West, Axel Niendorf Universität Leipzig, Josef Käs
<b>Zielgruppe</b>	Medizinische Diagnostik und Forschung, Kliniker, Patientinnen

---

#### Ziel des Arbeitspakets

- Entwicklung eines standardisierten Daten-Formats (FDO), das erlaubt medizinische Informationen (Bild- und medizinische/klinsche Daten) zu verknüpfen und diese für die Entwicklung KI basierter Lösungsansätze nutzbar macht.

#### Aufgaben

- Erstellung eines Datenraums Digitale Medizin am Beispiel des Mamma Karzinoms, einer bösartige Krebserkrankung weiblichen Brustgewebes.
- Digitalisierung von Gewebeproben und Verknüpfung mit diagnostischen, klinischen Informationen.
- Entwicklung einer FDO-basierten Datenbank-Infrastruktur.
- Überführung der Daten in eine strukturierte Datenbank.
- Auswertung der Daten zur Entwicklung eines Proof-of-Concepts zur KI-basierten Diagnostik, Prognostik oder Vorhersage.
- Erarbeitung eines einheitlich anwendbaren Struktur-Standards für weitere medizinische Anwendungsfälle.

#### Meilensteine

- FDO basierte Datenbankstruktur am Beispiel des Mamma Karzinoms aufgesetzt und für KI-PoC genutzt.

### AP<sub>p</sub> 3.2. Robotik

---

<b>Start-up</b>	Yardstick Robotics Sebastian Bartsch (CTO), Domenico Dagostino (CEO)
<b>Mission</b>	Entwicklung von KI basierten Robotik-Systemen.
<b>Partner</b>	DFKI GmbH, Robotics Innovation Center, Frank Kirchner, Sirko Straube
<b>Zielgruppen</b>	Industrie, Robotik, Forschung

---

#### Ziel des Arbeitspakets

- Roboterdaten mittels FAIR Digital Objects öffentlich speichern, um diese anderen zur Verfügung zu stellen und die Verwendung dieser Daten auf Basis von FDOs zu testen.

#### Aufgaben

- Aufnahme von Roboterdaten (Bilder, Punktwolken, GPS) auf einer Roboterteststrecke gemeinsam mit dem Partner Yardstick Robotics.
- Anreicherung der Daten mit Meta-Informationen zusammen mit dem Partner Yardstick Robotics.
- Die Daten mittels FDOs (wie z.B. Nanopublikationen) speichern und öffentlich zugänglich machen.
- Datenanalyse und verbesserte Visualisierung auf Basis von FDOs.

#### Meilensteine

- Aufgezeichnete Roboterdaten als FDOs abgelegt.

### AP<sub>p</sub> 3.3. Komplexe Systeme

---

<b>Start-up</b>	Coherence (in Planung) Marcus Kraus
<b>Mission</b>	Entwicklung eines übergeordneten konzeptionellen Designs für heterogene Daten mit dem Ziel der Erfassung von kausalen Zusammenhängen.
<b>Partner</b>	Frankfurt School of Finance & Management, Jan Nagler TU Dresden, Marc Timme U Bremen, Michael Beetz, Uwe Engel, Thorsten Fehr, Robert Porzel FHI für Digitale Medizin, Matthias Günther U Magdeburg, Wolfgang Marwan
<b>Zielgruppen</b>	Systemarchitektur Datenräume Gesellschaft

