FAIR DATA PUBLISHER

Strategie für den flächendeckenden Einsatz von FAIR Digital Objects

Im Auftrag von MISSION KI Nationale Initiative für Künstliche Intelligenz und Datenökonomie

FDO Connect wird im Rahmen der Initiative MISSION KI – Nationale Initiative für Künstliche Intelligenz und Datenökonomie entwickelt. MISSION KI wird vom Bundesministerium für Digitales und Staatsmodernisierung bis Ende 2025 gefördert und von acatech – Deutsche Akademie der Technikwissenschaften umgesetzt.

MISSION KI Nationale Initiative für Künstliche Intelligenz und Datenökonomie

acatech DEUTSCHE AKADEMIE DER TECHNIKWISSENSCHAFTEN

Projektpartner

GWDG Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen

://fdo/connect

gefördert durch

Bundesministerium für Digitales und Staatsmodernisierung

aufgrund eines Beschlusses des Deutschen Bundestages

INHALT 03

EXPOSITION 05

MISSION KI TREIBT DEN FAIR DATA PUBLISHER VORAN 06
STRATEGISCHE EINBETTUNG BEI ACATEC UND MISSION KI 06
WEITERENTWICKLUNG DES FAIR DATA PUBLISHER 07
DIE ÜBERBRÜCKUNG ZWEIER WELTEN UND NEUE FUNKTIONEN 07
FAIR DIGITAL OBJECTS MACHEN DARK DATA NUTZBAR 08
FAIR DIGITAL OBJECTS FÜR EUROPA 10
RISIKEN FEHLERHAFTER DATENINTERPRETATION 11

DAS KONZEPT DER FAIR DIGITAL OBJECTS 12
FAIR-PRINZIPEN 14
TREIBENDE AKTEURE ZU FAIR DIGITAL OBJECTS 16
WIE UNTERSCHEIDEN SICH FAIR DIGITAL OBJECTS ZU ANDEREN STANDARDS UND ERGÄNZEN DIESE? 17
FAIR DIGITAL OBJECTS IN BEZUG AUF INTERNET, DATENRÄUME UND INTELLIGENZ 18

FAIR DATA PUBLISHER 21

FDO-SERVICE-ARCHITEKTUR 22
KONZEPT UND IMPLEMENTIERUNG VON NANOPUBLIKATIONEN ALS FDOS 22
FDO MANAGER: NUTZUNGSOBERFLÄCHE UND SDK 23
AUTOMATISIERTE METADATEN-EXTRAKTION 23
DEZENTRALER ABFRAGEDIENST 26
ADAPTER VERBINDEN DIE SYSTEMWELTEN 26
FDO-PROFILE UND TYPE REGISTRY 26
AUTONOME FDOS (AFDOS) 28
QUERY-OPERATIONEN 28
INTEGRATION MIT DATENBANKEN UND DATENRÄUMEN 28
FAZIT UND EMPFEHLUNGEN 31
EVALUIERUNG DER TYPING-MECHANISMEN 34

TOOL HEBT UNSTRUKTURIERTEN DATENSCHATZ 36
TECHNOLOGISCHE KOMPONENTEN DER DATENEXTRAKTION 36
PROFILE-MATCHING UND FDO-GENERIERUNG 36
NUTZUNGSSCHNITTSTELLE UND INTERAKTION 38
OFFENER QUELLCODE, DOKUMENTATION UND EVALUIERUNG 38
FAZIT UND EMPFEHLUNGEN 39

USE CASES 41

DIGITALE MEDIZIN 42
AKTUELLE HERAUSFORDERUNGEN IN DER DIGITALEN MEDIZIN 43
ZIELE UND LÖSUNGSANSÄTZE DES ANWENDUNGSFALLS 43
DIE FDO-ARCHITEKTUR FÜR DIE UMSETZUNG DER EU-VORGABEN 44
INFORMATIONEN AUS SYNTHETISCHEN PATHOLOGISCHEN BEFUNDEN 44
GENERIERUNG VIRTUELLER GEWEBEFÄRBUNGEN 45
FAZIT UND EMPFEHLUNGEN 46

INDUSTRIELLE FERTIGUNG 49
DAS PROJEKT „BATCAT“ ALS DEMONSTRATOR 50
DIE ROLLE DES FDO-BASIERTEN KONVERTIERUNGSSERVICES 50
FDOS ALS ERMÖGLICHER VON INTEROPERABILITÄT UND MACHINE ACTIONABILITY 51
ANBINDUNG AN DIE ASSET ADMINISTRATION SHELL 52
STRUKTUR UND TECHNISCHE UMSETZUNG DES DEMONSTRATORS 52
KOMMERZIALISIERUNGSSTRATEGIEN UND WERTSCHÖPFUNG DURCH FDOS 54
FAZIT UND EMPFEHLUNGEN 54

ROBOTIK 56
DAS ROBIVAL-PROJEKT ALS DATENGRUNDLAGE 57
EXPERIMENTAL RESEARCH ONTOLOGY (ERO) 57
FDO-PROFILE FÜR ROBIVAL FDO-RECORDS 57
SOFTWARE UND AUTOMATISIERUNG 58
DATENTRANSFER UND DATENBANKINTEGRATION 58
IMPLEMENTIERTE PRODUKTE IM ANWENDUNGSFALL ROBOTIK 58
GRENZEN UND AUSBLICK 58
FAZIT UND EMPFEHLUNGEN 60

KOMPLEXE SYSTEME 63
DIE LÖSUNGSSTRATEGIE: FDOS UND KAUSALE INFERENZ 64
TECHNISCHE REALISIERUNG VON FDO-KAPSELN UND PCMCI-ALGORITHMUS 64
VALIDIERUNG MITTELS SURROGATDATEN 65
ERGEBNISSE DES ANWENDUNGSFALLS „KOMPLEXE SYSTEME“ 65
FAZIT UND EMPFEHLUNGEN 66

QUANTEN COMPUTING 67
DAS Q-FDO KONZEPT 68
INTEGRATIONSASPEKTE UND SYNERGIEN 70
VORTEILE DER Q-FDO NUTZUNG 70
FAZIT UND EMPFEHLUNGEN 72

STRATEGISCHE BETRACHTUNGEN 75

RECHTLICHE EINORDNUNG VON FDOS 76
FDO IM ALLGEMEINEN RECHTSRAHMEN 76
INTEROPERABILITÄT DURCH
TECHNISCHE STANDARDISIERUNG 76
DATENSCHUTZRECHTLICHE POTENZIALE
UND HERAUSFORDERUNGEN 77
CYBERSICHERHEIT IN DER MEHRSTUFIGEN
REGULIERUNGSARCHITEKTUR 77
KI-GOVERNANCE DURCH STRUKTURIERTE DATEN ...78
KI-INTEGRATION IN FDO-INFRASTRUKTUREN.....78
REGULATORISCHE ANFORDERUNGEN ALS
ENTWICKLUNGSPFAD – AUSGEWÄHLTE BEISPIELE ..79
DIGITAL SERVICES ACT: ARCHITEKTURBEDINGTE
REGULIERUNGSPFADE 80
FAZIT UND EMPFEHLUNGEN 81

FDO-INTEGRATION IN DIE
ÖFFENTLICHE REGISTERLANDSCHAFT 83
HERAUSFORDERUNGEN DER DATENLANDSCHAFT 83
DAS POTENZIAL VON FAIR DIGITAL OBJECTS 83
TECHNISCHE ASPEKTE
UND SICHERUNGSMECHANISMEN 83
RECHTLICHER RAHMEN UND HERAUSFORDERUNGEN ..83
DATENSCHUTZRECHTLICHE ASPEKTE
UND PERSÖNLICHKEITSPROFILE 83
MASSNAHMEN ZUR RISIKOBEGRENZUNG 84
ANFORDERUNGEN AN EINE
RECHTSSICHERE IMPLEMENTIERUNG 84
VERORDNUNGSERMÄCHTIGUNG
UND DATENARCHITEKTUR 84
FAZIT UND EMPFEHLUNGEN 85
EXTENDED FAIR DIGITAL OBJECTS (XFDO).....87

FDO-ANWENDUNGSFALL
NUTZUNGSRECHTE-MANAGEMENT 88
ENTWICKLUNG EINES KONZEPTS ZUR
VERWALTUNG ALLGEMEINER NUTZUNGSRECHTE 88
ABBILDUNG VON URHEBERRECHTEN
UND GESETZLICHEM STATUS 89
TECHNISCHE IMPLEMENTIERUNG
DES POLICY FRAMEWORKS 90
VALIDIERUNG UND ZUKÜNFTIGE ARBEITEN 92
FAZIT UND EMPFEHLUNGEN 93

ÖKONOMISCHE BETRACHTUNGEN 95
ÖKONOMISCHE AUSWIRKUNGEN
EINER FDO-BASIERTEN DATENÖKONOMIE 95
FAZIT UND EMPFEHLUNGEN 98

BETRIEB EINER FDO-SERVICE-ARCHITEKTUR 99
SPEZIALISIERTE GESCHÄFTSMODELLE
FÜR JEDEN AKTEUR 99
CHARAKTERISTIKA ETABLIERTER DATENRÄUME 99
FDOS ALS TEIL EINER ÖFFENTLICHEN
DATENINFRASTRUKTUR 100
FAZIT UND EMPFEHLUNGEN 102

TRANSFERSTRATEGIE FÜR NORMUNG
UND STANDARDISIERUNG 103
FAZIT UND EMPFEHLUNGEN 106

STRATEGISCHE ZUSAMMENFASSUNG
UND ANREGUNGEN 107
NOTWENDIGE WEITERE SCHRITTE,
DIE SICH DIREKT AUS DEM PROJEKT
FAIR DATA PUBLISHER ERGEBEN 107
ALLGEMEINE ANREGUNGEN
ZU WEITEREN SCHRITTEN 112

IMPRESSUM 119

EXPOSITION

Mission KI treibt den FAIR Data Publisher voran

Der FAIR Data Publisher stellt eine zentrale technologische Initiative dar, um die Auffindbarkeit, Zugänglichkeit, Interoperabilität und Nachnutzbarkeit von Daten bedeutend zu verbessern. Dieses Projekt ist nicht nur ein technisches Werkzeug, sondern trägt zu einer wichtigen Entwicklung in der deutschen und europäischen Dateninfrastruktur bei, die von einem Konsortium aus mehreren namhaften Organisationen gemeinsam vorangetrieben wird.

¬ Das Herzstück des FAIR Data Publisher ist die Implementierung von FAIR Digital Objects (FDOs). Ein FDO ist ein digitaler Container, der nicht nur die eigentlichen Daten enthält, sondern diese untrennbar mit reichhaltigen Metadaten und einem persistenten Identifikator (PID) verknüpft. Dieser PID, ähnlich einer ISBN für Bücher, sorgt dafür, dass das Datenobjekt dauerhaft und eindeutig adressierbar bleibt, selbst wenn sich sein Speicherort ändert.

Der FAIR Data Publisher ist ein Satz an Softwares, die den Prozess der Erstellung, Verwaltung und Veröffentlichung dieser FDOs vereinfachen. Er bietet eine benutzerinnen- und benutzerfreundliche Schnittstelle, die es auch Fachexpertinnen und -experten ohne tiefgreifende IT-Kenntnisse ermöglicht, ihre Daten nach den FAIR-Prinzipien aufzubereiten und in verteilten Datenräumen sicher und nachvollziehbar zu teilen. Er fungiert als Einstiegspunkt in eine vernetzte und intelligente Datenökonomie.

STRATEGISCHE EINBETTUNG BEI ACATECH UND MISSION KI

Das Projekt MISSION KI – Nationale Initiative für Künstliche Intelligenz und Datenökonomie der acatech – Deutsche Akademie der Technikwissenschaften, gefördert durch das Bundesministerium für Digitales und Staatsmodernisierung dient als zentrales Hebelprojekt der Digitalstrategie, um die digitale Wettbewerbsfähigkeit Deutschlands zu stärken und eine vertrauenswürdige KI „Made in Germany“ zu etablieren.

Die Initiative konzentriert sich auf drei strategische Säulen: Erstens die Verbesserung der Datenbasis für KI durch die Förderung der Vernetzung von Datenräumen über Sektoren- und Ländergrenzen hinweg sowie die Entwicklung von Open Source Services zur besseren Auffindbarkeit und Nutzung von Daten. Zweitens die Stärkung der Entwicklung vertrauenswürdiger KI durch die Implementierung eines freiwilligen, praxistauglichen KI-Qualitätsstandards und die Einrichtung von KI-Innovations- & Qualitätszentren, um das Vertrauen der Anwender zu erhöhen. Drittens die Unterstützung des Wachstums von KI-Innovationen durch Programme wie das AI Founder Fellowship und die gezielte Zusammenführung von KMU und KI-Startups, um den Transfer von Spitzenforschung in die wirtschaftliche Praxis zu beschleunigen.

Im Rahmen der Säule 1 hat sich MISSION KI zur Förderung der Vernetzung von Datenräumen der Entwicklung einer Architektur auf Basis von FAIR Digital Objects verschrieben und den FAIR Data Publisher ins Leben gerufen. Acatech treibt bereits seit Jahren die Vision einer souveränen und föderalen Dateninfrastruktur für Deutschland und Europa voran, prominent sichtbar im Gaia-X-Projekt. Der FAIR Data Publisher ist eine konkrete, praxisnahe Umsetzung dieser Vision. Er liefert die grundlegende Technologie, um Datenräume mit hochwertigen, interoperablen und vertrauenswürdigen Daten zu füllen.

Für die MISSION KI ist der FAIR Data Publisher von essenzieller Bedeutung. Künstliche Intelligenz ist auf riesige Mengen qualitativ hochwertiger und zugänglicher Trainingsdaten angewiesen. Oft scheitern KI-Projekte am Mangel solcher Daten oder an dem enormen Aufwand, heterogene Datensätze nutzbar zu machen. Indem der FAIR Data Publisher Daten von vornherein in einem standardisierten, maschinenlesbaren und interoperablen Format bereitstellt, legt er das Fundament für die Entwicklung und Skalierung von KI-Anwendungen in Deutschland.

WEITERENTWICKLUNG DES FAIR DATA PUBLISHERS

Das aktuelle Projekt FDO Connect, auf das sich dieser vorliegende Bericht bezieht, stellt nach FDO One, die zweite Entwicklungsphase des FAIR Data Publishers dar. Die Ergebnisse werden im Anschluss open source veröffentlicht.

FDO Connect setzt auf den Vorarbeiten von FDO One auf und konzentriert sich darauf, die etablierte Handle-basierte FDO-Infrastruktur unter anderem mit der Technologie der Nanopublikationen zu verknüpfen. Im Rahmen von FDO One wurde zunächst der FDO Manager entwickelt, der nun für den FAIR Data Publisher umfassend erweitert wurde. Dieser Manager dient als primäre „Oberfläche“ für menschliche Anwenderinnen und Anwender. Im Rahmen des viermonatigen Projekts FDO Connect wurde nun der FDO Manager unter anderem in die Lage versetzt, FDOs im Nanopublikations-Netzwerk anzulegen.

Schon bei FDO One diente der Mobility Data Space als begleitender Anwendungsfall zur Evaluierung von FDOs. Dieses Konzept – die unmittelbare Verzahnung von Infrastrukturentwicklung und praktischer Erprobung der Technologie in domänenspezifischen Kontexten – wurde bei FDO Connect im Rahmen des FAIR Data Publishers nicht nur fortgesetzt, sondern systematisch erweitert: Die Entwicklungen wurden parallel anhand konkreter Anwendungsfälle in den Bereichen Industrielle Fertigung, Robotik, Digitale Medizin und Quantencomputing als Proof of Concepts erprobt. Diese Vorgehensweise ermöglicht es, bereits während der Entwicklung Erkenntnisse zur jeweiligen Funktionsweise zu sammeln und die Technologie anzupassen. Zugleich liefert sie unmittelbare Aufschlüsse über längerfristige Anpassungsbedarfe und weitere notwendige Schritte in technischer wie rechtlicher Hinsicht – Erkenntnisse, die als Vorausschau in diesen Bericht einfließen konnten. In der technischen Tiefe wurde ein „Documentation-as-Code“-Ansatz angewendet.

Ermöglicht wurde dieses Projektdesign durch den Zusammenschluss des Teams aus FDO One mit der Initiative FDO4DE zu FDO Connect. FDO4DE widmet sich insbesondere der Identifizierung und Entwicklung von Anwendungsfällen für FDOs und wurde von der Universität Bremen ins Leben gerufen. Aus ihr ist die Mehrzahl der Use Cases in diesem Projekt hervorgegangen. Dadurch wurde auch eine starke interdisziplinäre Zusammenarbeit wichtiger Akteure aus der Entwicklung von FAIR Digital Objects verwirklicht.

DIE ÜBERBRÜCKUNG ZWEIER WELTEN UND NEUE FUNKTIONEN

Im Rahmen des jüngsten Projekts FAIR Data Publisher wurden nun zwei elementare Netzwerke zur Weiterentwicklung von FDO zusammengeführt und deren Interoperabilität sichergestellt: das Handle-Netzwerk und das Nanopublikations-Netzwerk. FDOs lassen sich nun mit denselben Werkzeugen und derselben Basisarchitektur auch im Nanopublikations-Netzwerk erstellen, ohne dass Benutzerinnen und Benutzer den technischen Unterschied zwischen Handle-basierten und Nanopublikations-basierten FDOs kennen müssen.

Die Idee der Nanopublikationen entstand Anfang der 2010er-Jahre im Umfeld der Wissenschaftskommunikation und der Forschung zum Semantischen Web. Ziel war es, wissenschaftliches Wissen nicht nur in langen Artikeln zu dokumentieren, sondern in kleinsten, maschinenlesbaren Einheiten verfügbar zu machen. Eine Nanopublikation beschreibt im Kern eine einzelne, überprüfbare wissenschaftliche Aussage – ergänzt durch Metadaten über Herkunft, Kontext und Vertrauenswürdigkeit. Dabei stellt eine Nanopublikation anhand reicher Metadaten bereits einen kleinen Wissensgraph dar.

EXPOSITION_FAIR DATA PUBLISHER

INTERVIEW MIT MANFRED RAUHMEIER, VORSTAND DER ACATECH STIFTUNG

>>Wieso setzt sich acatech so engagiert für das Thema FAIR Digital Objects ein?
<<Für ein leistungsstarkes Datenökosystem brauchen wir einen reibungslosen Datenaustausch zwischen Datenräumen. Der flächendeckende Einsatz von FAIR Digital Objects bietet eine große Chance, um Daten effizient und sicher auszutauschen.

>>Welche Chancen ergeben sich daraus für Deutschland?
<<Indem wir Datenschätze auf diese Weise heben, machen wir unsere Datenökonomie zukunftsfähig, treiben KI-Innovationen voran und stärken somit die Wettbewerbsfähigkeit unserer Unternehmen.

In der Welt der Nanopublikationen entspricht ein Metadateneintrag (Metadata Record) nicht nur einer einzelnen Nanopublikationen, sondern einer ganzen Versionshistorie oder Sequenz von Nanopublikationen. Jede Nanopublikation in dieser Kette verweist auf die vorige Version, wobei der neueste Eintrag den aktuellen Record repräsentiert. Der FDO Identifier, als Handle-PID (Persistent Identifier) vorliegend, leitet in der Nanopublikations-Welt immer zur neuesten Version der Nanopublikation weiter, während frühere Versionen über spezifische Anfragen, insbesondere mittels SPARQL, zugänglich sind.

Die technische Verknüpfung zwischen FDO One und Nanopublikationen erfolgt durch mehrere Komponenten. Ein Adapter-Modul auf Basis des DOIP-Protokolls wurde entwickelt, das die Verbindung zum Nanopublikations-Netzwerk herstellt. Des Weiteren wurde eine Nanopubs4FDO-Bibliothek konzipiert, die auf der bestehenden Nanopub-Softwarebibliothek aufbaut, um FDOs im Nanopublikations-Netzwerk verfügbar zu machen.