---

#### Ziel des Arbeitspakets

Komplexe Systeme erfordern zu ihrer Beschreibung eine Sprache, die der Hierarchieebene der Strukturordnung angemessen ist. Dies kann zum Beispiel entweder die qualitative Beschreibung eines soziologischen Zusammenhangs, wie auch die Modellierung eines komplexen Netzwerkes sein. In unserem Startup Coherence werden in der Interaktion zwischen Natur-

und Geisteswissenschaft Kausal-Paradigmen identifiziert, die benötigt werden, um synthetische Trainings-Datensätze zu erzeugen. Große Sprachmodelle, wie GPT-3, GPT-4 oder ChatGPT, basieren auf großen homogenen Daten an Texten, bestehend aus Wörtern, bestehend aus Buchstaben. Mit Hilfe von Repräsentations Deep Learning (wie bei OpenAI und ChatGPT) können für heterogene Daten diejenigen Features gelernt werden, die besonders relevant und effektiv als Digital Identifier und zusammengenommen als Daten-Konnektom dienen. Damit können die Daten, egal ob Bilder von Tumoren, Diagnosen in Form von Texten, Messreihen wie EKG so in einen hochdimensionalen Raum eingebettet werden, dass eine digitale hierarchische Repräsentation entsteht. Dies geht insofern über konventionelle Datenstrukturen hinaus weil Hierarchien für die Messung von kausalen Zusammenhängen dienen können. Damit wird nicht nur statisch komprimiert, sondern mittels KI erfasst, welche Features mit welchen anderen und wie genau zusammenhängen. Dieser Standard ermöglicht dann einen effektiven Austausch von Daten, weil es nicht mehr nötig ist, alles auf einen spezifischen Standard zu konvertieren, bevor die Daten ausgetauscht werden können. Darüber hinaus erwarten wir, dass sich mit Deep Learning aufgrund dessen Repräsentationen emergente Kausalitäts-Strukturen finden lassen werden. Dies wird dann Kollaborationen ermöglichen, die vorher undenkbar waren, oder bestehende suboptimale Datendialoge zwischen Institutionen und Datenzentren optimieren. Damit kann Effizienz auf allen Ebenen der Wissenschaft und sektoralen Gliederung des Staates geschaffen werden. Dies führt zu einer genuinen Transdisziplinarität.

### Aufgaben

- Hierarchische Repräsentation kodieren kausale Zusammenhänge. ← Systemarchitektur
- Etablierung von FDOs als universelle Schnittstellen (API), die alle Datenräume struktur analog kategorisieren und somit vergleichbar entwickeln, unabhängig von der speziellen Disziplin. ← Use Cases
- Kommunikation in einfacher Sprache. ← Gesellschaft

### Meilensteine

- Entwicklung einer genAI-Toolbox, die fähig ist, aus Surrogate-Daten bestimmte vorgegebene (kausale) Features aus heterogenen Datensätzen zu identifizieren.

### AP<sub>p</sub> 3.4. Quantentechnologie

---

<b>Start-up</b>	QBITFLOW GmbH Roland Becker, Hansjörg Dittus
<b>Mission</b>	Konzeption und Entwicklung von neuartigen Quantencomputern wie auch KI-basierten Anwendungen im Bereich des quantenmaschinellen Lernens
<b>Partner</b>	Hochschule Bremen, Patrick Draheim DESY, Karl Jansen Universität Saarbrücken, Giovanna Morigi
<b>Zielgruppe</b>	Industrie

---

### Ziel des Arbeitspakets

- Entwicklung und Anwendung von quantenspezialisierten FDOs (Q-FDO) mit dem Schwerpunkt von Quantenexperimenten für Algorithmen der Künstlichen Intelligenz und damit einhergehender und notwendiger Datenvorverarbeitungen. Die Q-FDO werden befüllt

durch Hyperparametrisierung der Lernmodelle, der Parametrisierung der Datenvorverarbeitung wie auch resultierender Performanz der quantengestützten Lernverfahren, so dass über die Metadateninformation des Q-FDO strukturierte Analysen der Experimente ermöglicht werden.

### **Aufgaben**

- Konzeptionalisierung der Ausgestaltung von Q-FDOs im Pilotprojekt hinlänglich Datenbestand, Hyperparametrisierung und Lernverhalten
- Durchführung der angedachten Quantenexperimente zur Generierung des benötigten Datenbestandes für den Q-FDO
- Annotation des zugrunde liegenden Datenbestandes und der Hyperparametrisierung und Überführung der Resultate in den Q-FDO

### **Meilensteine**

- Aufgezeichnete Quantenexperimente sind im Q-FDO abgelegt

## AP<sub>p</sub> 4. Gesellschaftliche Werkzeuge

Zentrale Aufgabe des Konsortiums FDO4DE soll es sein, eine transdisziplinäre Brücke über alle relevanten Disziplinen und Sektoren zu bauen. Integral für eine Zusammenführung von Informationen durch Multi-Indexierung (**transdisziplinäre Maßzahlen**), ist **semantische Analyse**, **Daten- und Makroökonomie**, eine enge und vorausschauende Kooperation des Infrastrukturaufbaues mit den Dimensionen des **Datenrechts**, sowie die akademische **Zukunftsforschung**. Dies sind Disziplinen, die bereits im Piloten als pivotale Elemente eines transdisziplinären Ansatzes dabei sein müssen. Ein Think Tank betreibt die Kommunikation mit dem Bürger und der Zivilgesellschaft, sowie der Wissenschaften untereinander.

### AP<sub>p</sub> 4.1. Semantische Analyse

<b>StartUp</b>	Sense H Hannover, Suat Can U Bremen, Uwe Engel
<b>Mission</b>	Web Data Mining & Analytics
<b>Partner</b>	U Bremen, Robert Porzel Frankfurt School of Finance, Jan Nagler TU Dresden, Marc Timme KIT, Frank Tristram
<b>Zielgruppe</b>	Politik, Gesellschaft & Datenräume

#### Ziele des Arbeitspakets

→ Soziale Medien stellen eine zentrale Quelle öffentlicher Meinungsbildung dar. Anders als standardisierte Umfragen produzieren sie qualitative Textdaten, in denen sich zu variablen Anteilen und Intensitäten Beschreibungen, Meinungen, Bewertungen und Emotionen mischen können. Diese Texte entstehen kontinuierlich und in erheblichem Umfang als Prozessdaten. Oftmals sind sie mit visuellen und auditiven Informationen verknüpft und können auch von Bots generierte Fälschungen sein. Darüber hinaus erfordern Textdaten aus Sozialen Medien ein sorgfältiges Datenmanagement zur Vorbereitung der Analyse sowie fortgeschrittene quantitative Analysemethoden, um sie in ihre vielfältigen Bedeutungskomponenten zu zerlegen.

Das Ziel des Arbeitspakets ist die Forschung und letztlich Entwicklung eines Minimum Viable Products (MVP), das den FDO-Ansatz nutzt, um Textdaten und deren Metadaten zu sammeln und diese mit fortschrittlichen Methoden aus den Bereichen des maschinellen Lernens und der Sprachverarbeitung in eine analysierbare Form umzuwandeln und analytisch in ihre Bedeutungskomponenten zu zerlegen. Eine weitere Funktion dieses MVP wird darin bestehen, aus qualitativen Textdaten soziodemografische Informationen zu gewinnen, die sonst nur durch Umfragen zugänglich wären. Drittens zielt das Vorhaben darauf ab, Textdaten aus sozialen Medien abstrakt-analytisch qua statistischem Matching mit Survey-Daten zu verknüpfen.

Umfrageforschung ist weltweit ein riesiger Markt, der sich selbst im Umbruch befindet: Einerseits entfernen sich gerade Online-Umfragen vom bisherigen Goldstandard (reine Wahrscheinlichkeitsauswahlen), andererseits bleiben Potentiale ungenutzt, die sich aus Datenverknüpfung und speziellen Sampling-Methoden ergeben. AP 4.1 reagiert auf diesen Umbruch und die ungenutzten Potentiale mit der Entwicklung und Erprobung eines Tools für Web Data Mining & Analytics.

## Aufgaben

- Entwicklung eines Tools für Web Data Mining & Analytics für den sich entwickelnden Markt für Computational Social Research.
- Explorative und feldexperimentelle Erprobungen dieses Instruments.