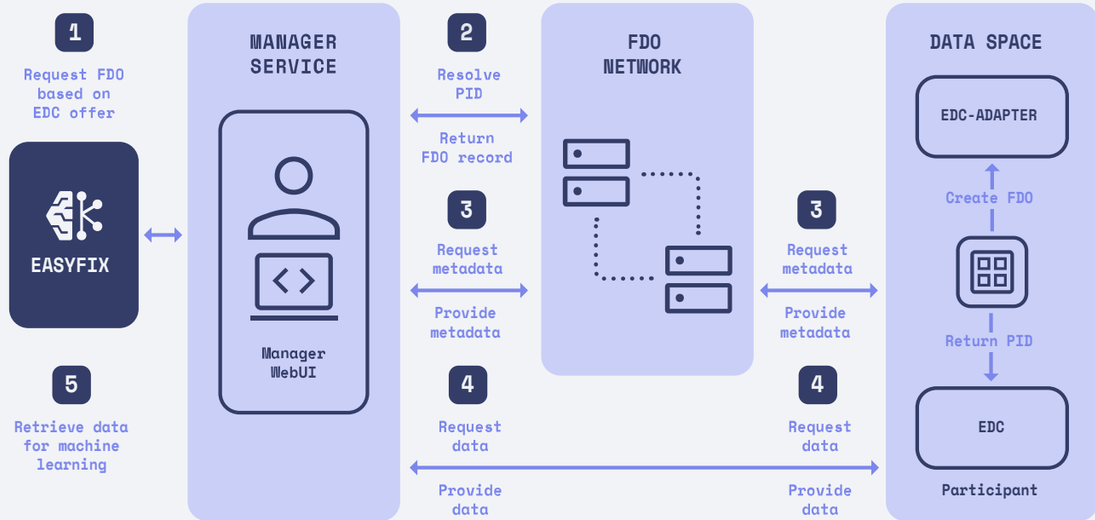
Zudem wurde ein Konverter (AP2) entwickelt, der als elementare Scharnierfunktion dient, um unstrukturierte Domänendaten, wie beispielsweise PDF-Dokumente, automatisch in strukturierte, domänenspezifische FAIR Digital Objects (FDOs) zu konvertieren. Dieser automatisierte Workflow ermöglicht die Transformation menschlich lesbarer Dokumente in maschinell verarbeitbare Daten und vereinfacht den FDO-Erstellungsprozess durch eine intuitive, browserbasierte Benutzeroberfläche. Zudem unterstützt das Tool eine KI-gestützte automatische Profilerkennung, welche Nutzenden passende FDO-Profile vorschlägt und dadurch den Bedarf an Expertenwissen reduziert.

Autonome FDOs (aFDOs) stellen ein besonders innovatives Element der FDO-Service-Architektur dar, da sie über interne Operationen verfügen und die Fähigkeit besitzen, eigenständig auf Ereignisse zu reagieren. Im Gegensatz zu klassischen FDOs, die Logik extern ausführen, sind aFDOs interaktiver und selbstständiger, da sie Operationen intern ausführen und Regeln für Zugriff und Interaktion verstehen.

FAIR DIGITAL OBJECTS MACHEN DARK DATA NUTZBAR

Derzeit wenden Menschen und Maschinen noch den Großteil ihrer Zeit und Mühen auf, Daten derart aufzubereiten, dass sie diese überhaupt verstehen und verarbeiten können. Kosten und Risiken entstehen dabei in mehreren Dimensionen. Unternehmen müssen erhebliche Ressourcen für die manuelle Datenbereinigung, -transformation und -integration aufwenden. Schätzungen zufolge verbringen Datenanalysten bis zu 80 % ihrer Arbeitszeit mit solchen Aufgaben anstatt mit wertschöpfender Analyse. Darüber hinaus entstehen Opportunitätskosten durch entgangene Geschäftsmöglichkeiten, da Entscheidungen auf Basis unvollständiger oder inkonsistenter Daten getroffen werden müssen. Forschungseinrichtungen können ihre Ergebnisse nicht effektiv teilen, Behörden arbeiten mit veralteten oder widersprüchlichen Informationen, und demokratische Prozesse leiden unter mangelhafter Datengrundlage. Moderne Data-First-Infrastrukturen bieten einen wesentlichen Nutzen, indem sie Defizite im Umgang mit wertvollen Daten beseitigen und die Entstehung von „Dark Data“, „Stranded Data“ und „Siloed Data“ verhindern. Dark Data erzeugen Speicherkosten, ohne als Entscheidungsgrundlage zu dienen, Stranded Data bleiben durch veraltete Formate unzugänglich, und Siloed Data entstehen durch isolierte Abteilungssysteme. FDOs lösen diese Probleme durch offene Formate, standardisierte Schnittstellen und einheitliche Metadaten.

FDOs zeichnen sich durch ihre umfangreichen Metadaten aus, die es ermöglichen, dass die Daten selbstbeschreibend sind – sowohl hinsichtlich ihres Kontextes als auch ihrer Verarbeitung. Diese Selbstbeschreibung macht die Daten automatisierbar und für Maschinen verstehbar. Durch die Integration in Wissensgraphen-Datenbanken und die Verwendung in digitalen Zwillingen



•FAIR DATA PUBLISHER ONE•

Die FDO-One-Version des FAIR Data Publishers schaffte bereits eine nahtlose Verbindung zwischen Datenräumen (Data Space) und maschinellem Lernen. Der Workflow beginnt mit der Anforderung eines FDO durch den Teilnehmer basierend auf EDC-Angeboten (1), gefolgt von der FDO-Erstellung und PID-Rückgabe (2). Anschließend erfolgt die PID-Auflösung mit Handle-Record-Rückgabe (3), die bidirektionale Metadatenanfrage und -bereitstellung zwischen Service- und Data Space-Ebene (3), sowie der entsprechende bidirektionale Datenaustausch (4). Der Prozess kulminiert im automatisierten Datenabruf für maschinelles Lernen (5).

EXPOSITION_FAIR DATA PUBLISHER

INTERVIEW MIT DR. SVEN BINGERT, GWDG

>>Welche Mehrwerte erzeugen FDOs für Forschung und die intelligente Vernetzung?
<<FDOs können neue Potenziale in einem global integrierten Datenraum erschliessen – beispielsweise durch den automatisierten Austausch von Forschungsdaten, plattformübergreifende Produktionsoptimierung oder die nahtlose Bereitstellung von Trainingsdaten für KI-Anwendungen.

>>Wie helfen FDOs der Dateninwertsetzung und KI-Systemen?
<<Die persistente Identifizierung von Informationen unter Berücksichtigung rechtlicher,vertrauens-bezogener und technologischer Aspekte ist eine notwendige Voraussetzung für die künftige Verwendung von Daten durch Nutzende und KI-Systeme. FDO-Prinzipien in einer verteilten Architektur ermöglichen zudem neue Anwendungen für den Austausch und die Wiederverwendung von Daten.

entfalten FDOs ihr volles Potenzial. Besonders bedeutsam sind dabei die emergenten Effekte, die entstehen, wenn Erkenntnisse über verschiedene Domänen und Disziplinen hinweg zusammengeführt werden. Dies schafft neue Möglichkeiten für maschinelles Transferlernen, bei dem Zusammenhänge und Kausalitäten aus einem Bereich auf andere übertragen werden können, und ermöglicht die Etablierung einer zuverlässigen „Ground Truth“ – einer fundierten Erkenntnisgrundlage, die als Referenz für weitere Analysen und Wissensproduktionen dient.

FAIR DIGITAL OBJECTS FÜR EUROPA

Die FAIR Digital Objects fügen sich als konzeptioneller und implementierungstechnischer Rahmen fundamental in die Europäische Digitalstrategie ein, insbesondere in die European Open Science Cloud (EOSC).

FDOs spielen auch eine unterstützende Rolle im Kontext der digitalen Souveränität und der PSI-Richtlinie (Richtlinie über die Weiterverwendung von Informationen des öffentlichen Sektors). Die Europäische Digitalstrategie strebt eine größere digitale Souveränität an, also die Fähigkeit Europas, in der digitalen Welt autonom zu handeln und zu gestalten. FDOs tragen dazu bei, indem sie eine neutrale, standardisierte und maschinenlesbare Grundlage für Forschungsdaten und andere digitale Objekte schaffen. Dadurch wird die Abhängigkeit von spezifischen, proprietären Plattformen oder Technologien aus Drittländern reduziert.

Das Once-Only Technical System (OOTS) ist eine weitere zentrale Komponente der digitalen Agenda, die im Rahmen der Single Digital Gateway-Verordnung (SDG-VO) implementiert wird. Das OOTS verfolgt das „Once-Only“-Prinzip, das sicherstellt, dass Bürger und Unternehmen bestimmte Standardinformationen nur einmal an eine öffentliche Verwaltung übermitteln müssen.

Somit bilden FDOs ein generisches Architekturprinzip, das die notwendige technische Basis für die gesamte Europäische Strategie legt, indem es sowohl die Wiederverwendung von Forschungsdaten (EOSC) als auch die effiziente und vertrauenswürdige grenzüberschreitende Verwaltung (OOTS) durch die Schaffung interoperabler, maschinenlesbarer digitaler Objekte unterstützt.

Risiken mangelnder Dateninteroperabilität

Fehlerhafte Dateninterpretation verursacht erhebliche wirtschaftliche und operative Schäden in nahezu allen Bereichen der modernen Gesellschaft. Diese Problematik entsteht, wenn verschiedene Systeme, Organisationen oder Personen Daten nicht einheitlich verstehen, austauschen oder verarbeiten können.

- **Beispielsweise drohen Schäden, wenn Patientendaten fehlerhaft interpretiert werden oder es entstehen Risiken bei Datenschutzverletzungen und Identitätsdiebstahl, wenn Systeme Berechtigungen oder Zugriffskategorien unterschiedlich interpretieren. Automatisierte Prozesse können fehlschlagen, wenn sie auf inkonsistente Dateneingaben treffen, was Kaskadenfehler in vernetzten Systemen auslösen kann.**

- **Bei der Einführung digitaler Fahrkartensysteme beispielsweise konnten verschiedene Verkehrsverbünde ihre Daten nicht austauschen. Reisende mussten für dieselbe Strecke mehrere separate Tickets kaufen, obwohl technisch ein durchgängiges Ticket möglich gewesen wäre. Dies führte zu Mehrkosten für Verbraucher und reduzierter Attraktivität des öffentlichen Verkehrsangebots.**

- **Bürgerinnen und Bürger müssen identische Informationen bei verschiedenen Behörden separat einreichen, da die Systeme nicht interoperabel sind.**

- **Wissenschaftliche Studien können nicht effektiv verglichen werden, wenn Forschungsdaten in unterschiedlichen, inkompatiblen Formaten vorliegen. Dies verzögert wissenschaftliche Durchbrüche und führt zu vermeidbaren Dopplungen bei Forschungsprojekten, die Millionen an Fördermitteln verschwenden.**

- **Beim Training von Künstlicher Intelligenz verstärken sich die Probleme mangelnder Dateninteroperabilität erheblich, da KI-Systeme besonders sensibel auf Datenqualität und -konsistenz reagieren. Wenn Datensätze aus verschiedenen Quellen unterschiedliche Kodierungen, Maßeinheiten oder Kategorisierungen verwenden, lernt das Modell diese Inkonsistenzen als vermeintliche Muster. Dies resultiert in systematischen Fehlern, die sich durch das gesamte System propagieren und schwer zu identifizieren sind.**

- **Ein KI-Unternehmen für Wetterprognosen sammelte beispielsweise Daten von verschiedenen nationalen Wetterdiensten. Deutsche Daten verwendeten andere Höhenangaben (Meter über NN) als britische (Fuß über Meeresspiegel), und Windgeschwindigkeiten wurden in verschiedenen Einheiten gemessen. Die KI lernte diese Inkonsistenzen als regionale „Wettermuster“ und produzierte systematisch falsche Vorhersagen für Grenzregionen.**

- **Wenn sich Datenquellen oder -formate ändern, müssen KI-Modelle häufig komplett neu trainiert werden, anstatt inkrementell aktualisiert zu werden.**

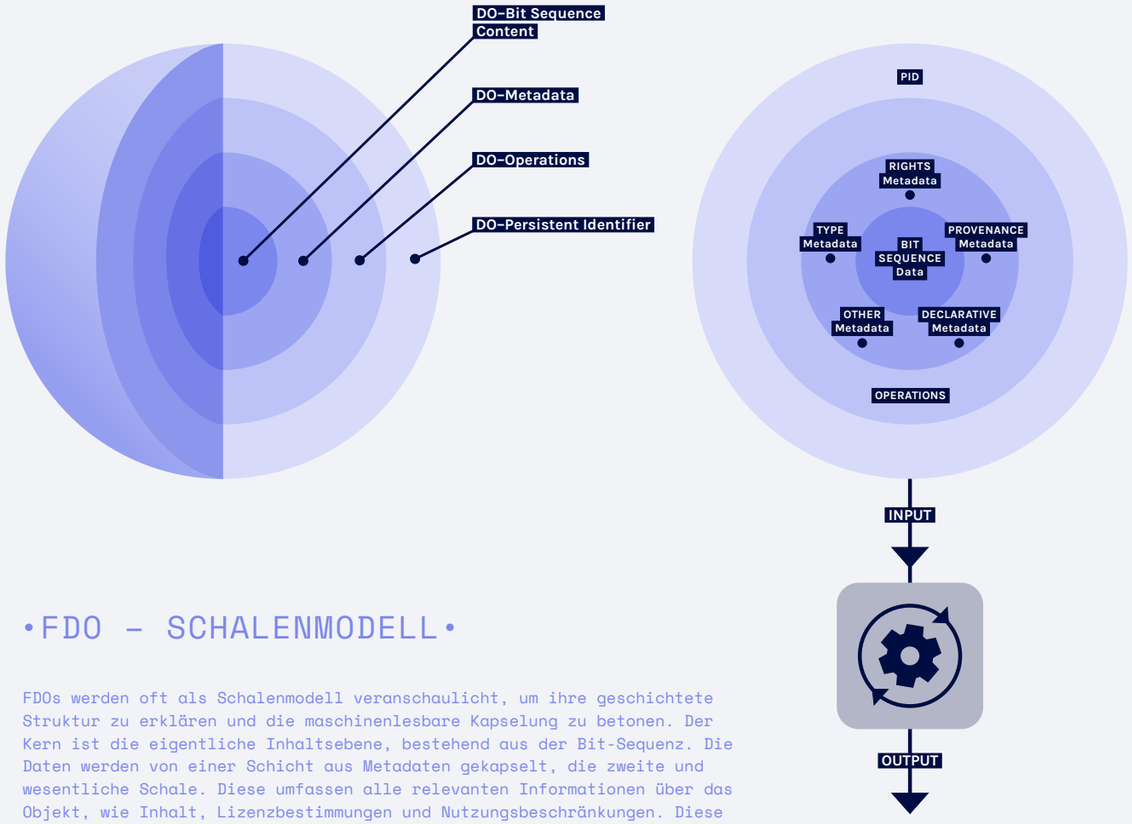
Das Konzept der FAIR Digital Objects

FAIR Digital Objects (FDOs) stellen einen bedeutenden Fortschritt in der digitalen Datenverwaltung dar. Sie verstehen sich als ein technologieunabhängiges Regelwerk, das digitale Objekte standardisiert austauschbar macht.

Die zentrale Idee wird durch einen Vergleich mit Containern im Schifffahrtsbereich verdeutlicht: Wie bei Containern geht es nicht um das Material selbst, sondern um die exakte Spezifikation der Dimensionen, Schnittstellen und Eigenschaften. Ein Kran in Hamburg kann genauso mit einem Container arbeiten wie einer in Peking, weil beide dasselbe Protokoll verstehen.

FDOs funktionieren genauso nach generalisierten Grundsätzen – sie definieren klare Regeln, wie digitale Objekte strukturiert sein müssen, welche Eigenschaften sie haben und wie sie miteinander interagieren. Jedes FDO hat eindeutige Identifier, definierte Operationen und Profile, die beschreiben, welche Eingabedaten benötigt werden und welche Ausgabedaten zu erwarten sind. Dadurch können verschiedene Systeme und Services nahtlos zusammenarbeiten, ohne die zugrundeliegende Technologie verstehen zu müssen. Die besondere Stärke liegt in der Protokoll- und Technologievielfalt: FDO-Regeln können mit verschiedenen existierenden Protokollen und Standards umgesetzt werden – sei es z. B. mit Nanopublikations-Protokollen, mit DOIP-basierten Binärprotokollen (Digital Object Interface Protocol) oder durch Linked-Data-Ansätze. Für die Entwicklung spezialisierter APIs bedeutet das, dass beispielsweise Services für Batterien, Bauwesen oder Medizinprodukte jeweils eigene APIs bereitstellen können, die alle auf denselben FDO-Grundregeln basieren.

Das ermöglicht nahtlose Interoperabilität: Ein System muss nicht die zugrundeliegende Technologie verstehen – es reicht, wenn es die FDO-Regeln beherrscht. Dadurch können verschiedene Implementierungen verlustfrei ineinander übersetzt werden. FDOs schaffen somit eine Abstraktionsebene, die es ermöglicht, komplexe digitale Services zu nutzen, ohne deren technische Details verstehen zu müssen. Die Funktionsweise basiert auf einer Kombination aus Standardisierung, Maschinenlesbarkeit und automatisierter Verknüpfbarkeit. Im Unterschied zu herkömmlichen Datenstrukturen integrieren FDOs ihre beschreibenden Informationen direkt in die Objektarchitektur, wodurch sie sich selbst erklären. Diese Kombination von Daten und Metadaten ermöglicht eine autonome Interpretierbarkeit durch maschinelle Systeme und reduziert externe Abhängigkeiten. Die Architektur eines FDO basiert auf einem hierarchisch strukturierten Komponentenmodell.



•FDO – SCHALENMODELL•

FDOs werden oft als Schalenmodell veranschaulicht, um ihre geschichtete Struktur zu erklären und die maschinenlesbare Kapselung zu betonen. Der Kern ist die eigentliche Inhaltsebene, bestehend aus der Bit-Sequenz. Die Daten werden von einer Schicht aus Metadaten gekapselt, die zweite und wesentliche Schale. Diese umfassen alle relevanten Informationen über das Objekt, wie Inhalt, Lizenzbestimmungen und Nutzungsbeschränkungen. Diese Kapselung ermöglicht je nach Typ auch die Definition von Operationen auf das Objekt. Die äußerste Schale bildet der Persistent Identifier (PID) als stabiler Verbindungspunkt zum gesamten Objekt. Für ein echtes FDO ist dabei entscheidend, dass es Metadaten bereitstellt, die sofort beim Auflösen des PIDs und vor allem für maschinengesteuerte Prozesse zugänglich sind. Je nach Betonung der Anwendung variieren Anzahl und Bezeichnung der Schalen.

Alle FDOs folgen einem standardisierten Aufbau und sind typisiert, wodurch ihr Inhalt und ihre Funktion von Natur aus maschinell erkennbar und interoperabel sind. Über vordefinierte Schnittstellen und Protokolle können sie verschiedene Informationen wie Metadaten, Rechte oder Nutzungsbefugnisse bündeln, die über Repositorien ausgetauscht und miteinander vernetzt sind.

Historisch fand der aus der Informatik stammende Begriff des „Digital Objects“ zunächst in verschiedenen objektorientierten Programmiersprachen Verwendung, bevor 1995 Robert Kahn – gemeinsam mit Vint Cerf, Entwickler der Internet-Grundlage, das TCP/IP (Transmission Control Protocol/Internet Protocol) – und Robert Wilensky den Begriff erstmals auf Datendienste im Internet übertragen haben. In ihrem Paper „A framework for distributed digital object services“ schlugen sie ein technisches Rahmenwerk vor, das es erlaubt, digitale Informationen, unabhängig von Ort, Format oder Softwareumgebung, als standardisierte digitale Objekte zu speichern, zu identifizieren, zu übertragen und zu verwalten.

Beispiele für digitale Objekte sind vielfältig – sie reichen von Textdateien über Websites mit Datenbankinhalten bis hin zu komplexen Excel-Listen mit vielen verschiedenen Parametern und Referenzen. Es handelt sich dabei im Kern um Bitsequenzen bzw. Datenstrukturen, deren zentrale Bestandteile aus Daten oder digitalem Material (Key-Metadata) bestehen. In einfacher Form könnte ein einfach strukturiertes Digital Object etwa eine XML-Datei mit Postleitzahlen sein, bei der jede Ziffernfolge für einen konkreten Ort in

Deutschland steht. Deutlich komplexere Digital Objects hingegen zeichnen sich durch hierarchische Strukturen und ihre gegenseitige Vernetzung aus, wie es beispielsweise bei digitalen Konstruktions- oder Bebauungsplänen im parametrischen Design der Fall ist.

Weiterhin verfügen sie über eine spezifische und dauerhafte Kennzeichnung (Digital Object Identifier), die aus einer eindeutigen Zeichenkette hervorgeht (Name der Kennzeichnungsstelle sowie zugehörige Zeichenfolge) und von einem Handle-Generator zugewiesen wird.

Technisch betrachtet folgt die Architektur der Digital Objects einem klaren Aufbau, der die langfristige Auffindbarkeit, Zugänglichkeit und Nutzbarkeit über unbestimmte Zeit sicherstellt.

Den wichtigsten Baustein bildet die Identifizierungsschicht (Identifier/Resolution System). Sie vergibt jedem Digital Object einen Persistent Identifier (PID), das Fundament jedes FDO. Er macht das FDO jederzeit auffindbar – auch wenn sich der Inhalt oder der Dateiname ändert. Solche Identifikatoren sind eine essenzielle Voraussetzung für beispielsweise die wissenschaftliche Zitierbarkeit und die langfristige Datenverfügbarkeit. Für das Beispiel des digitalen Bebauungsplans würde das bedeuten, dass dieser einen eindeutigen und dauerhaften Identifikator (z. B. eine DOI) erhält. Dieser PID dient nunmehr als offizielle Adresse des Bebauungsplans und jeder, der sie kennt, kann direkt zum Digital Object mit dem konkreten Plan gelangen.

Hinter dem Identifikator verbirgt sich ein strukturierter Metadatensatz, der die eigentlichen Objektinformationen enthält und als Sammlung von Attributen organisiert ist, wobei jedes Attribut aus einer Schlüssel-Wert-Kombination besteht und die Werte unterschiedliche Datentypen annehmen können. Dies vollzieht sich in der Modellschicht (Repository System) als zentrale Verwaltungseinheit. Hier werden die eigentlichen Digital Objects (DOs) gespeichert und ihre strukturellen und semantischen Eigenschaften definiert. Beim Beispiel des Bebauungsplans befinden sich hier also die eigentlichen Daten und alle zugehörigen Dokumente.

Die dritte Ebene, die Ressourcenschicht (Registry System), übernimmt die Rolle eines Metadatenkatalogs. Im Gegensatz zur Modellschicht speichert sie nicht die digitalen Objekte und ihre Inhalte selbst, sondern ausschließlich deren Metadaten (Titel, Beschreibung, Herausgeber etc.). Diese Informationen ermöglichen eine effiziente Suche und Filterung von Digital Objects.

Ein herausragendes Merkmal der FDO-Architektur ist die selbstreferenzielle Definition der Attribute. Jeder Attributschlüssel verweist auf eine formale Beschreibung, die die Eigenschaften des zugehörigen Werts festlegt. Diese Definitionen existieren selbst als eigenständige FDOs in öffentlich zugänglichen Verzeichnissen und schaffen so eine vollständige Selbstdokumentation des Systems. Spezialisierte Attributtypen ermöglichen zudem die Verlinkung zwischen verschiedenen FDOs, indem sie persistente Identifikatoren als Werte nutzen und ein vernetztes Geflecht digitaler Objekte schaffen. Profilverweise spielen eine besondere Rolle, da sie auf standardisierte Strukturvorgaben zeigen (z. B. durch Sign Posting) und die Konformität eines FDOs zu definierten Schemata sicherstellen.

Die Profildefinitionen legen fest, welche Attribute in einem FDO obligatorisch sind und welche optional hinzugefügt werden können. Diese Strukturvorgaben ermöglichen die Standardisierung von FDOs für spezifische Anwendungsdomänen, bieten aber gleichzeitig Flexibilität für individuelle Erweiterungen. Die Validierung gegen diese Profile stellt die Konsistenz und Vollständigkeit der Datenobjekte sicher. Die besondere Stärke dieser Architektur liegt in ihrer rekursiven Selbstbeschreibungsfähigkeit: Maschinelle Systeme können durch die Auflösung der verknüpften Definitionen die vollständige Semantik eines FDOs autonom erschließen. Diese Eigenschaft eliminiert die Abhängigkeit von externen Dokumentationen oder proprietären Spezifikationen, da die gesamte Interpretationslogik im verteilten FDO-Ökosystem kodiert und dauerhaft zugänglich bleibt.

Indem FDOs den FAIR-Grundsätzen folgen, welche sich als Akronym aus den Wörtern Findable, Accessible, Interoperable und Reusable ableiten, gehen sie inhaltlich deutlich weiter als DOs, da sie die spezifischen Herausforderungen moderner Datenökosysteme unter diesen Grundsätzen adressieren. FAIR wurde ursprünglich für das Management wissenschaftlicher Daten entwickelt und ist technologieneutral.

FAIR-PRINZIPIEN

Zurückgehend auf einen Artikel von Marc Wilkinsons und seinem Team aus dem Jahr 2016 („The FAIR Guiding Priciples for scientific data management and stewardship“) wurde das Prinzip als technische Lösung für ein zukunftsfähiges Datenökosystem in der Wissenschaft entwickelt.

Wilkinsons Forschung zeigte, dass in naturwissenschaftlichen Disziplinen wie den Biowissenschaften oder der Weltraumforschung bereits gut organisierte Repositorien für Forschungsdaten vorhanden sind, die den wissenschaftlichen Output erheblich verbessern.

Eines der zentralen Probleme dieser Repositorien bestand jedoch darin, dass die enthaltenen Daten oft hoch spezialisiert waren und nicht alle relevanten Informationen abbildeten.

Mit der zunehmenden Digitalisierung fehlten jedoch standardisierte Strukturen, die einen maschinellen Datenzugriff ermöglicht hätten. Gleichzeitig wuchs die Notwendigkeit, diesen maschinellen Zugriff effizient und unkompliziert zu gestalten. Während Menschen Daten häufig durch begleitende Kontextinformationen erschließen können, benötigen Maschinen hierfür klar definierte, formale Standards.

Die FAIR-Grundsätze setzen genau an diese Problematik an, indem sie den Austausch von Daten durch standardisierte, interoperable und maschinenlesbare Formate fördern. Ziel ist es hierbei, die Wiederverwendbarkeit, Transparenz und automatisierte Verarbeitung digitaler Objekte signifikant zu verbessern.

Während die FAIR-Grundsätze in der europäischen Forschungsförderung inzwischen als verbindlicher Standard gelten, zeigen sich Wirtschaft und Verwaltung mit der Implementierung noch zögerlich. In einer zunehmend datengetriebenen Wissensgesellschaft reicht es aber nicht mehr aus, dass Daten lediglich lokal gespeichert werden. Vielmehr müssen sie so aufbereitet, beschrieben und veröffentlicht werden, dass sie langfristig auffindbar, abrufbar, in neue Kontexte einbettbar und jederzeit erneut verwendet werden können. Die FAIR-Grundsätze schaffen damit einen effizienten Rahmen, ohne sensible Daten unkontrolliert zugänglich zu machen. Denn FAIR bedeutet nicht automatisch Open Data. Das Prinzip selbst verhält sich neutral zu der Frage, wer unter welchen Bedingungen Zugriff zu den digitalen Objekten erhält und für welche Zwecke er sie nutzen darf, wodurch sie sich auch für proprietäre Anwendungskontexte eignen.

→ FAIR-Prinzipen bei Digital Objects

FINDABLE
PIDs gewährleisten durch eine dauerhafte, einmalige Kennzeichnung die Auffindbarkeit von Daten und Metadaten in digitalen Räumen. Zusätzlich werden die Daten mit umfangreichen Metadaten anhand standardisierter Schemata beschrieben. Dabei enthalten die Metadaten stets die eindeutige Kennung der Daten, die sie beschreiben. Abschließend werden sowohl die Daten als auch ihre Metadaten in durchsuchbaren Repositorien registriert. Das ermöglicht es autonomen Systemen, die relevanten Daten unkompliziert wiederzufinden.

ACCESSIBLE
Die Zugänglichkeit der Daten stellt sicher, dass diese über standardisierte Kommunikationsprotokolle abrufbar sind. Diese Protokolle sind offen, frei verfügbar und universell einsetzbar. Dadurch lässt sich ein Authentifizierungs- und Autorisierungsverfahren durchführen. Auch beim Abhandenkommen der Daten im digitalen Raum bleiben die zugehörigen Metadaten weiterhin zugänglich.

INTEROPERABLE
Interoperabilität setzt voraus, dass die (Meta-)Daten in einer formalen und maschinenlesbaren Sprache beschrieben sind. Zudem enthalten sie qualifizierte Verweise auf andere Daten, um die Einbettung in komplexe Analysekontexte zu ermöglichen.

REUSABLE
Die Wiederverwendbarkeit von Daten zielt darauf ab, dass die Daten umfassend beschrieben (z. B. durch Herkunftsangaben) und mit entsprechenden Nutzungslizenzen versehen sind.

EXPOSITION_FAIR DIGITAL OBJECTS

Treibende Akteure zu FAIR Digital Objects

- **EOSC Association:** Ein zentraler Akteur in Europa. Sie zielt darauf ab, eine föderierte, disziplinübergreifende Umgebung für Forschungsdaten und -services in Europa zu schaffen. FDOs spielen eine grundlegende Rolle in der Implementierung von EOSC, da sie als das technische Rückgrat für den interoperablen Datenaustausch und die Schaffung einer „Global Digital Object Cloud“ angesehen werden. Die EOSC Association hat sogar eine „FAIR Metrics and Digital Objects Task Force“ eingerichtet.

- **GO FAIR Initiative**: Eine treibende Kraft bei der Entwicklung und Umsetzung der FAIR-Prinzipien. Sie koordiniert verschiedene Implementation Networks (INs). Das FAIR Digital Objects (FDO) Forum IN innerhalb von GO FAIR arbeitet aktiv an der Spezifizierung des FDO-Frameworks, der Entwicklung von Validatoren für die FAIRness und der Förderung von Pilotprojekten zur Realisierung einer nutzbaren Infrastruktur.

- **Research Data Alliance (RDA):** Eine internationale Organisation, die sich durch ihre Arbeits- und Interessengruppen der Entwicklung von globalen Dateninfrastrukturen und Best Practices widmet. Innerhalb der RDA wurden wichtige Grundlagen für FDOs geschaffen, beispielsweise durch die RDA Kernel Information Group, die einen ersten Kernsatz relevanter Attribute definierte, und weitere Gruppen, die die Natur von Typ-Registern spezifizierten.

- **DONA Foundation:** Verantwortlich für das Handle-System, einem persistenten Identifikator-System. Da FDOs die Nutzung persistenter Identifikatoren (PIDs) erfordern, um ihre Auffindbarkeit und Zitierfähigkeit zu gewährleisten, ist die DONA Foundation indirekt eine Schlüsselorganisation, die die notwendige technische Infrastruktur für die Adressierung und Lokalisierung von FDOs bereitstellt.

- **FAIR Digital Objects (FDO) Forum:** Eine zentrale, internationale Initiative, deren Hauptziel die Standardisierung und weltweite Implementierung des FDO-Konzepts ist. Es fungiert als Drehscheibe für Experten aus Wissenschaft, Industrie und Politik, um eine kohärente und globale Datenlandschaft zu schaffen. Es verbindet einige der hier genannten Akteure.

- **FDO4DE (FAIRe Digitale Objekte für Deutschland):** Eine User-basierte Anwendungsinitiative und ein Konsortium, das sich der nationalen und anwendungsorientierten Implementierung der FAIR Digital Objects (FDOs) in Deutschland verschrieben hat. Während das internationale FDO Forum die Spezifikationen festlegt, konzentriert sich FDO4DE auf die praktische Überführung des Konzepts in die reale Anwendung über verschiedene Sektoren hinweg.

- **acatech:** Maßgeblich in Deutschland an der MISSION KI – Nationale Initiative für Künstliche Intelligenz und Datenökonomie beteiligt. Im Rahmen dieser Initiative werden mehrere Projekte gefördert, die sich direkt mit FDOs beschäftigen. acatech fungiert als Umsetzungsorganisation und treibt so die Entwicklung einer FDO-Servicearchitektur und die Erarbeitung von Handlungsempfehlungen für den europaweiten und sektorübergreifenden Einsatz voran (z. B. im Projekt FAIR Data Publisher).

- **GWDG (Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen):** Führt als Hauptauftragnehmer das Projekt FAIR Data Publisher und entwickelt zusammen mit einem Konsortium von 17 Partnern eine modulare FDO-Servicearchitektur.

- **Fraunhofer-Institute** (z. B. ISST, FIT): Aktiv in Projekten wie FDO One , um prototypische technische Infrastrukturen zu implementieren und die Zusammenführung von FDOs mit bestehenden Konzepten wie den International Data Spaces (IDS) und der Asset Administration Shell (AAS) zu untersuchen.

- **Helmholtz Metadata Collaboration (HMC):** Arbeitet an der Realisierung von FAIR Digital Objects innerhalb der Helmholtz-Gemeinschaft.

- **RWTH Aachen University:** Ausgezeichnet für ihre herausragende Umsetzung der FAIR-Prinzipien im Forschungsdatenmanagement mit dem EUNIS Award (Projekt „FAIR Digital Objects: FAIRtilizer for the Digital Harvest“), was ihre Rolle als wichtige Implementierungs-Organisation unterstreicht.

Wie unterscheiden sich FAIR Digital Objects zu anderen Standards und ergänzen diese?

Die vier Konzepte – FAIR Digital Objects, Data Catalog Vocabulary (DCAT), Gaia-X Federation Services sowie IDSA RAM/EDComponents – verfolgen unterschiedliche Schwerpunkte beim Umgang mit Daten, weisen jedoch auch Überschneidungen auf.

FDOs sind stark auf die Umsetzung der FAIR-Prinzipien ausgerichtet. Sie kapseln Daten in digitalen Objekten, die mit persistenten Identifikatoren versehen sind, und sorgen so für Rückverfolgbarkeit und Wiederverwendbarkeit. Das zugrunde liegende Datenmodell trennt streng zwischen Daten und Metadaten, was die Interoperabilität erheblich verbessert. Auch im Bereich der Zugriffskontrolle bieten FDOs präzise Richtlinienmechanismen, die eine sehr feingranulare Steuerung erlauben. Besonders hervorzuheben ist ihre Skalierbarkeit, da sie gut für große und verteilte Datensätze geeignet sind.

Der DCAT-Standard verfolgt einen anderen Ansatz. Es handelt sich hierbei um ein ontologiebasiertes Vokabular, das speziell für die Beschreibung von Datensätzen und -diensten in Katalogen entwickelt wurde. DCAT ermöglicht damit ein semantisches Verständnis von Daten, was vor allem in Metadaten-Katalogen große Vorteile bringt. In Bezug auf Interoperabilität liegt DCAT im Mittelfeld: Innerhalb von Katalogumgebungen sorgt es für Kompatibilität, ist aber weniger für den breiten Datenaustausch geeignet. Zugriffssteuerung ist nur rudimentär vorgesehen, da das Resource Description Framework (RDF) primär auf Datenbeschreibung fokussiert ist. Auch die

	FAIR Digital Objects	Data Catalog Vocabulary (RDF*)	Gaia-X Federated Services	IDSA RAM / EDComponents
Anwendungs-bereich	**Kapselung und Rückverfolgbarkeit** von Daten durch digitale Objekte und persistente Identifikatoren	**Vokabular** zur Beschreibung von Datensätzen und Datendiensten in **Katalogen**	Ökosystem-Ansatz für den Aufbau **föderierter Dateninfrakstrukturen**	Referenzarchitektur und Komponenten für den Aufbau von **B2B-Datenräumen**
Datenmodell	**Klare Trennung** von Daten und Metadaten, konzipiert zur Einhaltung der FAIR-Prinzipien	**Ontologiebasiert** und zugeschnitten auf die Beschreibung von Datensätzen in Katalogen	**Verteiltes Modell,** entwickelt für föderierte Infrastrukturen	**Datenspezifisches Modell** mit Fokus auf sicheren Datenaustausch
Interoperabilität	**Hohe Interoperabilität** durch strikte Einhaltung der FAIR-Prinzipien	**Mittel** – Sicherstellung der Kompatibilität in Katalogumgebungen	**Hoch** – basiert auf standardisierten Schnittstellen und föderierten Ansätzen	**Mittel** – hohe Interoperabilität innerhalb IDSA-konformer Ökosysteme
Zugriffskontrolle	**Sehr stark** – präzise Zugriffskontrolle durch Richtliniendefinitionen	**Schwach** – unterstützt Zugriffskontrolle nicht von Haus aus	**Stark** – ermöglicht Datensouveränität durch föderiertes Identitätsmanagement	**Stark** – legt starken Fokus auf sicheren Datenaustausch
Skalierbarkeit	Gut geeignet für die Verwaltung **verteilter und großer Datensätze**	**Nicht** für die Verarbeitung großer Datensätze ausgelegt	Unterstützt **groß angelegte, verteilte Architekturen**	Skalierbarkeit ist an die **spezifische IDS-Architektur gebunden**

*DCAT ist ein RDF-Vokabular, das die Interoperabilität zwischen im Web veröffentlichten Datenkatalogen erleichtern soll.

Skalierbarkeit ist eingeschränkt, da das Vokabular nicht für die Verarbeitung großer Datenmengen optimiert wurde.

Die Gaia-X Federation Services verfolgen einen Ökosystem-Ansatz und adressieren die Herausforderung, föderierte Dateninfrastrukturen zu schaffen. Das Datenmodell ist auf ein verteiltes System ausgelegt, in dem Daten bei den Anbietern verbleiben, aber über standardisierte Schnittstellen föderiert genutzt werden können. Dadurch erreicht Gaia-X eine hohe Interoperabilität. Auch die Zugriffskontrolle ist stark, da föderiertes Identitäts- und Berechtigungsmanagement die Datensouveränität sicherstellt. Zudem sind Gaia-X-Dienste klar auf Skalierbarkeit ausgerichtet und unterstützen groß angelegte, verteilte Architekturen.

Das IDSA RAM/EDComponents-Modell wiederum ist primär auf den sicheren B2B-Datenaustausch zugeschnitten. Es stellt eine Referenzarchitektur sowie konkrete Komponenten für Datenräume bereit. Das Datenmodell ist datenspezifisch und betont den sicheren Datenaustausch. Interoperabilität wird auf mittlerem Niveau gewährleistet, da sie innerhalb IDSA-konformer Ökosysteme hoch ist, außerhalb dieser Architektur jedoch eingeschränkt sein kann. Der Zugriffsschutz ist ein zentrales Element und stark ausgeprägt, da Sicherheit und Vertrauen im Vordergrund stehen. Die Skalierbarkeit ist zwar grundsätzlich gegeben, allerdings an die spezifische IDS-Architektur gebunden.