## Meilensteine

- Entwicklung eines FDO-basierten Tools für Web Data Mining & Analytics

## AP<sub>p</sub> 4.2. Transdisziplinäre Maßzahlen

<b>Start-up / Verein</b>	IndexNow (in Gründung) Oliver Rack
<b>Mission</b>	Zusammenführung von PID-markierten Informationen aus verschiedenen Objekten und Disziplinen zu transdisziplinären Maßzahlen durch Multivariante Indices ermöglichen.
<b>Partner</b>	Transparency International Deutschland e.V., Helena Peltonen-Gassmann Next Learning e.V., Oliver Rack U Bremen, Uwe Engel, Torben Klarl H Hannover, Suat Can Land Bremen, Senatskanzlei, Annette Lang (tbc) Land Hamburg, Staatskanzlei, Norbert Kuhn (tbc) Statistisches Landesamt Bayern (tbc) Statistisches Landesamt Bremen (tbc) Statistik Nord (Hamburg, Schleswig-Holstein) (tbc) Kommission zur Zukunft der Statistik, Walter Radermacher (LMU)
<b>Zielgruppe</b>	Bürger, Gesellschaft und Regierungen

## Ziele des Arbeitspakets

- In dem gemeinnützigen Verein **Index Now e.V** bringen unsere Stimmen der Zivilgesellschaft, getragen von *Transparency, International Deutschland* und *Next Learning*, wesentliche funktionale Element in Ausgestaltung und Design des gesellschaftlichen Werkzeuges Indexierung in die Wertschöpfung ein. Sie arbeiten eng mit dem **KIT Start-up Universal Data Interface** zusammen, um ausgewählte Pilotprojekte zur Bereitstellung von Indices relevanter sozio-kultureller, ökologischer und ökonomischer Maßzahlen zu realisieren und mit FIZ Karlsruhe zu repositionieren. Das Start-up **Coherence** liefert eine systematische Koordination und die universale Struktur der Indexerstellung und Erhebung in Kollaboration mit dem Datenraum Komplexe Systeme und dem statistischen Verbund. Es leistet eine transdisziplinäre Vermittlung und Definition der Begrifflichkeiten im Kontext unter den Disziplinen, zB als Grundlage emanzipatorischer Infrastrukturen und als konzeptionelles Design des Gesamtprozesses.
- Mit der Strukturierung der Indexierung über diverse Themenfelder öffnen sich diese für eine Kommerzialisierung der Zuarbeit für staatliche Verwaltungen und Behörden, sowie für Beratung von Wirtschaft und Industrie. Wir werden dieser erwarteten Entwicklung durch Gründung von **Index Now Commercial** Rechnung tragen.
- Multi-Indizierte öffentliche Datensätze sind als (Open Data) Grundlage einer wissensbasierten offenen Regierungsarbeit (Open Government) und einer informierten Gesellschaft.

Unser Ziel ist es, die FDOs in den 5. Nationalen Aktionsplan im Rahmen der Teilnahme Deutschlands an der Open Government Partnership (OGP) für 2025-2027 einzubringen.<sup>19</sup>

## Aufgaben

- Fachliche Konstruktion der Indices
- Definition eines individuellen Preissteigerungsindex (Inflation) und eines Index zur lokalen Biodiversität von Vögeln (im urbanen Raum) in Zusammenarbeit mit dem Verbund Statistik
- Quantifizierung der Nachhaltigkeitsziele der UN (SDGs) adaptiert auf die Stadtstaaten Bremen und Hamburg im Zusammenarbeit mit den senatorischen Behörden<sup>20</sup>
- Evaluierung der FAIRness in Deutschland durch Befragung einzelner Institutionen.