Zusammenfassend lässt sich sagen, dass FAIR Digital Objects besonders bei Interoperabilität, Zugriffskontrolle und Skalierbarkeit überzeugen, während DCAT seine Stärken in der semantischen Datenbeschreibung hat, aber Schwächen bei Sicherheit und Skalierung zeigt. Gaia-X bietet eine flexible und skalierbare Lösung für föderierte Dateninfrastrukturen mit starker Datensouveränität, wohingegen IDSA RAM/EDComponents eine praxisnahe Architektur für vertrauensvolle B2B-Datenräume bereitstellt, jedoch stärker an ein spezifisches Ökosystem gebunden ist.

FAIR Digital Objects in Bezug auf Internet, Datenräume und Intelligenz

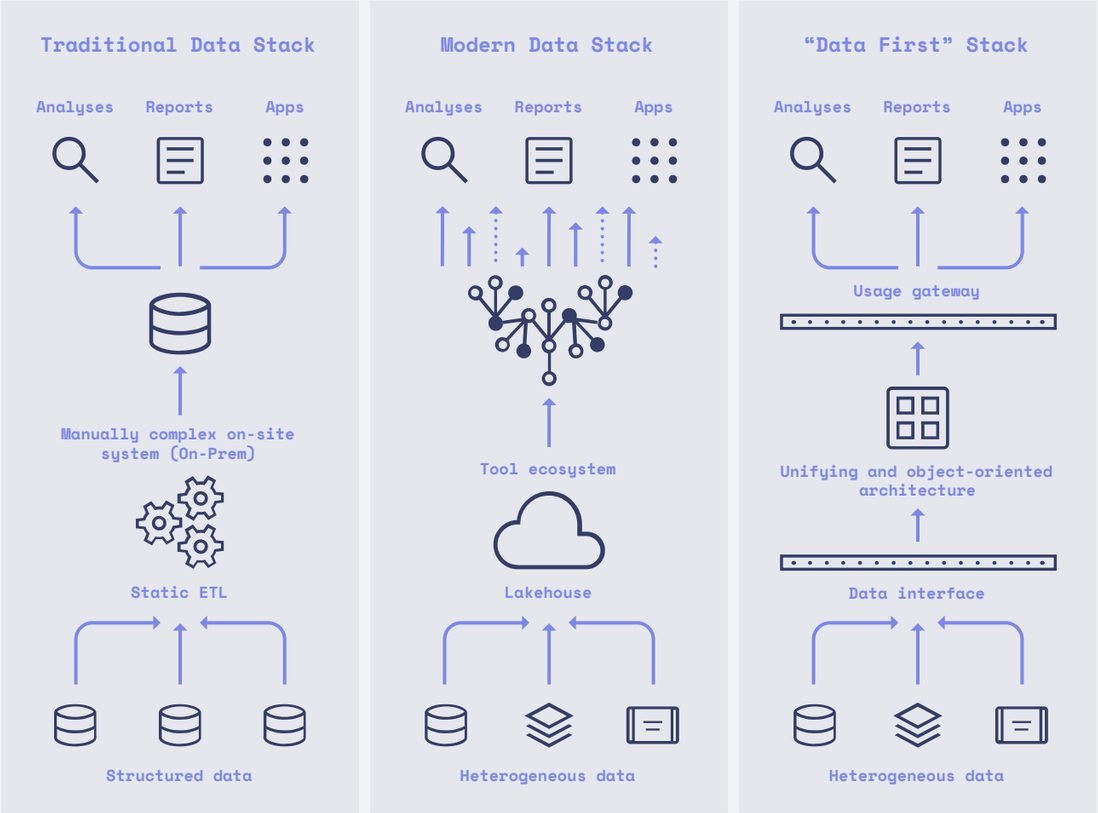
Die Entwicklung der Dateninfrastrukturen hat in den letzten zwei Jahrzehnten eine bemerkenswerte Transformation durchlaufen. Getrieben von exponentiell wachsenden Datenmengen, veränderten Geschäftsanforderungen und technologischen Fortschritten entstanden drei distinkte Evolutionsstufen, die jeweils neue Mehrwerte und Möglichkeiten erschlossen haben. Der Trend geht zum „Data First Stack“, der heterogene Daten als primären Geschäftswert ins Zentrum aller Architekturentscheidungen stellt. FDOs können durch ihre selbstreferenziellen Eigenschaften einen wichtigen Beitrag zu dessen Realisierung leisten.

DER TRADITIONELLE DATA STACK: STATISCHE FUNDAMENTE

Der traditionelle Legacy Data Stack (LDS) war geprägt von strukturierten Daten und wartungsintensiven On-Premises-Systemen. Diese Architektur basierte auf dem bewährten ETL-Paradigma (Extract-Transform-Load), bei dem Daten zunächst aus verschiedenen Quellsystemen extrahiert, dann transformiert und schließlich in relationale Data Warehouses geladen wurden. Die strikte Vorab-Modellierung der Datenstrukturen erforderte eine hohe Abhängigkeit von teurer Hardware und lokaler Infrastruktur, was zu langwierigen Entwicklungszyklen für neue Datenintegration führte und die Skalierbarkeit sowie Flexibilität erheblich begrenzte.

DER MODERNE DATA STACK: FLEXIBLE SKALIERBARKEIT

Der moderne Data Stack (MDS) ist das Resultat eines Paradigmenwechsels von ETL zu ELT (Extract-Load-Transform), was zu deutlich mehr Konnektivität und flexiblerer Nutzung verschiedener Datenservices führte. Diese moderne Datenarchitektur vereint die Skalierbarkeit von Data Lakes mit den strukturierten Abfrage- und Transaktionsfähigkeiten von Data Warehouses. Die Integration heterogener Datenquellen ermöglichte es, strukturierte, semi-strukturierte und unstrukturierte Daten gemeinsam zu verarbeiten, während Cloud-native Architekturen eine elastische Skalierung gewährleisteten.



•ENTWICKLUNG DATA STACK•

Der transformative Übergang von traditionellen, systemzentrierten Ansätzen zu modernen, datenzentrischen Architekturen, die interoperable, metadatenreiche, objektorientierte und automatisierte Datenflüsse ermöglichen.

DER „DATA FIRST“ STACK: DATEN ALS STRATEGISCHES GUT

Die neueste Evolutionsstufe positioniert heterogene Daten als primären Geschäftswert im Zentrum aller Architekturentscheidungen. Im Gegensatz zu gängigen Ansätzen, die oft als „Data Last“ beschrieben werden und sich stark auf das Management und die Wartung der Dateninfrastrukturen konzentrieren, de-priorisiert der Data First Stack diese Aufgaben entweder durch Abstraktionen oder durch intelligente Designarchitekturen.

Der DFS betrachtet Daten als Software und schließt Code mit ein, um objektorientierte Fähigkeiten wie Abstraktion, Kapselung, Modularität, Vererbung und Polymorphie über alle Komponenten des Data Stacks hinweg zu ermöglichen. Dies ist entscheidend, da das Daten-Ökosystem nicht nur unter Datensilos, sondern auch massiv unter „Data Code Silos“ leidet, bei denen Code-Komponenten nicht miteinander kommunizieren können. Die Architektur basiert auf Vereinheitlichung durch Modularisierung, wobei zusammensetzbare primitive Komponenten miteinander sprechen, um komplexere Datenlösungen zu bilden. Der DFS ist durch eine vereinheitlichte Architektur-Ideologie (unified architecture ideology) in Verbindung mit vertraglichen Handshakes (contractual handshakes) realisierbar. Diese Kombination ermöglicht die Umkehrung der Datenverantwortlichkeit näher an die Datenquelle, was eine Voraussetzung für ein echtes Datenprodukt ist.

WEITERENTWICKLUNG DES INTERNETS

Der Schwerpunkt des Internets, wie wir es bislang kennen, liegt auf Inhalten wie Dokumenten und Medien. Bezüglich strukturierter Daten und insbesondere Metadaten (Kontexte, Anweisungen) weist es Defizite auf. Ein Data First-Denken wie bei FDOs kann zur Weiterentwicklung von Internet und WWW von einem Netz der Datenübertragung, über ein Netz der Dokumente zu einem Netz der Daten, Register und Bedeutungen beitragen, um einen „Global Integrated Data Space“ (GIDS) aus atomaren und komplexen, vertrauenswürdigen, persistenten und maschinenlesbaren Informationseinheiten zu erschaffen. Die Entwicklung des GIDS soll mit einer geschichteten Architektur erfolgen, um die Komplexität zu bewältigen, vergleichbar mit dem ISO/OSI-Modell (International Organization for Standardization/Open Systems Interconnection) des Internets. Im Gegensatz zum Internet jedoch, bei dem nur Pakete ausgetauscht werden, geht es beim GIDS um bedeutungsvolle Objekte (Digital Objects), die eine generische Integrationsbasis bilden und spezifische Anforderungen wie z. B. Verarbeitungskontexte, Provenienz, Datensouveränität und datenrechtliche Aspekte adressieren. Sie werden in der Architektur als Basiselemente für eine hierarchische Informationsinfrastruktur in einem zukünftigen GIDS berücksichtigt.

Dies verwandelt das Internet vom Kommunikationsnetz über einen noch relativ statischen Wissensspeicher zu einem komplexen Lern-Graphen, der den Weg für echtes kausales Lernen und laterales Denken bei Maschinen ebnet. Denn wenn Kontext-Fakten in einem Fakten-Graph in der Form zueinander perkolieren, sodass daraus eine Idee und ein Modell emergieren, das es auf einer allgemeineren, abstrakteren Ebene entweder modellierbar macht oder auf ein komplett anderes Problem übertragbar macht, wird echtes Lernen erreicht. Voraussetzung ist dabei, dass im Quell- und Zielsystem die Kontexte jeweils derart perkolieren, dass ein logisches Matching möglich wird. Ein Kontextfenster eines LLM ist eben nicht ein echtes Kontextfenster und noch lange nicht „Ground Truth“. Es ist zwar die Buchstabensuppe eines Kontextbereichs, aber nicht der wirklich kausale Kontext und noch lange nicht der breitere Kontext, um als Grundlage für Intelligenz zu dienen.

Oder wie es Richard S. Sutton, Preisträger des 2024 Turing Award, ausdrückt: „Learning from the Stream“ ist etwas anderes als „Learning about the Stream“.

FAIR DATA PUBLISHER

FDO-Service-Architektur

Aufbauend auf der vorangegangenen Entwicklung einer Handle-basierten Architektur für FAIR Digital Objects im Rahmen von FAIR Data Publisher wurde diese nun mit der Technologie der Nanopublikationen verknüpft und mit der Welt von Linked Data und semantischen Modellen vereint.

¬ Die Kombination von Handle-basierten FDOs mit Nanopublikationen und Linked Data bietet signifikante Vorteile: sie überbrückt unterschiedliche Communitys (Handle vs. Web/Linked Data/Semantic Web) und nutzt Technologien der Linked Data Welt wie Vokabularien, Ontologien und Logik zur Weiterentwicklung des FDO-Konzepts. Nicht-digitale Konzepte sind repräsentierbar.

Die Architektur nutzt die Unveränderlichkeit (Immutability) von Nanopublikationen als Stabilitätsgarantie und ermöglicht gleichzeitig über Such-Endpunkte schnelle Abfragen und eine umfängliche Suchfunktionalität. Die einzelnen Komponenten – von der GUI-basierten Verwaltung über automatisierte Metadatenextraktion bis hin zu dezentralen Abfragediensten und autonomen FDOs – greifen somit ineinander und eröffnen dadurch eine umfassende Nutzung in Forschungs- und Datenökosystemen. Offene Schnittstellen, quelloffene Implementierungen und die Integration in bestehende Plattformen wie z. B. Coscine oder datenraumbasierte Systeme sichern zudem eine nachhaltige Anschlussfähigkeit und bilden die Grundlage für eine breite Anwendung und Weiterentwicklung der Architektur.

Dadurch bildet sie ein leistungsfähiges Fundament für die Weiterentwicklung digitaler Objekte, indem sie die Interoperabilität und maschinelle Nutzbarkeit von Datenobjekten deutlich erhöht und dabei stabile, skalierbare und nachvollziehbare Mechanismen für Identifikation, Versionierung und Abfrage etabliert.

KONZEPT UND IMPLEMENTIERUNG VON NANOPUBLIKATIONEN ALS FDOS

In diesem Modell entspricht ein Metadaten-Record einer Nanopublikation, wobei ein FDO Metadaten-Record in der Nanopublikations-Welt eine ganze Versionshistorie oder Sequenz von Nanopublikationen repräsentiert. Jede Nanopublikation in dieser Kette verweist auf die vorherige Version, und der neueste Eintrag bildet den aktuellen Record. Frühere Versionen sind über spezielle Abfragen zugänglich.

Die Auflösung eines FDO-Identifikators, der als Handle-PID vorliegt, führt stets zur neuesten Version der Nanopublikation. Die Identifikatoren älterer Versionen sind im Record enthalten und können ebenfalls direkt abgefragt werden. Komplexere Abfragen über die Versionshistorie sind über SPARQL möglich. Es wurde ein Mapping zwischen Handle-IDs und URIs konzipiert: Ein Präfix wie https://handle.net/ wird der Handle-ID vorangestellt, um einen URI zu erzeugen. Umgekehrt können URIs eines Nanopublikations-basierten FDOs über Erweiterungen der Tools auf der Handle-Seite unter einem Handle verfügbar gemacht werden. Die Schnittstellen sind dabei in Java und Python implementiert.

FDO MANAGER: NUTZUNGSOBERFLÄCHE UND SDK

Der FDO Manager stellt eine grafische Benutzerinnen- und Benutzeroberfläche bereit, die sowohl die manuelle Erstellung von FDOs über profildefinierte Formulare als auch die automatisierte Metadaten-Extraktion mittels großer Sprachmodelle ermöglicht. Er basiert auf einer früheren Version aus dem Vorprojekt FDO One und wurde für die Steuerung der Funktionen des Systems erweitert. Die Landingpage des Managers ermöglicht das Durchsuchen bestehender FDOs und die Anzeige eines Feeds von Nanopublikationen mit grundlegenden Informationen wie Änderungsdatum und Dienst.

Das Erstellen von FDOs ist auf zwei Arten möglich: über automatische Metadaten-Extraktion oder durch manuelle Eingabe. Bei der manuellen Erstellung wählt die Benutzerin bzw. der Benutzer ein FDO-Profil aus, woraufhin das System automatisch ein Formular generiert. Ausfüllhinweise zum Formular werden direkt aus Beschreibungen des FDO-Profils generiert (sofern diese vorhanden sind). Nach erfolgreicher Veröffentlichung wird eine Handle-PID des Objekts zurückgegeben und das publizierte FDO angezeigt.

Ein zentraler Aspekt der Architektur ist die Integration des Handle Systems. Records der Handle-PIDs enthalten PIDs aller Attribute, was eine eindeutige Identifikation des Attributs gewährleistet. Der FDO Manager übersetzt diese PIDs in menschenlesbare Namen für die Anzeige. Diese Informationen werden aus den öffentlichen Attribut-Definitionen gezogen.

Die FDOs können durch die im Projekt FAIR Data Publisher erfolgte Zusammenführung der Systeme mit denselben Werkzeugen und derselben Basisarchitektur auch im Nanopublikations-Netzwerk erstellt werden. Dabei werden Konzepte der Handle-Welt auf die Nanopublikations-Registrierung abgebildet, sodass URIs/URLs anstelle von Handle-PIDs verwendet werden. Der Handle-Identifikator bleibt dabei weiterhin in den Assertionen der Nanopublikationen enthalten. Insbesondere wurde ein Adapter zum Nanopublikations-Netzwerk entwickelt, sodass für die Erstellung von FDOs ein DOIP-Endpunkt (Digital Object Interface Protocol) bereitsteht, der sich nicht von den DOIP-Endpunkten unterscheidet, mit denen man FDOs in Handle-basierten Repositorien erstellt.

> »Ich fand es wichtig, dass sich FDOs ungehindert entfalten und verbreiten können, indem wir eine interoperable und technologie-unabhängige Basis schaffen.«
>
> **Tobias Kuhn**
> Knowledge Pixels

AUTOMATISIERTE METADATEN-EXTRAKTION

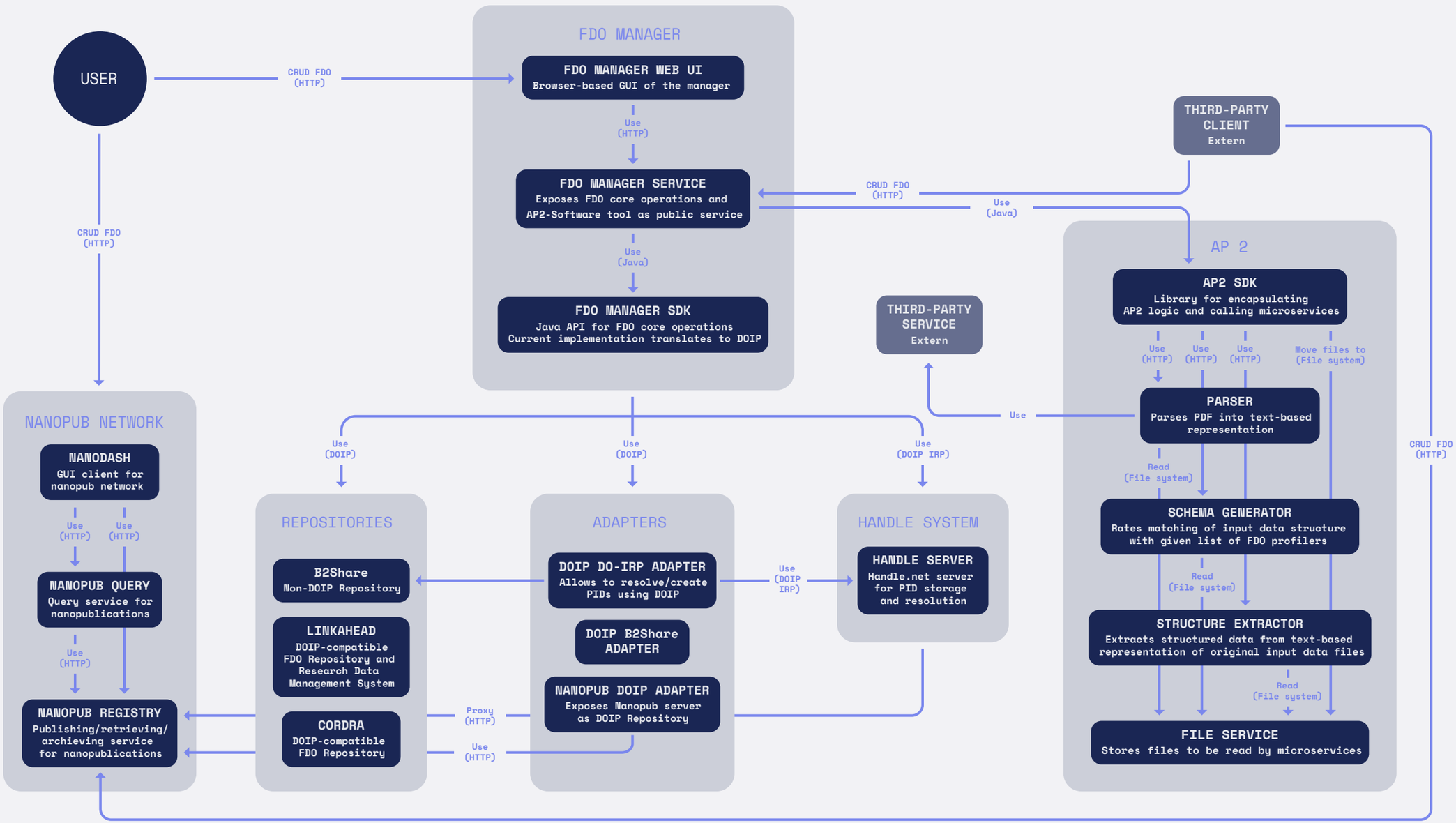
Die automatisierte Metadaten-Extraktion ist eine Funktion, die die Erstellung von FDOs durch das Hochladen von Dateien ermöglicht. Benutzerinnen und Benutzer wählen ein Profil zur Steuerung der Datenextraktion oder lassen das System Vorschläge machen, das Profil automatisch zu erkennen. Das Backend nutzt hier maschinelles Lernen.