## Meilensteine

- Etablierung einer universellen Struktur der Projekte im Piloten

## AP<sub>p</sub> 4.3. Recht der Digitalisierung

---

<b>Koordination</b>	Datenrecht Dennis-Kenji Kipker (IT Sicherheit) David Wagner (Datenschutz)
<b>Mission</b>	Einbettung der FDO-Architektur in die datenrechtlichen Rahmenbedingungen; Abstimmung zwischen technischen Lösungsansätzen und rechtlichen Vorgaben; Ableitung rechtspolitischer Forderungen und gesetzgeberischer Konsequenzen aus der FDO-Mission.
<b>Partner</b>	U Augsburg, Benedikt Buchner, Dennis-Kenji Kipker Hochschule Bremen Kanzlei NN Universität Bremen, Uwe Engel, Lorenz Kähler Universität der Bundeswehr München, Mario Martini (tbc), David Wagner
<b>Zielgruppe</b>	Regelungsadressaten (Bürger:innen, Verwaltung, Unternehmen); Politik und Gesetzgebung auf Landes-, Bundes und EU-Ebene; rechtswissenschaftliche Literatur und Rechtsprechung

---

## Ziel des Arbeitspakets

- Disruptive digitale Entwicklungen geraten typischerweise mit der bestehenden Rechtsordnung in Konflikt. Dies haben jüngst zahlreiche Klagen unterstrichen, die sich gegen Entwickler von KI richten: Die Kläger gehen auf Grundlage des Urheberrechts und des Datenschutzrechts gegen die Gewinnung und die Nutzung von Trainingsdaten vor, siehe noyb und Zeit Online. Haben diese Klagen Erfolg, sehen die betroffenen LLMs einer ungewissen Zukunft entgegen. Denn das sog. „unlearning“ von LLMs steckt noch in den Kinderschuhen. Es ist diese Art von Sand in dem Getriebe, die FDO4DE zu vermeiden sucht, indem das Projekt von Beginn an auf eine umfassende datenrechtliche Begleitung setzt. Damit beugt es auch technischen Designentscheidungen vor, die sich später nur mit erhöhtem Aufwand umkehren lassen. Deswegen beginnen wir bereits im Piloten mit einer grundlegenden Betrachtung aller im Förderkontext relevanten Rechtsbereiche und ihrer Implikationen für

<sup>19</sup>Hier der laufende Plan für 2023-2025

<sup>20</sup>Sustainable Development Goals

FDO's. Dabei sind Aspekte der Datensicherheit ebenso essentiell, wie Datenzugangs- und Nutzungsregime, die jüngst auf europäischer Ebene in Gestalt des Data Act, des Data Governance Act und dem European Health Data Space (EHDS) entstanden sind. Die neuen Datenzugangs- und Nutzungsregime stehen auf dem Fundament von einem ganzen Kranz an Rechtsakten, mit denen der Gesetzgeber Geschäftsgeheimnisse, Urheber und Betroffene bei der Verarbeitung ihrer personenbezogenen Daten schützt. Die zentralen datenrechtlichen Herausforderungen, das bisherige Fehlen einheitlicher Standards, Formate und Verfahren für eine gemeinsame Datennutzung, muss durch eine möglichst passgenaue symbiotische Abstimmung der FDO-Architektur auf die neuen und bekannten datenrechtlichen Regulierungsansätze angegangen werden. Gleichzeitig bedarf jegliche aktuelle und künftige Neuregulierung im Datenrecht einer rechtspolitischen Begleitung, indem das Innovationspotential einer FDO-Architektur bei jeder Gesetzgebung so weit wie möglich Berücksichtigung findet, um insbesondere für künftige Vorhaben im Datenrecht weitere Innovationspotenziale zu heben. Darüber hinaus stellen sich verschiedene datensicherheitsrechtliche Fragestellungen im Projektkontext: Die technischen, organisatorischen und rechtlichen Anforderungen an die sichere Datenhaltung werden immer größer, was eine disziplinenübergreifende und rechtssystematische Betrachtung voraussetzt.

### Aufgaben

- Systematisierung des datenrechtlichen Anwendungsrahmens im konkreten Förderkontext
- Validierung ermittelter Anforderungen unter Berücksichtigung von projektbasierter Technologiegestaltung
- Ableitung rechtspolitischer Anpassungsszenarien zur Optimierung datenbezogener Rechtsgrundlagen

### Meilensteine

- Road Map europäisches und nationales Datenrecht
- Datenrechtlicher Synergieraport zur Ermittlung optimaler Potenziale der Technikgestaltung

### AP<sub>p</sub> 4.4. Makroökonomie

---

<b>Werkzeug</b>	Datenökonomie U Bremen, Torben Klarl
<b>Mission</b>	Entwicklung und (empirische) Analyse eines dynamischen allgemeinen Gleichgewichtsmodells zur Analyse anreizkompatibler Wirtschaftspolitik für eine Datenökonomie mit technologischen Standards

---

<b>Partner</b>	FDO4DE
----------------	--------

---

<b>Zielgruppe</b>	Politik & Gesellschaft
-------------------	------------------------

---

### Ziel des Arbeitspakets

- In diesem Projekt sollen Standards aus einer makroökonomischen Wachstumsperspektive untersucht werden.
- Insbesondere sollen die Anreize zur Produktion neuer Technologien auf Basis von Standards gegenüber der Nutzung von Standards im Endgutsektor herausgearbeitet werden. Weiterhin wird untersucht, unter welchen Bedingungen Standards einen Effekt auf das langfristige Wachstum haben.
- Zudem werden Wohlfahrtseffekte von Standards herausgearbeitet.

## Aufgaben

- Aufbereitung der Literatur und Modellierung des Grundmodells
- Modellerweiterung und Ableitung wirtschaftspolitischer Handlungsempfehlungen
- Simulation der Modelle
- Datensatzerstellung
- Empirische Validierung

## Meilensteine

- Entwicklung eines sog. semi-endogenen Wachstumsmodells in welchem ökonomische Aspekte von Standards im Forschungs- und Entwicklungsprozess integriert werden
- Quantitative Modellevaluation und empirische Validierung des Modells.
- Herausarbeitung der Bedingungen, wie die Standardisierung von Technologien das langfristige Wachstum erhöhen kann.
- Erforschung, wie Wirtschaftspolitik ausgestaltet sein muss, um eine effiziente Marktlösung zu erreichen. Auch zu diesem Punkt wird das Projekt eine Antwort liefern, insbesondere um der Wirtschaftspolitik Handlungsempfehlungen anzubieten.

## AP<sub>p</sub> 4.5.Think Tank

---

<b>Veranstaltung</b>	Think Tank U Bremen, HG Döbereiner, Uwe Engel, Lorenz Kähler
<b>Mission</b>	Reflektion und Interaktion aufgeworfenen Thematiken in einem Think Tank.
<b>Partner</b>	FDO4DE
<b>Zielgruppe</b>	Politik & Gesellschaft

---

## Ziel des Arbeitspakets

- Stabile Demokratie durch kuratiertes Wissen

## Aufgaben

- Organisation von Veranstaltungen zur Diskussion mit der Zivilgesellschaft und dem Bürger
- Transdisziplinärer Austausch zwischen den Wissenschaften

## Meilensteine

- Gesellschaftliches Bewusstsein für nachhaltige Informationsinfrastruktur.

## AP<sub>p</sub> 4.6. Strategische Vorausschau

---

<b>Verein</b>	D2030 Klaus Burmeister
<b>Mission</b>	Eröffnung eines Diskussionsraumes für Zukunftsentwicklungen in Deutschland
<b>Partner</b>	Scenario Management International AG, Foresightlab, FDO4DE
<b>Zielgruppe</b>	Politik & Gesellschaft

---

### Ziel des Arbeitspakets

- Der gemeinnützige Verein *D2030 - mehr Zukunft wagen* entwickelt auf der Grundlage der Methodik der strategischen Vorausschau mögliche Leitszenarien der gesellschaftlichen und politischen Entwicklung Deutschlands, die den Hintergrund der Arbeit unseres Konsortiums bilden.