Mit den Ergebnissen der Extraktion wird das Formular vorab befüllt, was Benutzerinnen und Benutzern die Möglichkeit gibt, die Metadaten vor der Veröffentlichung zu erweitern oder zu korrigieren. Die Architektur der Extraktion umfasst einen Parser, der Textdokumente (PDF, HTML, DOCX, u. a.) in ein leicht zu verarbeitendes Textformat (Markdown) umwandelt. Der „Structure Extractor“ nutzt dieses vorverarbeitete Dokument und befüllt ein von den Nutzerinnen und Nutzern angegebenes Profil automatisiert mit Werten. Falls kein Profil angegeben wird, werden relevante Key-Value-Paare extrahiert und an den „Schema Generator“ weitergegeben, um passende Profile zu finden. Die am besten passenden Profile werden danach befüllt, als wären sie von den Nutzerinnen und Nutzern angegeben worden. Die genutzten NLP-Technologien schließen eine LLM-Infrastruktur der GWDG für die Datenextraktion und den Dienst „docling“ für das Parsen von Dateien ein. Dabei wurde das Interface so allgemein gehalten, dass unterschiedliche Modelle konfiguriert werden können. Eine „Save Locally“-Funktion ermöglicht das Speichern der extrahierten Daten, ohne sie als FDO zu publizieren.

↗ Weitere Details im Kapitel „Tool hebt unstrukturierten Datenschatz“

FAIR DATA PUBLISHER_FDO-SERVICE-ARCHITEKTUR

•FDO-SERVICE-ARCHITEKTUR•



Die verteilte FDO-Service-Architektur integriert verschiedene etablierte Standards und Protokolle. Sie zielt darauf ab, durch eine klare Trennung von Kern-Services, spezialisierten Operations-Diensten und Adaptern eine hohe Flexibilität und Erweiterbarkeit zu ermöglichen. Der FDO Manager mit Web UI bildet das primäre Interface für Nutzerinnen und Nutzer. Als Service stellt er öffentliche FDO Core Operations bereit. Das Nanopub Network ist ein weiteres Herzstück und dient der Verwaltung semantischer Nanopublikationen. Die Repositories-Schicht integriert Systeme wie Cordra und LinkAhead mit dem Handle System, welches persistente Identifikatoren (PIDs) vergibt. Die Gruppe der Operations beinhaltet spezialisierte Services, die verschiedene Aufgaben bei der Datenverarbeitung und -konvertierung übernehmen. Der PDF2CSV-Dienst ist für die Konvertierung von PDF-Dokumenten in ein FDO-konformes Format zuständig. Der Structure Extractor dient der Extraktion strukturierter Daten, während der Schema Generator FDO-Profile abgleicht. Ein File Service unterstützt diese Operations-Dienste, indem er Dateien speichert, die von den Mikroservices gelesen werden. Die Adapter stellen eine entscheidende Komponente dar, um die Interoperabilität zwischen den verschiedenen Systemen und Protokollen zu gewährleisten. Die Autonomen FDOs (aFDOs) können selbstständig miteinander interagieren und komplexe Berechtigungsstrukturen verwalten.

FAIR DATA PUBLISHER_FDO-SERVICE-ARCHITEKTUR

DEZENTRALER ABFRAGEDIENST

Die Architektur ist mit einem dezentralen Abfragedienst ausgestattet, der sowohl für Entwicklerinnen und Entwickler gedachte SPARQL-Schnittstellen als auch vereinfachte APIs für spezifische Anwendungsfälle bereitstellt. Verschiedene Adapter vermitteln zwischen den unterschiedlichen Protokollen und ermöglichen die nahtlose Integration der Backend-Systeme.

Der Dienst „Nanopub Query FDO“ ist eine FDO-spezifische Version des „Nanopub Query“, der als leistungsstarkes und vielseitiges Werkzeug zum Abfragen und Filtern von FDOs dient. Er bietet verschiedene Repositorien, darunter „full“ (speichert alle Nanopublikationen), „pubkey“ (pro öffentlichem Schlüssel) und „type“ (pro Nanopublikations-Typ).

Eine allgemeine SPARQL-Abfrageschnittstelle ist vorhanden, die für Entwicklerinnen und Entwickler gedacht ist, um die API zu verstehen und beliebige SPARQL-Abfragen auszuführen. Um die Komplexität von SPARQL für reguläre Entwicklerinnen und Entwickler zu reduzieren, sind spezifische, fokussierte APIs bereitgestellt. Dazu gehören „FDO Text Search“ für Freitext-Suchen, „Find FDO by Reference“ zur Suche nach FDOs anhand von Referenz-IDs (Handles oder URIs) und „FDO Feed“, eine Liste der neuesten FDOs/Nanopublikationen, optional filterbar nach Ersteller.

Die spezifischen APIs basieren auf einer allgemeineren API über die Repositories. Eine Nanopublikation selbst kann eine SPARQL-Vorlage mit Platzhaltern definieren, die zur Generierung benutzerinnen- und benutzerdefinierter APIs verwendet wird. Dadurch bleibt die zugrunde liegende SPARQL-Komplexität verborgen. Der FDO Manager nutzt ebenfalls diesen Abfragedienst über einen Nanopub-Adapter. Die Funktionalität, die neueste Version einer Nanopublikation zu ermitteln, ist im Kern implementiert.

ADAPTER VERBINDEN DIE SYSTEMWELTEN

Adapter bilden eine essenzielle Schicht zur Vermittlung zwischen verschiedenen Protokollen. Sie ermöglichen die nahtlose Integration heterogener Backend-Systeme unter dem einheitlichen Digital Object Interface Protocol (DOIP).

Der DOIP-DO-IRP-Adapter verbindet die DOIP-Welt mit dem nativen Handle-System. Er übersetzt DOIP-Anfragen in Handle-Protokoll-Kommandos und transformiert Handle-Responses zurück ins DOIP-Format. Die Übersetzung erfolgt auf mehreren Ebenen: Protokoll-Transformation (JSON zu Handle-Protokoll), Attribut-Mapping (DOIP-Felder zu Handle-Typen) und Datentyp-Transformation (strukturierte Objekte zu Byte-Arrays). Der Adapter berücksichtigt die Besonderheiten des Handle-Systems, wie nicht-atomare Updates, und implementiert eine Quasi-Transaktionslogik.

Der Nanopub-DOIP-Adapter löst die konzeptionelle Herausforderung, unveränderliche Nanopublikationen, über das auf Veränderbarkeit ausgelegte DOIP-Protokoll zugänglich zu machen. Bei Update-Operationen erstellt der Adapter neue Nanopublikationen, da bestehende nicht modifiziert werden können. Die neue Version referenziert ihre Vorgänger in der Provenienz, wodurch eine Versionskette entsteht. Der Adapter verwaltet einen Index dieser Versionen und präsentiert über DOIP stets die aktuelle Version, während die Historie im Hintergrund erhalten bleibt. Für die Datenextraktion konstruiert der Adapter SPARQL-Anfragen.

FDO-PROFILE UND TYPE REGISTRY

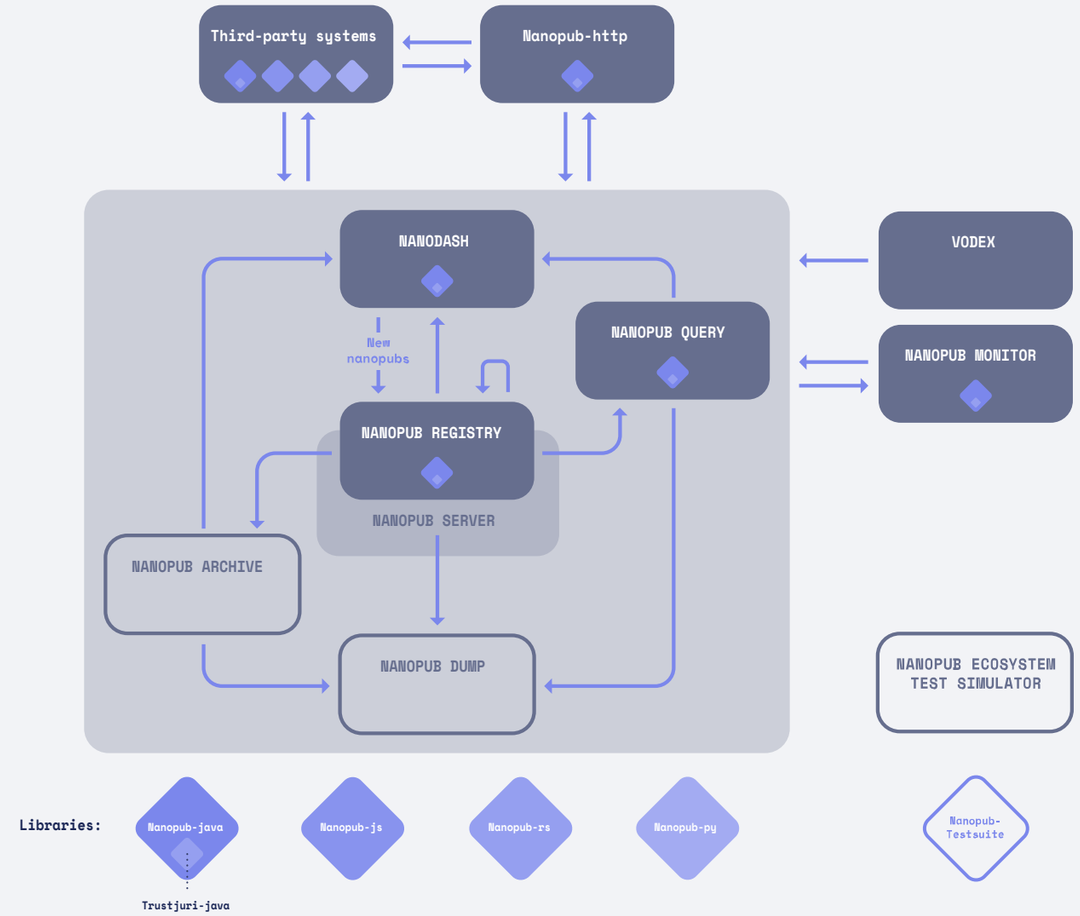
Das Profil- und Typsystem wird über eine eigene Type Registry verwaltet, die sowohl JSON-Schema-basierte Profile für Handle-FDOs als auch SHACL-basierte Bedingungen für Nanopublikations-FDOs unterstützt. Diese läuft in einer Testbed-Umgebung und dient als unabhängige Software für die Erstellung von Profilen und Attribut-Definitionen. Ihre praktische Anwendbarkeit wird durch die Integration in bestehende Systeme wie den EDC-basierten (Eclipse Dataspace Components) Datenraum „BatCAT“ und die Forschungsdatenplattform „Coscine“ demonstriert.

↗ Weitere Details im Kapitel zu den Use Cases

Profile haben einen Bezeichner, einen Namen und eine Beschreibung und definieren Attribute wie „Creators“ oder „Open Access“. Sie sind im Manager oder anderer Software nutzbar, um mit FDOs zu interagieren. Jedes FDO muss auf ein FDO-Profil verweisen, das es über ein dafür vorgesehenes Attribut in seinem FDO-Record einhält. Dieses Attribut besitzt einen PID, der auf eine Profil-Attribut-Definition verweist und dessen Wert ein PID ist, der eine Profildefinition referenziert.

Eine FDO-Profildefinition enthält eine Liste von obligatorischen und optionalen Attributen, die im FDO-Record vorhanden sein müssen. Denn das FDO muss seiner Profildefinition entsprechen, was bedeutet, dass alle obligatorischen Attribute vorhanden sein müssen. FDO-Profildefinitionen und Attribute innerhalb der Profildefinitionen müssen in offenen Registern definiert sein. Im Handle-System werden Profile oft als JSON-Schema repräsentiert. Die Data Type Registry kann zur Registrierung neuer Profil- und Attribut-Definitionen basierend auf JSON-Schema verwendet werden.

Alle Handle-basierten FDO-Profildefinitionen in FAIR Data Publisher enthalten mindestens die Attribute „FdoProfile“ (verweist auf die Profildefinition) und „FdoService“ (verweist auf einen Service FDO zur DOIP-Nutzung). Optional können Attribute wie „FdoStatus“ (Status des FDOs), „FdoRights“ (Rechtespezifikation), „FdoMimeType“ (MIME-Typ), „DataRefs“ (Referenzen zu anderen FDOs mit Bit-Sequenzen) und „DataUris“ (URIs zu Bit-Sequenzen) hinzugefügt werden. Semantische Flags werden verwendet, um anzugeben, ob ein Attributwert ein anderes FDO oder eine Datenbitsequenz referenziert. Für Nanopublikations-basierte FDOs sind rdf:type und dct:conformsTo minimale Attribute. Profile für Nanopublikationen nutzen SHACL zur Definition von Bedingungen.



•NANOPUBLIKATIONEN•

Das Nanopub Network ist ein weiteres Herzstück der Architektur und dient der Verwaltung semantischer Mikropublikationen. Es umfasst den Nanopub Server, der verschiedene spezialisierte Dienste koordiniert. Dazu gehören die Nanopub Registry für die Registrierung und Verwaltung, der Nanopub Query als Abfrage-Service und das Nanopub Archive für die Langzeitarchivierung. Eine grafische Benutzerinnen und Benutzeroberfläche, Nanodash, fungiert als Client für das Nanopub Network.

AUTONOME FDOS (AFDOS)

Ein besonders innovatives Element bilden „autonomen FDOs“ (aFDO). Sie verfügen über interne Operationen und können eigenständig auf Ereignisse reagieren. Diese aFDOs durchlaufen einen neunstufigen Entscheidungsprozess und protokollieren ihre Aktivitäten vollständig, wodurch sie interaktiver und selbstständiger werden als bisherige FDOs.

Als Proof-of-Concept ist hierzu ein in sich geschlossenes System implementiert (alles in einem Container), das sowohl die FDO-Spezifikation einhält als auch FDOs autonomer werden lässt. Klassische FDOs führen Logik extern aus, während autonome FDOs Operationen intern ausführen und Regeln für Zugriff und Interaktion verstehen. Sie protokollieren Ereignisse und Laufzeit und benötigen daher eine Laufzeitumgebung für Interaktion und Operationsausführung.

Es wurden vier Beispiel-FDOs erstellt: zwei Software-FDOs und zwei Datensatz-FDOs. Jedes FDO besitzt ein Profil, einen Typ, Metadaten und verweist auf das Bit-Sequenz-Objekt. Ein neues, wichtiges Konzept ist die Kompatibilität von Operationen, die als Attribut-Wert-Paar im Profil der Operation definiert ist. Ohne diese Kompatibilität wäre unklar, welches FDO welche Operation auf einem anderen ausführen kann.

Eine Echtzeit-Visualisierung zeigt den Status des Systems und der FDOs. Autonome FDOs besitzen Zertifikate für den Zugriff und führen Logs (z. B. „Analysis Log“ für Software, „Access Log“ für Datensätze). Ein Ereignis wie das Erstellen oder Aktualisieren eines Datensatzes löst einen Aufruf eines autonomen FDOs aus. Der Prozess umfasst damit eine Kaskade autonomer Entscheidungen: Ereignisauslösung, Suche nach kompatiblen FDOs, Aufruf durch Software-FDO, Authentifizierung (derzeit als Mocking realisiert), Operationsausführung, Ergebnisrücksendung und Protokollierung von Zugriff und Ergebnis. Protokolleinträge enthalten Zeitstempel, Status, angewandte Operation/Methode und Ergebnis. Das System validiert die Existenz von Zertifikaten für alle FDO-Operationen und erzwingt autonome Schutzmaßnahmen.

QUERY-OPERATIONEN

Das System bietet verschiedene Suchmechanismen mit unterschiedlichen Abstraktionsebenen. Generische Abfragen (Op.QuerySparql) akzeptieren uneingeschränkte SPARQL-Abfragezeichenfolgen als Eingabe und führen sie auf der RDF-Repräsentation der FDO-Records aus. Vorlagenbasierte Abfragen (Op.QueryTemplate) verwenden Abfragevorlagen, die durch ihre PIDs identifiziert werden und Platzhalter enthalten können. Eine Abfrageanforderung an eine vorlagenbasierte Abfrage besteht aus dem PID der Abfragevorlage und gegebenenfalls einer Reihe von Parametern.

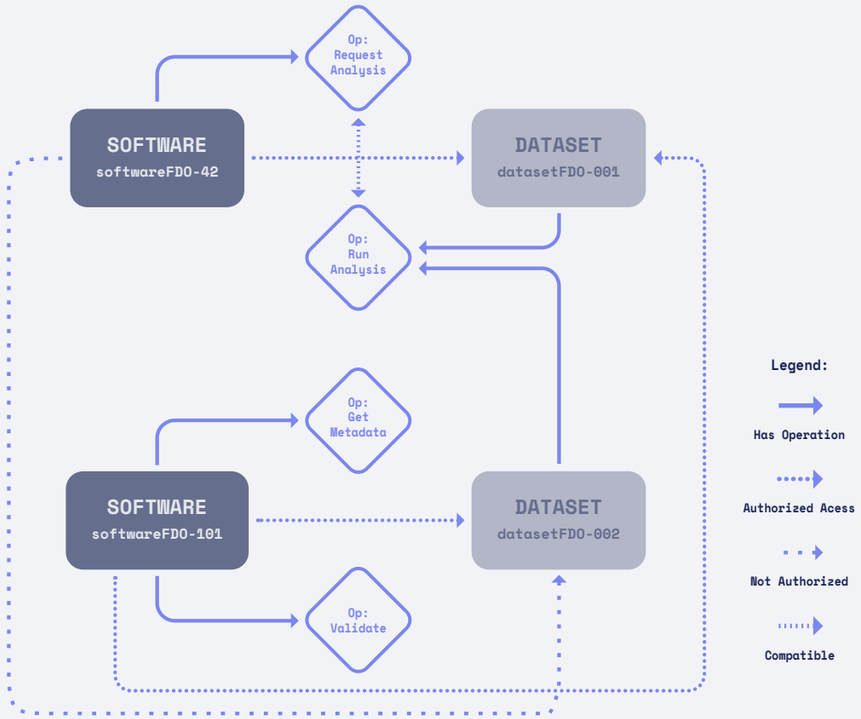
Spezifische Abfragen sind noch granularer als vorlagenbasierte Abfragen und Fragen nach einer festen Art von Daten mit einem festen Satz von Parametern. Dazu gehören: Op.QueryFreeText für Freitextsuchen in FDOs und Metadaten-Records, Op.QueryRef zum Abfragen von FDOs anhand von PIDs/URIs, auf die in ihren Metadaten-Records verwiesen wird, Op.QueryFeed zur Bereitstellung einer Liste von Abfragen pro Quelle (z. B. Tool oder Benutzerinnen und Benutzer) nach Aktualität sortiert, und Op.QueryFavorites, die alle vom Benutzerinnen und Benutzer favorisierten FDOs oder anderen Ressourcen zurückgibt.

INTEGRATION MIT DATENBANKEN UND DATENRÄUMEN

Der EDC-basierte Datenraum BatCAT dient als Beispiel, um zu testen, wie sich ein FDO-Service, der die AP2-Extraktionsfunktionalität bereitstellt, in einen EDC-Datenraum einbringen lässt. Dazu wurde ein DOIP-Endpunkt für die Schnittstelle des FDO Managers entwickelt und dieser über eine publizierte FDO-Operation beschrieben. Der Service wurde so in den Datenraum integriert, dass die Nutzung einen vorherigen Vertragsabschluss über die EDC-Komponenten voraussetzt.

Ergänzend wurde die Anbindung an die NRW-weite, generische Forschungsdatenplattform Coscine demonstriert. Coscine erstellt bereits persistente Identifikatoren (PIDs) für Projekte und Ressourcen unter Verwendung des Handle-Systems. Ein Prototyp wurde in die Produktionsversion von Coscine integriert, und alle bestehenden PIDs wurden gemäß der Prototyp-Abbildung von Eigenschaften innerhalb einer Datentyp-Registrierung aktualisiert. Abhängig von den Zugriffsrechten werden Metadatenfelder aus Projekten und Ressourcen zum PID-Record hinzugefügt. Neu erstellte PIDs in Coscine werden automatisch mit den erforderlichen Informationen zur Darstellung eines FDOs ergänzt.

Alle im Projekt entwickelten Softwarekomponenten, einschließlich der FDO-Service-Architektur, sind für die quelloffene Bereitstellung vorgesehen. So soll eine breitere Gemeinschaft die Architektur selbst nutzen und erweitern können. Eine umfassende Dokumentation der Software, einschließlich eindeutiger Schnittstellenbeschreibungen, einer Beschreibung der

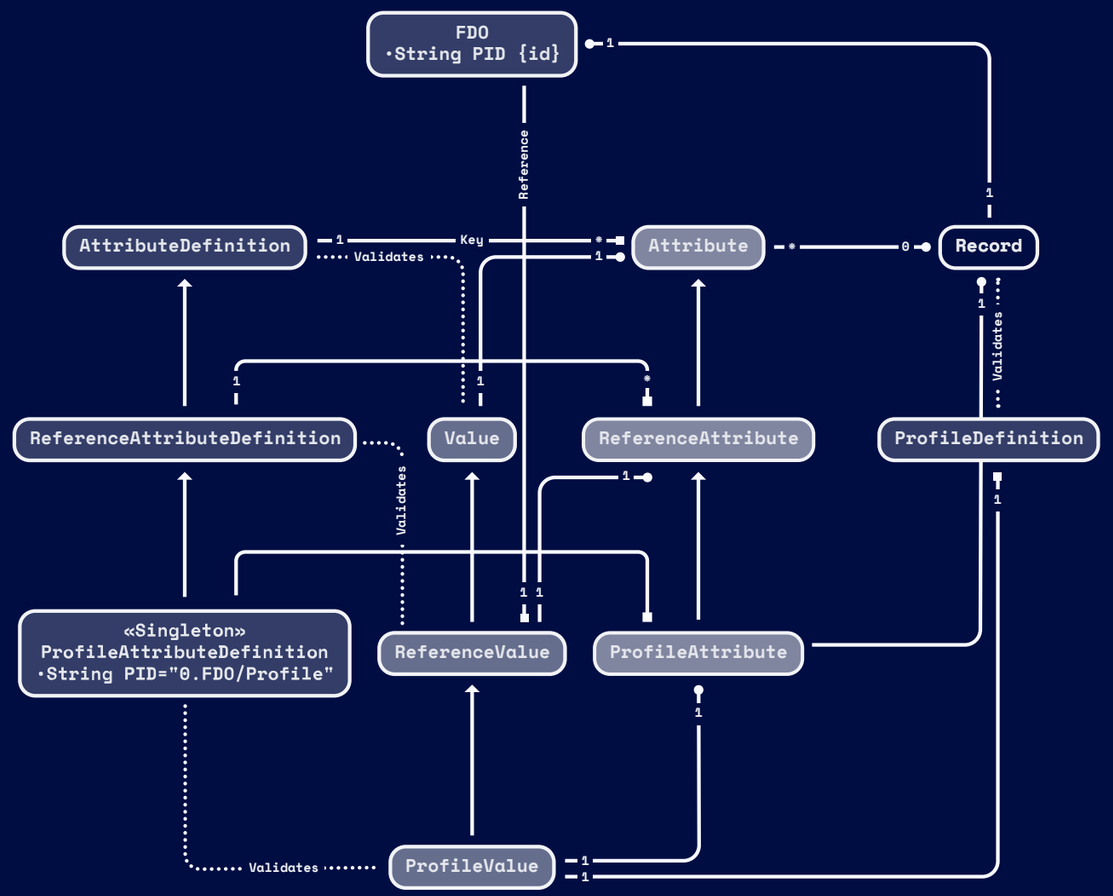


•AUTONOME FDOS•

Die Interaktion autonomer FAIR Digital Objects (FDOs) anhand zweier Software-FDOs sowie zwei Dataset-FDOs, die über definierte Operationen miteinander verbunden sind. SoftwareFDO-42 kann Analysen anfordern und ausführen. Es interagiert mit datasetFDO-001, einem FDO mit geschützten Gesundheitsdaten, das Analyseanfragen eigenständig verarbeitet. Die Verbindungen zeigen autorisierten Zugriff und die Möglichkeit zur Ausführung von Operationen wie Request Analysis und Run Analysis. SoftwareFDO-101 kann Metadaten abrufen und Validierungen durchführen. Es hat Zugriff auf beide Datensätze, darf aber selbst keine Analyseoperationen ausführen. DatasetFDO-002 (ein FDO mit Industrie-Monitoringdaten unter restriktiven Zugriffsregeln) wird in diesem Zusammenhang für Validierungsoperationen eingebunden. Die Grafik macht sichtbar, wie autonome FDOs selbstständig Operationen ausführen, Zugriffsrechte prüfen und in einem Netzwerk zusammenarbeiten.

Software-Architektur und einer Installationsbeschreibung, ist erstellt. Eine systematische Evaluierung der Software-Performance und -Usability wurde durchgeführt. Diese umfasste Funktionstests und automatisierte Regressionstests, die in einer eigenen Integrationsumgebung stattfinden. Dies stellt die Konformität mit den Anforderungen und Benutzerinnen- und Benutzerbedürfnissen sicher.

Das Konzept vereint die Stärken etablierter Technologien und adressiert gleichzeitig bestehende Herausforderungen im Datenmanagement, indem es eine zukunftsfähige Infrastruktur schafft. Die verteilte FDO-Service-Architektur auf Basis von Nanopublikationen und Handle-Records generiert so eine Reihe signifikanter Mehrwerte, welche die Verwaltung und Interaktion mit digitalen Objekten in vielfältigen Kontexten



Definition of the classes
- FDO Any dark-blue class is an FDO.
- Attribute (mid-blue): Attributes are components of a Record. Attributes consist of a key, which is a reference to an Attribute Definition, and a Value.
- Value (light-blue) Attribute values. This could be a scalar value like a simple string, a reference to another FDO or a set, list or tuple.

•FDO-SPEZIFIKATIONEN (FDO-SPECS)•

Die vereinfachten FDO Spezifikationen ohne die Servicekomponenten. Im Zentrum stehen die Records, die aus verschiedenen Attributen aufgebaut sind. Jedes Attribut verweist über eine Attribut-Definition auf seinen Typ und enthält einen Wert, der entweder ein einfacher skalarer Datentyp, eine komplexe Struktur wie Liste oder Tupel oder auch eine Referenz auf ein anderes FDO sein kann. Profile bilden dabei den Rahmen, indem sie die obligatorischen und optionalen Attribute festlegen, die ein FDO enthalten muss. Ergänzt werden sie durch Profil-Attribute und Profil-Werte, die sicherstellen, dass FDOs gültig und interoperabel bleiben. Beziehungen wie „validates“ oder „references“ verdeutlichen, wie Attribute, Werte und Definitionen miteinander verknüpft sind. Dunkelblau steht für FDOs, Mittelblau für Attribute und Hellblau für Werte. Die Grafik macht damit deutlich, wie FDOs konzeptionell strukturiert sind, wie Profile die Konsistenz sichern und wie Referenzen den Aufbau komplexer, verknüpfter Datenobjekte ermöglichen.

→ Autonome FAIR Digital Objects

„aFDOs“ repräsentieren einen Paradigmenwechsel in der Verwaltung digitaler Objekte. Während klassische FDOs statische Datenkapseln mit strukturierten Metadaten darstellen, die über standardisierte Protokolle wie DOIP passiv auf Anfragen warten, besitzen aFDOs eigene Entscheidungs- und Handlungskompetenzen. Sie evaluieren autonom die Berechtigung der Anfrage, treffen Entscheidungen über die Gewährung oder Verweigerung des Zugriffs und protokollieren jede Interaktion für spätere Audits. Ein aFDO vereint die strukturelle Integrität eines klassischen FDO mit der Fähigkeit zur eigenständigen Aktion. Es wird vom passiven Datenobjekt zum aktiven Teilnehmer im digitalen Ökosystem.

grundlegend verbessern. Diese Fortschritte sind entscheidend, um den Anforderungen einer zunehmend digitalisierten Wissenschaft, Verwaltung und Wirtschaft gerecht zu werden.

Obwohl signifikante Fortschritte erzielt wurden, konnten im Rahmen des Projektes nicht alle Entwicklungen abgeschlossen werden, wodurch sich konkrete Bedarfe für weitere Schritte ergeben. Die identifizierten Bereiche erfordern weitere Konzeptionierung, Implementierung und Professionalisierung, um das System vollständig und produktionsreif zu gestalten.

FAZIT UND EMPFEHLUNGEN

MEHRWERTE

+ **Verbesserte Interoperabilität:** Die Architektur überbrückt erfolgreich die „Handle-Welt“ der FDOs mit der „Linked Data-Welt“ der Nanopublikationen und schafft dadurch eine technologisch unabhängige Basis für den systemübergreifenden Austausch und die Nutzung digitaler Objekte.

+ **Ermöglichung maschineller Aktionen:** Durch die Strukturierung als FDOs und die Integration semantischer Web-Technologien sind digitale Objekte nun maschinenlesbar und -aktionsfähig, was eine hochautomatisierte Verarbeitung und Analyse von Daten ermöglicht.

+ **Garantierte Datenpersistenz und Unveränderlichkeit:** Die Nutzung eines dezentralen Nanopublikations-Netzwerks gewährleistet die Unveränderlichkeit (Immutability) von Daten und Identifikatoren, wodurch eine langfristige und manipulationssichere Verfügbarkeit sichergestellt wird, ergänzt durch institutionelle Vereinbarungen.

+ **Hohe Skalierbarkeit für große Datenmengen:** Das Design der Nanopublikations-Architektur ermöglicht die effiziente Verwaltung sehr großer Mengen digitaler Objekte durch horizontale Skalierung und dezentrale Bearbeitung von Anfragen.

+ **Dezentrale Datengovernance durch autonome FDOs:** Autonome FDOs (aFDOs) ermöglichen selbstständige Entscheidungsfindung, Interaktion und die Durchsetzung granularer Zugriffsrechte ohne eine zentrale Kontrollinstanz, was die Datensicherheit und -integrität maßgeblich erhöht.

INTERVIEW MIT TOBIAS KUHN, KNOWLEDGE PIXELS

>>Was war denn die größte Motivation, die Konzepte aus dem Vorgängerprojekt FDO One mit Nanopublikationen zu verknüpfen?
<<Die Motivation war, dass die Idee der FDOs momentan ihre Wirkung nicht vollständig entfalten kann, da verschiedene Ansätze bestehen, die sich überschneiden, aber konzeptionell wie technologisch nicht vollständig interoperabel sind. Auf der einen Seite gibt es den Handle-basierten Ansatz und auf der anderen Seite den Ansatz basierend auf Semantic-Web-Technologien. Dass diese Welten bis jetzt nicht zusammengefunden haben, hat mich gestört. Ich fand es wichtig, dass wir diese Diskrepanz auflösen, sodass sich FDOs dann ungehindert entfalten und verbreiten können, indem wir eine interoperable und technologieunabhängige Basis schaffen.

+ Leistungsstarke, flexible Abfragefunktionen: Ein integrierter, mächtiger SPARQL-Endpunkt und spezifische APIs ermöglichen schnelle, vielseitige Such- und Filterfunktionen für FDOs, einschließlich Freitextsuche und versionshistorischer Abfragen.

NOTWENDIGE WEITERE SCHRITTE

> *Eine dauerhaft implementierte Laufzeitumgebung für autonome FDOs sowie eine tiefgehende Analyse der benötigten Ressourcen stehen noch aus.*

> *Die Testbed-Infrastruktur bedarf einer Professionalisierung und der Schaffung redundanter Dienste, um volle Produktionsreife zu erreichen.*

> *Die umfassenden API-Spezifikationen für die verschiedenen Abfrage-Operationen müssen noch vollständig dokumentiert werden.*

> *Die Skalierung und tiefergehende Integration autonomer FDOs mit weiteren Use Cases und dem FDO Manager ist ein zukünftiger Entwicklungsschritt.*

> *Eine Weiterentwicklung, sodass Abfragen neben Nanopublikationen auch Handle-basierte FDOs berücksichtigen, ist zentral für die Akzeptanz dieses dezentralen Ökosystems.*

> *Ein generisches und vollständig spezifiziertes Typensystem zur Assoziierung von FDOs mit ihren Operationen ist noch zu erarbeiten, um die Interoperabilität zu maximieren.*

> *Verbesserte Sichtbarkeit von Profilen: Zusätzlich zu den Referenzen auf die Registratur, in der die Profile veröffentlicht sind, sollten weitere Informationen zu den Profilen im FDO Manager leicht sichtbar gemacht werden.*

> *Das entwickelte Werkzeug ist zwar in der Lage, komplexe Dokumente mit dutzenden Seiten und mehreren MB Größe in wenigen Minuten zu verarbeiten, für einen produktiven Betrieb sollte die Performance jedoch verbessert werden, entweder durch Unterstützung der Extraktionsaufgaben der KI durch klassische Algorithmen oder durch Spezialisierung (Finetuning) kleiner LLMs.*

>>Und welche technische Herausforderung war bei dieser Zusammenführung von FDO One und Nanopublikationen am schwierigsten zu meistern?
<<Die größte Herausforderung bestand darin, eine konzeptionell solide und technisch machbare Lösung zu entwickeln, welche die teilweise nicht 100 % kompatiblen Konzepte der Handle-Welt und der Nanopublikations-Welt überbrückt. Besonders die Verbindung von veränderbaren FDO-Records mit unveränderlichen Nanopublikationen erforderte Integrationsarbeit, mit dem Resultat, dass unser Konzept die Stärken beider Ansätze kombiniert.

>>Welchen konkreten Nutzen zieht man denn aus der Kombination dieser beiden Ansätze für die zukünftige Handhabung digitaler Objekte?
<<Durch diese Kombination entstehen stabilere und hochskalierbare FDOs, welche die Vorteile von Linked Data-Technologien und mächtigen SPARQL-Abfragen nutzen können, um maschinelle Aktionen und eine verbesserte Interoperabilität zu gewährleisten.

Verweise FDO-Service-Architektur:
↗ Handle System Technical Manual: Technische Details des Handle Systems
↗ Nanopublikationen Architektur: Informationen zur Architektur des Nanopublikationsnetzwerks
↗ W3C Anatomy of a Nanopublication: Referenz zur Struktur von Nanopublikationen
↗ Riojournal Nanopublikationen: Weitere Informationen zu Nanopublikationen
↗ Data Type Registry (Testbed): Registrierung von Profil- und Attributdefinitionen
↗ Type API: Generiert JSON-Schemata für Profile und Attributdefinitionen
↗ DOIP Specification: Spezifikation des Digital Object Interface Protocol
↗ DO-IRP Specification: Spezifikation des Digital Object Identifier Resolution Protocol
↗ FAIR Data Publisher Architecture Documentation (Konzept und Methodik): Dokumentation des Konzepts und Modells für AP1
↗ FDO Local Testbed GitLab: Ermöglicht die lokale Bereitstellung notwendiger Tools/Services
↗ nanopub-java-fdo: Java-Bibliothek für Nanopublikationen mit FDO-Unterstützung
↗ nanopub-py-fdo: Python-Bibliothek für Nanopublikationen mit FDO-Unterstützung
↗ EDC-FDO-Adapter GitHub: Repository für den Dataspace Demo Quick-Start Guide
↗ FDO One Homepage: Homepage des FDO One Vorprojekts

FAIR DATA PUBLISHER_FDO-SERVICE-ARCHITEKTUR

¬

Evaluierung der Typing-Mechanismen

¬

Das Thema Typisierung (Typing) ist fundamental für die Realisierung maschinenaktionsfähiger, interoperabler Dateninfrastrukturen durch FAIR Digital Objects. Die Möglichkeit zur automatisierten Verarbeitung und Ausführung von Operationen auf diesen Objekten ist dabei zentral.

¬

Das Typsystem für FDOs, das Regeln zur Assoziierung von Objekten mit anwendbaren Operationen festlegt, ist inspiriert von Prinzipien der Objektorientierten Programmierung (OOP) wie Abstraktion und Kapselung. Die Typisierung muss Informationen darüber liefern, wie das Objekt präsentiert wird (Kodierungsformate), welche Rolle es als informationelles Artefakt spielt und worauf es in der realen Welt verweist. Um ein interoperables Ökosystem zu gewährleisten, war eine strukturierte Analyse generischer Mechanismen zur Operations-Assoziierung notwendig.

¬

Kernmechanismen zur Operations-Assoziierung

¬

Das Paper „A Comparative Analysis of Modeling Approaches for the Association of FAIR Digital Objects Operations“ (Blumenröhr et al., 2025) bewertet drei Kernmodelle, die als Typsysteme für FDOs fungieren:

¬

Record Typing (Datensatz-Typisierung): Dies ist der strukturell einfachste Ansatz. Eine Operation wird direkt im Informationsdatensatz des FDO spezifiziert und dem individuellen Objekt zugeordnet. Dies ähnelt dem Duck Typing in OOP. Der Typ wird hierbei rein durch die Konstellation der anwendbaren Operationen definiert. Da Operationen statisch im Datensatz festgelegt sind, können sie zur Instanziierungszeit nicht dynamisch geändert werden.

¬

Profile Typing (Profil-Typisierung): Operationen werden abgeleitet, indem sie mit dem Profil assoziiert werden, welches das FDO instanziiert. Das Profil fungiert als eine Abstraktion oder Klasse, deren Anforderungen alle konformen FDOs erfüllen müssen. Jedes FDO muss ein Profil als Pflichtattribut in seinem Informationsdatensatz haben. Dieser Ansatz entspricht der nominalen Typisierung in OOP.

¬

Attribute Typing (Attribut-Typisierung): Die Assoziierung einer Operation erfolgt implizit durch das Vorhandensein spezifischer Attribute im Informationsdatensatz des FDO. Operationen können selbst als FDOs (Operation FDOs) dargestellt werden und definieren die notwendigen Anforderungen an Attribute des Ziel-FDOs, um anwendbar zu sein. Dies nähert sich einer dynamischen Typisierung, da die Anwendbarkeit zur Laufzeit durch Attribute bestimmt wird.

¬

Gewichtete Ergebnisse und Vergleich der Modelle

¬

Die vergleichende Analyse bewertet die Modelle anhand von Komplexität, Effizienz, Flexibilität, Granularität und Vielseitigkeit:

¬

Effizienz (Operationssuche): Record Typing ist am effizientesten, um alle Operationen zu finden, die einem FDO zugeordnet sind, da die Suche direkt im Datensatz erfolgt. Attribute Typing ist am wenigsten effizient, da es die Iteration über die Attribute aller Operationen im FDO-Ökosystem erfordert, um alle einem FDO zugeordneten Operationen zu finden.