### Aufgaben

- Entwicklung von Szenarien im Rahmen eines Open-Foresight-Projektes unter Mitwirkung eines rund 50-köpfigen Teams, bestehend aus Zukunftsforschenden, sowie Experten und Expertinnen aus Wissenschaft, Wirtschaft, Verwaltung und Zivilgesellschaft. Eine breitere Öffentlichkeit wird zudem in zwei Online-Dialogen beteiligt.

### Meilensteine

- Das Projekt *Neue Horizont 2045* liefert Leitszenarien für Deutschland bis 2045, die sich nach ihrer politischen Grundhaltung unterscheiden: i) Szenarien, die zwar eine Veränderung postulieren, aber wesentliche Grundlagen beibehalten ii) solche die skizzieren, wie Zukunft mit einem neuen Wohlstands- und Wirtschaftsmodell aussehen könnte und schlieslich iii) jene, in denen die Veränderungsbereitschaft eher gering ist und stattdessen Stabilität und Sicherheit im Vordergrund stehen.

## Budget Pilot: Not for public distribution!

Entwurf des Budgets: <https://nc.uni-bremen.de/index.php/s/TWby5kRnR6TWzKZ>

Entwurf

## Informationstheoretischer Hintergrund

### Duale Hierarchie von Daten und Disziplinen

Jeder Wissenschaft oder Technologie ist ein bestimmter Typus von Daten zugeordnet, der sich durch charakteristische Eigenschaften beschreiben lässt. Diese Eigenschaften lassen sich hierarchisch ordnen in Ebenen zunehmender Komplexität. Es entsteht ein duales Gebilde aus hierarchischen Datenstrukturen und diese Daten erzeugenden Wissenschaften, siehe die grafische Darstellung zur kausalen Emergenz.

### Transdisziplinarität: Von Syntax zu Semantik

Der durch KI induzierte Brückenschlag zwischen den Disziplinen, lässt sich nur durch eine transdisziplinäre Arbeitsweise verstehen und erklären. Transdisziplinär bedeutet, ausgehend von der Methodik einer Wissenschaft, die Phänomene einer anderen Wissenschaft zu verstehen. So genannte qualitative und quantitative Wissenschaften lassen sich durch die Quantifizierung des Qualitativen, d.h. durch semantische Metriken verbinden. Semantische Metriken ordnen einem Inhalt Bedeutung zu, indem sie der inhaltlichen Information eine Entscheidung (Handlung oder Intention, d.h. zeitverzögert und eventuell konditionale Handlung) zuweisen und diese je nach Anzahl der Möglichkeiten bewerten.

### Konzeptionelles Design einer nachhaltigen Datenökonomie

Wir begreifen die Welt durch Wahrnehmung und Kreation als hierarchische Dualität von Daten und ihnen zugeordneten Wissenschaften oder erzeugenden Technologien. Wir bauen auf Luciano Floridis' Logik und Philosophie der Information. Floridi bereitet eine konstruktivistische Grundlage für Kuration von Information als konzeptionelles Design. Damit vollzieht Floridi einen Paradigmenwechsel von einem repräsentativen zu einem konstruktiven Verständnis von Wissen.

*Der Mensch als Schöpfer von Wissen nutzt eine Logik des Machens, d.h., eine Logik der Gestaltung der semantischen Artefakte, für die er epistemisch verantwortlich ist.*

Konzeptuelles Design, so Floridi, ist eine Logik der Anforderungen, die es uns ermöglicht, über Faustregeln und Rezepte für beste Praktiken hinauszugehen. Wir gestalten FAIRe Daten konzeptionell, um Wissen auf nachhaltige Weise zu konstruieren und zu kuratieren.