¬

¬

Flexibilität (Updates): Attribute Typing ist am flexibelsten bei der Durchführung notwendiger Updates (z. B. Hinzufügen einer neuen Operation zu einer Menge von FDOs), da die Assoziierung implizit über Attribute im Operations-FDO definiert wird und keine expliziten Änderungen an den FDOs selbst notwendig sind. Record Typing ist in dieser Hinsicht am ineffizientesten, da für jede Assoziierung ein individuelles Update am FDO erforderlich ist.

¬

Granularität und Client-Wissen: Record Typing bietet die höchste Granularität, da jede Assoziierung individuell pro FDO festgelegt wird, was jedoch ein hohes Domänen- und Mechanismuswissen seitens des Clients erfordert. Profile Typing ist am wenigsten granular, da Operationen einer ganzen Klasse von FDOs zugeordnet werden. Dies reduziert das erforderliche Client-Wissen, da der Client lediglich ein Profil auswählen muss.

¬

Vielseitigkeit: Attribute Typing weist potenziell die größte Vielseitigkeit auf, da die Attributdefinitionen, welche die Assoziierungsbedingungen bilden, über Profile und FDOs hinweg wiederverwendet werden können. Record Typing hat die geringste Vielseitigkeit, da jede Assoziierung explizit deklariert werden muss.

¬

Interoperabilität: Die drei Assoziationsmodelle sind miteinander kompatibel und können gleichzeitig auf dasselbe FDO angewendet werden. Das Typsystem fungiert als gemeinsame Sprache, die eine konsistente Interpretation gültiger Operationen über die verschiedenen Modelle hinweg ermöglicht. Die Profile sind für die Interoperabilität von zentraler Bedeutung, da sie eine minimale, standardisierte Metadatenstruktur für alle FDOs bereitstellen.

¬

Für die Implementierung wird vorgeschlagen, die Komponenten und Regeln der Assoziationsmodelle in Graphen-Datenstrukturen zu speichern, um komplexe Prozeduren wie die beim Attribute Typing (Suche über alle Operationen) durch Caching und einfache Graphenabfragen zu kompensieren. Die Wahl des optimalen Modells hängt letztlich von den spezifischen Anforderungen der Datenumgebung, den Client-Erwartungen und den rechnerischen Einschränkungen ab.

¬

¬

¬

Zur Studie

FDO Operations Comparative Analysis (arXiv) https://arxiv.org/abs/2504.05361

¬

¬

¬

Verweise zu Typing

Digital Object Interface Protocol Specification https://www.dona.net/sites/default/files/2018-11/DOIPv2Spec_1.pdf
RDA Webseite https://www.rd-alliance.org/
FDO Forum FDO Requirement Specification https://zenodo.org/records/7782262
FDO Forum FDO Configuration Types https://zenodo.org/records/7825703
FDO Technical Overview https://zenodo.org/records/7824714
RDA Recommendation on PID Kernel Information https://www.rd-alliance.org/system/files/RDA%20Recommendation%20on%20PID%20Kernel%20Information_final.pdf

¬

¬

¬

¬

¬

¬

FAIR DATA PUBLISHER_FDO-SERVICE-ARCHITEKTUR

Tool hebt unstrukturierten Datenschatz

Das im Rahmen des Projekts FAIR Data Publisher entwickelte Software-Tool zur Generierung von FDOs aus unstrukturierten Daten hat die elementare Aufgabe, als Scharnierfunktion zwischen generischen FDO-Infrastrukturen und domänenspezifischen Datenformaten zu wirken.

¬ Ein Großteil der in spezifischen Domänen vorliegenden Daten ist unstrukturiert und nicht direkt für das Training von KI-Modellen verwertbar. Das primäre Ziel des Tools ist es daher, unstrukturierte Domänendaten automatisch in strukturierte, domänen-spezifische FDOs zu konvertieren. Dieses Tool wird als zusätzlicher Service innerhalb der im Projekt geschaffenen FDO-Service-Architektur angeboten, um die Menge verfügbarer FDOs im FDO-Datenraum zu erhöhen. Es soll die Möglichkeit einer generischen Datenstrukturierung durch Künstliche Intelligenz in verschiedenen Anwendungsfällen demonstrieren.

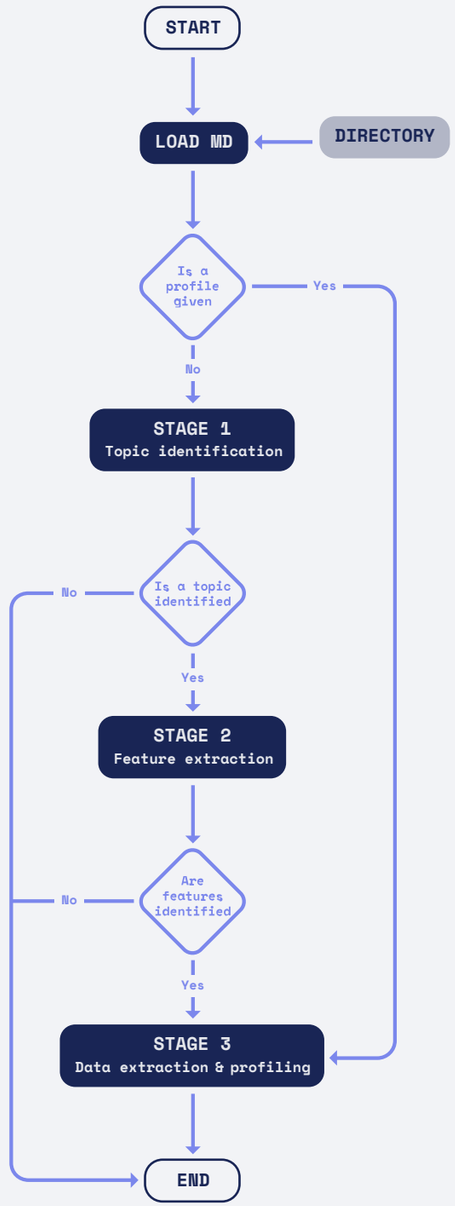
Die entwickelte Lösung basiert auf einer modularen Microservice-Architektur, die eine flexible und skalierbare Verarbeitung unterschiedlicher Datenformate ermöglicht. Der Verarbeitungsprozess gliedert sich in mehrere sequentielle Phasen. Zunächst werden die Eingabedaten, z. B. PDF-Dokumente, in ein generisches Textformat überführt. Aus diesem Textformat werden gezielt die strukturierten Daten extrahiert und in dem gängigen strukturierten Format JSON zwischengespeichert, bevor sie über den FDO Manager als FDO veröffentlicht werden können. Dedizierte Microservices übernehmen spezifische Schritte im Bereich des Data-Preprocessings und der Wissensexploration und kommunizieren über standardisierte Service-APIs, spezifiziert auf Basis der OpenAPI-Spezifikation.

TECHNOLOGISCHE KOMPONENTEN DER DATENEXTRAKTION

Für die Textextraktion aus PDF-Dokumenten kommen moderne Optical Character Recognition (OCR)-Technologien zum Einsatz, um sowohl maschinenlesbaren Text als auch gescannte Inhalte zu verarbeiten. Die extrahierten Textinhalte werden zunächst als Markdown-Dateien formatiert, um die ursprüngliche Dokumentstruktur zu bewahren und eine effiziente Weiterverarbeitung durch nachgelagerte Services zu ermöglichen. Ein zentraler Baustein ist der Einsatz von Large Language Models (LLMs) in Kombination mit einem generischen Prompt-Generator, um spezifisches, für die jeweiligen Anwendungsfälle relevantes Wissen zu identifizieren und zu strukturieren. Das durch die LLMs extrahierte Wissen wird in einem strukturierten JSON-LD-Format gespeichert (Text2JSON).

PROFILE-MATCHING UND FDO-GENERIERUNG

Von Nutzerinnen und Nutzern angegebene FDO-Profile werden eingesetzt, um gezielt benötigte Informationen zu extrahieren. Wird auf die Angabe eines FDO-Profils verzichtet, werden existierende Profile mit den im Dokument enthaltenen



•STRUCTURE EXTRACTOR•

Die intelligente Feedback-Architektur des Structure Extractors ermöglicht durch iterative Validierung eine hohe Qualität der Datenextraktion. Besonders hervorzuheben ist die schrittweise Strukturierung von unstrukturierten Daten durch die drei spezialisierten Verarbeitungsstufen, die jeweils eine eigene Qualitätskontrolle implementieren.

FAIR DATA PUBLISHER_TOOL ZUR DATENSTRUKTURIERUNG

INTERVIEW MIT FRANK TRISTRAM, KIT

>>Was war denn die größte technische Herausforderung bei dieser automatisierten Extraktion?
<<Technische Datenblätter von über 30 Seiten mit komplex verschachtelten Tabellen, fachlichen Abkürzungen und unklaren Bezügen stellen selbst Fachleute manchmal vor Übertragungs- und Verständnisprobleme. Dass die LLMs relevante Informationen aus solchen Quellen sogar fachübergreifend extrahieren und sie korrekt mit FDO-Profilen abgleichen ist schon beeindruckend. Gelingt deshalb aber auch nicht immer. Jedoch erreicht die Fehlerquote bereits fast die von Menschen.

>>Welchen Nutzen bringt diese Technologie letztendlich für die Anwender und die zukünftige Datenlandschaft?
<<Diese Technologie ermöglicht die effiziente Generierung von FDOs aus unstrukturierten Daten und bildet eine entscheidende Grundlage für zukünftige autonome KI-Systeme. Mit immer besseren LLMs wird die Zuverlässigkeit selbst für hochkomplexe Fragen bald die menschliche Zuverlässigkeit übersteigen.

>>Müssen User das FDO-Profil manuell auswählen, oder kann das System das am besten passende Profil automatisch erkennen?
<<Man muss das Profil nicht zwingend manuell auswählen, denn das System nutzt eine Funktion zur automatischen Profilerkennung, die auf der Analyse der hochgeladenen Datei basiert. Dazu wird ein hybrides Profile-Matching-Verfahren verwendet, das semantische Ähnlichkeiten und strukturelle Übereinstimmungen kombiniert.

Informationen in einem hybriden Profile-Matching-Verfahren abgeglichen. Zum Einsatz kommen hier subsymbolische und symbolische Methoden, um die Genauigkeit bei der Profilzuordnung zu optimieren. Das subsymbolische Verfahren basiert auf der Berechnung von Cosinus-Distanzen von Worteinbettungen (Embedding) zur Quantifizierung semantischer Ähnlichkeiten, während das symbolische Verfahren String- und Graph-Edit-Distanzen für strukturelle Übereinstimmungen verwendet. Basierend auf diesen Matching-Ergebnissen erstellt das System ein Ranking der verfügbaren Profile und generiert Mapping-Vorschläge. Im Falle geringer Übereinstimmung können über Anbindungen an den FDO-Manager neue FDO-Profile erstellt werden. Im letzten Schritt wird das FDO über die API des FDO-Managers generiert.

NUTZUNGSSCHNITTSTELLE UND INTERAKTION

Das Software-Tool wird durch eine webbasierte grafische Benutzerinnen- und Benutzeroberfläche (GUI) ergänzt, die eine intuitive Bedienung ohne spezielle IT-Kenntnisse ermöglicht. Die GUI ermöglicht es, Dateien hochzuladen und die automatische Datenextraktion durch die AP2-Maschinerie zu nutzen. Abgesehen von der optionalen Angabe eines FDO-Profils müssen Nutzende keine Eingaben machen, um die automatisierte Extraktion zu nutzen. Nach der Extraktion werden die Informationen zur Überprüfung, Korrektur oder Ergänzung präsentiert. Es ist die lokale Speicherung extrahierter Daten ohne direkten Publikationsprozess möglich. Ein wesentliches Feature für die User-Experience, das die Einstiegshürde senkt, ist die oben genannte automatische Profilerkennung und Präsentation passender FDO-Profile basierend auf der Analyse der hochgeladenen Dateien. Die GUI unterstützt zudem die gleichzeitige Veröffentlichung von Dateien, in dem diese in ein Repositorium hochgeladen werden und im FDO-Record referenziert werden.

OFFENER QUELLCODE, DOKUMENTATION UND EVALUIERUNG

Eine Bereitstellung der im Projekt entwickelten Software als offener Quellcode ist vorgesehen. Alle für die lokale Ausführung des Tools benötigten Software-Komponenten werden ebenfalls aus Open Source-Repositorien zur Verfügung stehen, sodass die Verwertbarkeit der Arbeitsergebnisse nicht eingeschränkt wird. Die Dokumentation der Software, einschließlich ihrer Bedienung, Schnittstellen und Installationsanweisungen, erfolgt nach dem „Documentation-as-Code“-Ansatz und wird ebenfalls auf einem Git-Repositorium veröffentlicht. Zur Evaluierung von Performance und Usability wurde ein Nutzungs-Anforderungskatalog erstellt und die Nutzungs-Performance sowie die User Experience gemessen. Das Feedback der Nutzerinnen und Nutzer konnte überwiegend noch vor Projektende bei der Implementation berücksichtigt werden.

FAZIT UND EMPFEHLUNGEN

MEHRWERTE

+ **Automatisierte FDO-Generierung:** Es wurde eine generische Software-Lösung entwickelt, die unstrukturierte Daten, z. B. PDF-Dokumente, automatisch analysiert, aufbereitet und als strukturierte FAIR Digital Objects (FDOs) einschließlich vollständiger Metadaten bereitstellt.

+ **Effiziente Datenstrukturierung:** Die Hauptfunktion des AP2-Tools ist die effiziente Extraktion strukturierter Daten aus unstrukturierten Dateien, was eine Transformation menschlich lesbarer Dokumente in maschinell verarbeitbare Daten ermöglicht.

+ **Vereinfachter FDO-Erstellungsprozess:** Die in AP1 und AP2 entwickelten Funktionalitäten vereinfachen den Workflow zum Erstellen von FDOs insbesondere für menschliche Nutzerinnen und Nutzer.

+ **Benutzerinnen- und Benutzerfreundliche GUI:** Eine intuitive, browserbasierte grafische Benutzerinnen- und Benutzeroberfläche (GUI) ermöglicht es Nutzerinnen und Nutzern, den entwickelten Extraktionsservice zu nutzen und Ergebnisse zu speichern oder nach Prüfung zu veröffentlichen.

+ **Intelligente Wissensextraktion:** Durch den Einsatz von Large Language Models (LLMs) werden relevante Informationen intelligent extrahiert und in strukturiertem JSON-LD-Format gespeichert, ohne dass domänenspezifische Anpassungen der Software notwendig sind.

+ **KI-gestützte Profilauswahl (Autodetection):** Das System bietet eine automatische Profilerkennung und Vorschläge für passende FDO-Profile, basierend auf der Analyse der hochgeladenen Datei, wodurch der Bedarf an Expertenwissen reduziert wird.

FAIR DATA PUBLISHER_TOOL ZUR DATENSTRUKTURIERUNG

+ **Hohe Datenqualität durch Validierung:** Der dreistufige Workflow des Structure Extractor mit integrierten Validierungsschleifen und Feedback-Systemen gewährleistet eine hohe Qualität der Informationsextraktion und die zuverlässige Strukturierung heterogener Datenbestände.

+ **Flexible und skalierbare Architektur:** Eine modulare Microservice-Architektur mit standardisierten OpenAPI-Schnittstellen sorgt für Flexibilität, Skalierbarkeit und einfache Erweiterbarkeit des Systems.

+ **Interoperabilität und Integration:** Das Projekt demonstriert die Zusammenführung der Handle- und Nanopublikationsnetzwerke und stellt eine elementare Scharnierfunktion zwischen verschiedenen Datenformaten und der generischen FDO-Infrastruktur dar.

+ **Development SDK:** Die Bereitstellung eines Java SDKs ermöglicht die einfache Integration der Business-Logik der Extraktion in andere Anwendungen.

NOTWENDIGE WEITERE SCHRITTE

> *Unterstützung zusätzlicher Eingabeformate ist für die zukünftige Entwicklung des Systems wünschenswert, um die Vielseitigkeit der Datenquellen zu erweitern. Beispielsweise könnten auch Grafiken in die Informationsgewinnung einfließen.*

> *Verbesserung der KI-Algorithmen ist notwendig, um die Genauigkeit und Effizienz der intelligenten Wissensextraktion aus unstrukturierten Daten kontinuierlich zu optimieren.*

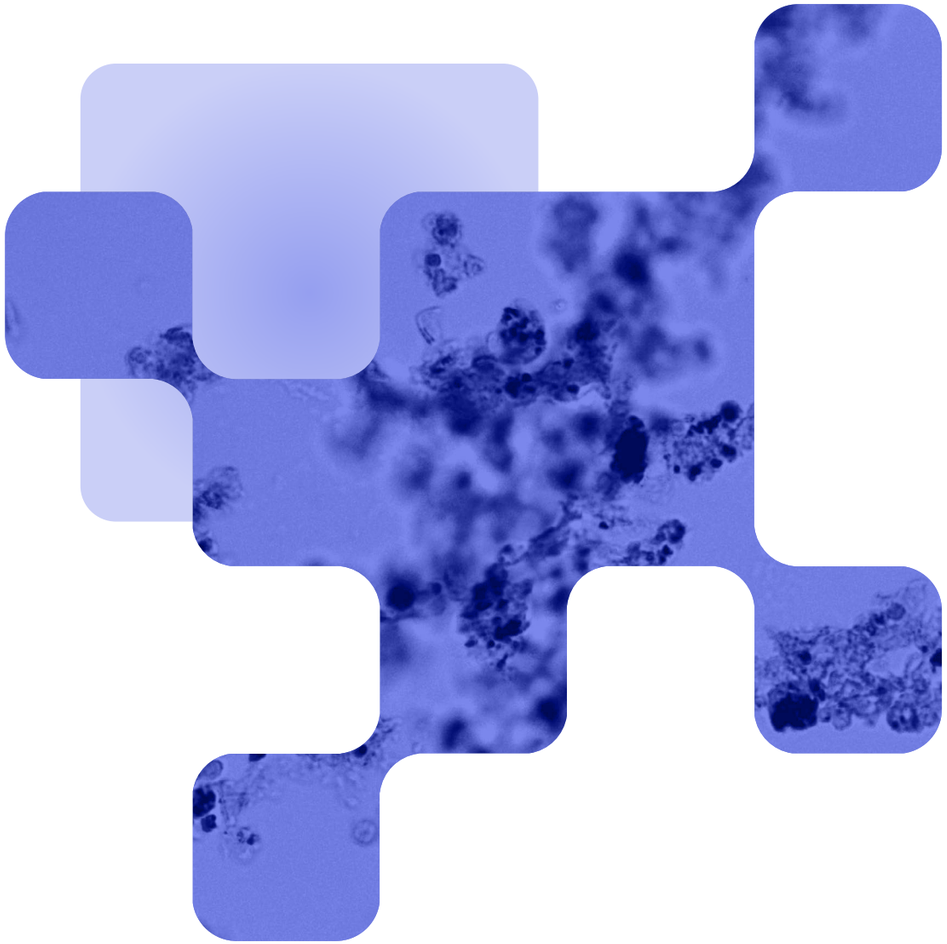
> *Parametrisierung der Extraktion: In manchen Fällen ist eine Extraktion nicht ohne zusätzliche Information möglich (z. B. mehrere Komponenten sind in einem Dokument spezifiziert, aber nur die Information zu einer sollen extrahiert werden). Der Workflow sollte dies zukünftig unterstützen.*

> *Erweiterung der Profile-Matching-Verfahren ist erforderlich, um die Zuordnung unstrukturierter Daten zu FDO-Profilen noch präziser und adaptiver zu gestalten.*

→ KI bringt Licht in „Dark Data“

Bis vor wenigen Jahren stützten sich Datenanalysen und die daraus folgenden Entscheidungen im Wesentlichen auf strukturierte Daten, obwohl der ganz überwiegende Teil existierender Informationen in unstrukturierter, schlimmstenfalls sogar in nicht maschinenlesbarer Form vorliegt. Zu aufwendig war es für Menschen, diese in eine Struktur zu überführen. Man nennt diese Daten (wie Texte, Bilder, Videos und Audio) deswegen auch „Dark Data“. Mit KI hat sich ihre Bedeutung binnen kürzester Zeit geändert. Ein direktes Resultat des Aufstiegs von Generativer KI (GenAI), Natural Language Processing (NLP) und insbesondere Large Language Models (LLMs). Erst die Fortschritte im NLP zum Beispiel haben es Maschinen ermöglicht, menschliche Sprache und somit einen Großteil dieser unstrukturierten Daten zu verstehen. LLMs wie GPT-4 oder Gemini, die auf gigantischen Mengen unstrukturierter Daten trainiert wurden, sind der entscheidende Katalysator: Sie können den Kontext und die Semantik dieser Daten erfassen, darauf basierend neue Inhalte generieren (GenAI) und dafür genutzt werden, diese Daten in eine Struktur zu überführen, wie innerhalb des AP2-Tools des FAIR Data Publishers. Somit kann KI nicht nur künftig von FDOs zu ihrer Qualitätssteigerung profitieren, sondern hilft jetzt dabei, in einer Übergangsphase in eine Data-First-Welt aus unstrukturierten Daten FAIR Digital Objects zu generieren.