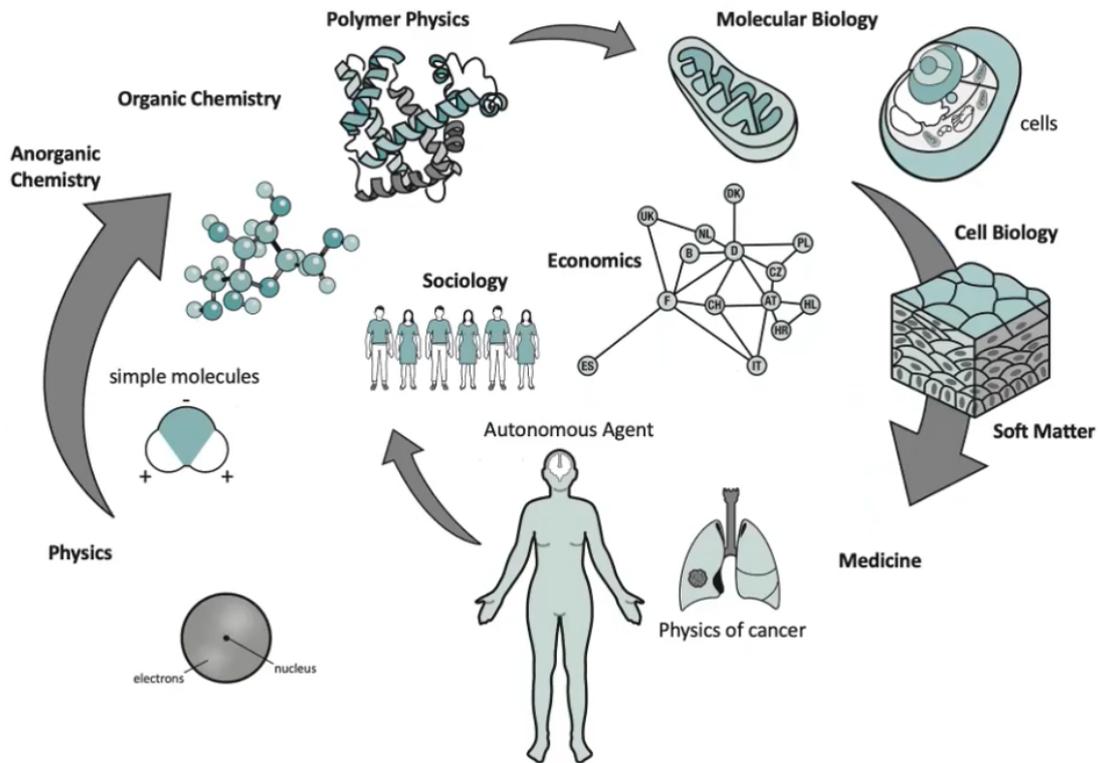


Abbildung 2: Kausale Emergenz erzeugt in der Natur eine hierarchische Struktur, die sich durch sukzessiv höhere Informationsebenen beschreiben lässt. Auf jeder Ebene bilden sich diejenigen effektiven Objekte (physikalisches Konzept der Quasiteilchen), die maximal kausal miteinander wechselwirken. Auf jeder Ebene entsteht so eine neue Eigenschaft, die sich durch FDOs darstellen lässt. Physik beschreibt fundamentale Wechselwirkungen zwischen Teilchen. Chemie betrachtet Reaktionen zwischen Molekülen. Biologie untersucht die Funktionen von Makromolekülen oder Zellen. Ein Ingenieur erzeugt aus Teilen eine neue Struktur. Symbole, die sich durch syntaktische Strukturen in Materie (Körper) codieren lassen, erhalten durch aktionsbasierte Semantik eine Bedeutung (Geist). Das ist das Prinzip eines Verkehrsschildes. Die Geisteswissenschaften beschäftigen sich also nicht mit den Dingen selbst, sondern mit der Bedeutung der Dinge, die für autonome Agenten entsteht, wenn diese handeln oder dies beabsichtigen.