USE CASES



Digitale Medizin

Der Use Case „Digitale Medizin“ konzentriert sich auf die Entwicklung von KI-Modellen für die prognostische und prädiktive Therapieerkennung bei Brustkrebs, um optimale und individuelle Therapieansätze zu ermöglichen. Dabei werden Klinikinformationssystemdaten extrahiert und in eine FDO-konforme Struktur gewandelt, um eine einfache Bereitstellung zu gewährleisten. Die Sensibilität der Daten und die technischen Anforderungen spielen hier eine besonders kritische Rolle.

¬ Der Anwendungsfall demonstriert die Vorteile von FDOs als Datenstrukturstandard in der Präzisionsmedizin und evaluiert die Möglichkeiten zur Erhöhung der verfügbaren Datenmenge und deren Nutzung, insbesondere zur Unterstützung der Entwicklung von KI-Modellen. Die Aufbereitung zu FDOs, einschließlich der Konvertierung strukturierter und unstrukturierter Daten, ist dabei entscheidend.

AKTUELLE HERAUSFORDERUNGEN IN DER DIGITALEN MEDIZIN

Die medizinische Domäne strebt im Zuge der Digitalisierung nach personalisierter Diagnostik und Therapie. Hierfür sind Daten unerlässlich, die sowohl von Menschen als auch Maschinen interpretiert und verarbeitet werden können. Aktuell bestehen jedoch erhebliche Herausforderungen bei der Aufbereitung und Bereitstellung dieser medizinischen Daten. Hinzu kommen komplexe Fragestellungen bezüglich der Datennutzung unter Wahrung des Datenschutzes und des Medizinrechts. Ein weiteres Hemmnis ist das Fehlen eines einheitlichen Standards und eines Vorgehensmodells für die Nutzung medizinischer Informationen in der digitalen Medizin und für die Entwicklung von KI-Modellen.

In der Onkologie manifestiert sich diese Fragmentierung besonders deutlich: Bildgebende Verfahren verbleiben in spezialisierten PACS-Systemen (Picture Archiving and Communication System) bei Bildgebungsverfahren mit proprietären DICOM-Erweiterungen und pathologische Befunde existieren als unstrukturierte Freitexte in heterogenen Laborinformationssystemen. Die molekulare Diagnostik fragmentiert sich über verschiedene Sequenzierungsplattformen mit inkompatiblen Ausgabeformaten, und die klinische Dokumentation verteilt sich über Krankenhausinformationssysteme unterschiedlicher Hersteller, ambulante Praxisverwaltungssysteme und spezialisierte Tumorregister mit eigenen Datenmodellen.

ZIELE UND LÖSUNGSANSÄTZE DES ANWENDUNGSFALLS

Der Anwendungsfall „Digitale Medizin“ evaluiert die praktische Anwendbarkeit von FDOs und entwickelt FDO-basierte Lösungsansätze. Dies geschieht exemplarisch am Krankheitsbild Brustkrebs (Mammakarzinom), wobei individuelle Prognostik und Prädiktion im Fokus stehen. Das Hauptziel ist die Entwicklung eines standardisierten Datenstruktur-Formats, genannt „FDO-Digitale-Medizin“, um medizinische Informationen wie Bild- und klinische Daten miteinander zu verknüpfen. Dies soll diese Daten für die Entwicklung KI-basierter Lösungsansätze in der personalisierten Medizin nutzbar machen.

Diese Fragmentierung behindert die klinische Praxis generell, insbesondere aber die Vorbereitung interdisziplinärer Tumorboards.

> »FDOs ermöglichen als selbstbeschreibende digitale Objekte eine dauerhafte Referenzierbarkeit und nahtlose Interoperabilität medizinischer Daten.«
>
> **Nils Niendorf**
> YottaSen

Gleichzeitig schafft die European Health Data Space-Regulierung (EHDS) einen verbindlichen Rechtsrahmen für die grenzüberschreitende Nutzung von Gesundheitsdaten, die die allgemeinen Vorgaben des Data Governance Act und Data Act für Gesundheitsdaten konkretisiert und damit die fragmentierte Systemlandschaft durch verbindliche, sektorspezifische Standards ersetzt. Mitgliedstaaten müssen Zugangsstellen einrichten, die als zentrale Anlaufstellen für die Sekundärnutzung elektronischer Gesundheitsdaten dienen. Gesundheitsdateninhaber werden verpflichtet, ihre Daten für genehmigte Forschungs- und Innovationszwecke bereitzustellen.

Kernstück der technischen Umsetzung der Regulierung bildet das europäische Austauschformat für elektronische Gesundheitsdaten (EHR). EHR-Systemanbieter müssen ihre Software so gestalten, dass sie Gesundheitsdaten in diesem standardisierten Format exportieren und importieren kann.

{EHDS}

Der Europäische Raum für Gesundheitsdaten ist eine Verordnung und trat am 26. März 2025 in Kraft. Sie ist ein Eckpfeiler der europäischen Gesundheitsunion und stellt den ersten gemeinsamen EU-Datenraum in einem spezifischen Bereich dar, der aus der EU-Datenstrategie hervorgeht.

DIE FDO-ARCHITEKTUR FÜR DIE UMSETZUNG DER EU-VORGABEN

Zur Erreichung dieser Ziele sind nun mehrere Lösungen entwickelt. Dazu gehören ein Datenstruktur-Format für die Digitale Medizin am Beispiel des Mammakarzinoms und die Ableitung einer Empfehlung für einen einheitlich anwendbaren Struktur-Standard für weitere medizinische Anwendungsfälle. Des Weiteren sind Beispiele möglicher FDOs, die Bild- und klinische Daten kombinieren, erarbeitet.

Die FDO-Interoperabilitätsarchitektur basiert auf drei technischen Säulen, die das Potenzial zur systematischen Umsetzung der EHDS-Vorgaben bieten. Persistente Identifikatoren (Format: 21.T11966/[UUID]) gewährleisten die dauerhafte Referenzierbarkeit medizinischer Daten über Jahrzehnte hinweg – unabhängig vom physischen Speicherort – und könnten damit die Integritätsanforderungen erfüllen. Standardisierte Metadatenprofile machen jedes FDO selbstbeschreibend: Ein onkologischer Befund kann neben den Messwerten die vom EHDS geforderten Terminologien (SNOMED CT, LOINC), präzise Messverfahrensdokumentation mit Gerätetypen und Kalibrierungsdaten sowie Qualitätsindikatoren mit Konfidenzintervallen enthalten. Maschinenlesbare Compliance-Layer können die rechtlichen Verarbeitungsgrundlagen automatisiert dokumentieren und durchsetzen – von der Erlaubnis zur Verarbeitung unter der DSGVO bis zu Anonymisierungsmetriken (k-Anonymität, l-Diversität) und angewandten Privacy-Preserving-Technologien (Differential Privacy, Secure Multi-Party Computation) sowie expliziten Nutzungsbeschränkungen gemäß den EHDS-Zweckbindungen.

Die konkrete Umsetzung für Brustkrebsdaten demonstriert die FDO-Prinzipien praktisch. Ein Tumorbefund-FDO wurde entwickelt mit medizinischen Daten wie Tumorgröße nach TNM-Klassifikation, Rezeptorstatus für ER/PR/HER2/Ki67 und histopathologischen Gewebeeigenschaften. Die Struktur ermöglicht perspektivisch die Repräsentation des Östrogenrezeptor-Status in drei Kodierungssystemen: SNOMED CT für klinische Systeme, LOINC für Laborsysteme und NCIT für Forschungsdatenbanken. Diese geplante Mehrfachkodierung würde Übersetzungsverluste eliminieren und nahtlose Integration ermöglichen. Im Rahmen des Pilotprojekts wurden drei FDO-Profile definiert und als Nanopublikationen publiziert. Das FDO-Profil Dataset Brustkrebs enthält die Parameter der „Tumorformel“ mit TNM-Klassifikation, Rezeptorstatus und Bildintegration. Das FDO-Profil Data Entry fungiert als Sub-Profil für ML-Modell-Entwicklung. Die FDO Data Entries enthalten die extrahierten Informationen mit Bildreferenzen und Rezeptorstatus für ER, PR, HER2 und Ki67. Diese Profile wurden als Nanopublikationen unter persistenten URIs publiziert und sind referenzierbar.

Die praktische Herausforderung der Interoperabilität zeigt sich in der Vielfalt medizinischer Datenquellen. Ein Brustkrebsfall aggregiert Daten aus mindestens sechs Systemen: radiologische Bildgebung (PACS), Laborwerte (LIS), pathologische Befunde, molekulare Diagnostik aus Sequenzierungsplattformen, klinische Dokumentation (KIS) und Therapiedaten. Diese Heterogenität verhindert effiziente Datennutzung und erschwert Interoperabilitätsanforderungen gerecht zu werden und praktische Hindernisse für Nutzerrechte zu minimieren.

INFORMATIONEN AUS SYNTHETISCHEN PATHOLOGISCHEN BEFUNDEN

Die im Proof-of-Concept entwickelte Konvertierungspipeline nutzte den Structure-Extraction-Service der Toolchain in der Architektur des FAIR Data Publisher für die Extraktion von Informationen aus synthetischen pathologischen PDF-Befunden und deren Überführung in FDO-Strukturen mittels des FDO Managers. Der Service extrahierte erfolgreich Rezeptorstatus-Werte (ER, PR, HER2, Ki67) und erzeugte strukturierte FDO Data Entries, die als Nanopublikationen publiziert wurden. Für die Produktivsetzung sind Natural Language Processing zur Freitextverarbeitung und die Integration standardisierter Ontologien (SNOMED CT, LOINC, ICD-O) vorgesehen, um semantisch angereicherte und international interoperable Datenstrukturen zu schaffen. Die Verwendung ausschließlich synthetischer pathologischer Befunde war eine bewusste Entscheidung zur Wahrung der DSGVO-Konformität – die besondere Sensibilität realer medizinischer Daten erforderte diese Vorsichtsmaßnahme während der Entwicklungsphase. Die Qualitätssicherung nutzte die FDO-Metadatenstruktur mit integrierten Feldern für Provenienz und Qualitätskennzeichnung. Der Proof-of-Concept etablierte die Grundfunktionalität der Datenextraktion und schuf damit die Basis für künftige Erweiterungen. Die FDO-Architektur ermöglicht die modulare Integration zusätzlicher Validierungsebenen: strukturelle Pflichtfeld-Prüfungen, semantische Ontologie-Checks und Cross-Source-Verifikationen können auf der bestehenden Infrastruktur aufbauen. Diese Skalierbarkeit gewährleistet die evolutionäre Anpassung an steigende Qualitätsanforderungen in der Produktivumgebung.

GENERIERUNG VIRTUELLER GEWEBEFÄRBUNGEN

Der Proof-of-Concept zur Generierung virtueller Gewebefärbungen demonstrierte Potenzial und Herausforderungen gleichermaßen. Die entwickelte Pipeline prozessierte Whole-Slide-Images durch Gewebesegmentierung, Zellkernerkennung, Kernklassifikation und Merkmalsaggregation. Die Nukleusklassifikation erreichte vielversprechende Average-Precision-Werte zwischen 0.886 und 0.984, während die Meta-Modelle zur Gesamtklassifikation die Performanceziele verfehlten. Besonders für HER2 und Ki67 generierten die Modelle keine verlässlichen Vorhersagen – sie klassifizierten undifferenziert Patches als „hoch“. Die Projektanalyse identifizierte als Ursache die Diskrepanz zwischen den Verteilungen der Trainings- und Testdaten sowie die begrenzte Datenbasis synthetischer Befunde. Diese transparente Limitationsdokumentation liefert wertvolle Erkenntnisse für Folgeentwicklungen. LinkAhead wurde als Datenbankinfrastruktur gewählt, da es lokale Datenhaltung sensibler Gesundheitsdaten ermöglicht und damit auf die DSGVO-Compliance einzahlt. Die Datenbank wurde erfolgreich aufgesetzt und die FDO-Integrationsarchitektur konzipiert, wodurch essenzielle Voraussetzungen für

→ Tumore mit Daten und KI präzise bekämpfen

Die Tumore sind keine homogenen Zellhaufen, sondern hochkomplexe biologische Systeme mit einzigartigen Signaturen, die sich auf molekularer, genetischer und epigenetischer Ebene manifestieren. Diese Signaturen, wie spezifische Genmutationen oder Genexpressionsmuster, machen jeden Tumor individuell und sind der Grund dafür, dass Standardtherapien oft versagen. Darüber hinaus existiert ein Tumor nie isoliert, sondern interagiert intensiv mit seinem Zellumfeld (dem Tumor-Mikromilieu). Hier manipuliert er unter anderem Immunzellen zur Immun-Evasion und induziert die Bildung neuer Blutgefäße (Angiogenese), um Wachstumssignale zu erhalten und sich zu versorgen. Diese komplexen und individuellen Wechselwirkungen bestimmen maßgeblich das Tumorwachstum, die Metastasierung und das Therapieansprechen. Angesichts dieser enormen biologischen Heterogenität ist eine pauschale Behandlung ineffektiv. Daher ist die Präzisionstherapie nötig, die exakt auf diese individuelle Charakteristik ausgerichtet ist. Hierzu sind KI-Modelle dringend nötig, um die Muster des Therapieansprechens erkennen zu können.

USE CASES_DIGITALE MEDIZIN

INTERVIEW MIT NILS NIENDORF, YOTTASEN

>>Warum ist die Standardisierung medizinischer Daten so entscheidend für die Weiterentwicklung der digitalen Medizin?
<<Weil die derzeit stark fragmentierte Datenlandschaft eine effektive Nutzung für personalisierte Diagnostik und Therapie massiv erschwert und viele manuelle Schritte erfordert.

>>Welche waren die größten praktischen Hürden, um FDOs im medizinischen Bereich zu etablieren?
<<Die größte Hürde war die Fragmentierung der medizinischen Datenquellen und die Notwendigkeit, eine DSGVO-konforme Anonymisierungspipeline zu entwickeln, die im Projekt nur konzeptionell umgesetzt werden konnte.

>>Was sind die konkreten Hauptvorteile, die FDOs für die Zukunft der digitalen Medizin versprechen?
<<FDOs ermöglichen als selbstbeschreibende digitale Objekte eine dauerhafte Referenzierbarkeit und nahtlose Interoperabilität medizinischer Daten, was die Grundlage für KI-Anwendungen und regulatorische Compliance schafft.

die technische Umsetzung in einem möglichen Folgeprojekt geschaffen wurden.

ERGEBNISSE, LIMITATIONEN UND PERSPEKTIVEN

Das Pilotprojekt erreichte trotz zeitlicher Begrenzungen substantielle Teilerfolge: Drei FDO-Profile für Brustkrebsdaten wurden definiert und als Nanopublikationen publiziert. Die Toolchain extrahierte erfolgreich strukturierte Informationen aus synthetischen pathologischen Befunden. Eine konzeptionelle Anonymisierungspipeline wurde entworfen, die auf etablierte Privacy-Preserving-Techniken wie die Methoden nach El Emam und Kniola referenziert, deren praktische Validierung jedoch aussteht. Die Nukleusklassifikation erreichte vielversprechende Average Precision-Werte bis 0.984, während die Meta-Modelle für HER2 und Ki67 die Performanceziele verfehlten. Die nicht realisierten Komponenten markieren verbleibende Entwicklungsfelder: Die Integration mit realen klinischen Systemen und die Validierung mit echten Patientendaten erfolgten bewusst nicht, um Datenschutzrisiken während der Entwicklungsphase auszuschließen. Die vollständige LinkAhead-FDO-Integration und die Optimierung der Meta-Modelle bleiben künftigen Projektphasen vorbehalten.

FAZIT UND EMPFEHLUNGEN

PROOF OF CONCEPT

- **Im Rahmen eines Proof of Concept zur Generierung virtueller Gewebefärbungen wurde das Potenzial der FDO-Nutzung für onkologische Daten aufgezeigt.**

- **Drei spezifische FDO-Profile für Brustkrebsdaten, inklusive eines Sub-Profils für die ML-Modellentwicklung, wurden definiert und als Nanopublikationen unter persistenten URIs publiziert.**

- **Eine Konvertierungspipeline auf Basis der entwickelten Toolchain aus der Architektur des FAIR Data Publisher, wurde erfolgreich demonstriert, die strukturierte Informationen aus synthetischen pathologischen PDF-Befunden extrahierte und in FDO-Strukturen überführte.**

- **Konzeptionell wurde eine DSGVO-konforme Anonymisierungspipeline entwickelt, um den Schutz sensibler Patientendaten zu gewährleisten, während LinkAhead als FDO-Datenbankinfrastruktur für lokale Datenhaltung aufgesetzt wurde.**

- **Die technische FDO-Architektur für die Präzisionsmedizin wurde so gestaltet, dass sie die Auffindbarkeit, Zugänglichkeit, Interoperabilität und Wiederverwendbarkeit von medizinischen Daten verbessert und regulatorische Anforderungen des EU-Digitalrechtsrahmens adressiert.**

- **Diese Arbeiten bilden eine Grundlage für die Entwicklung KI-basierter Lösungen auf Basis von FDOs in der personalisierten Medizin, trotz der Limitationen durch die ausschließliche Verwendung synthetischer Daten und noch nicht optimierter Meta-Modelle.**

→ Differential Privacy

Differential Privacy ist ein mathematisches Verfahren, um die Analyse von Daten zu ermöglichen, ohne die Privatsphäre einzelner Personen zu verletzen. Es funktioniert, indem statistische Ergebnisse durch gezieltes Rauschen verfälscht werden, sodass kein Rückschluss auf individuelle Daten möglich ist. Dies gewährleistet eine datenschutzkonforme Forschung und Entwicklung in der Onkologie mit sensiblen Tumor- und Patientendaten und von KI-Modellen, beispielsweise zur Krebsdiagnose oder Medikamentenwirksamkeit.

MEHRWERTE

+ **Deutliche Verbesserung der Dateninteroperabilität:** FDOs überwinden die systemische Fragmentierung der medizinischen Datenlandschaft, indem sie eine technisch standardisierte Infrastruktur nach FAIR-Prinzipien bieten, die eine effektive Nutzung heterogener Daten ermöglicht.

+ **Stärkung der regulatorischen Compliance:** Die entwickelte FDO-Architektur ist darauf ausgelegt, die Anforderungen des EU-Digitalrechtsrahmens, wie den European Health Data Space, den Data Act, den AI Act und die DSGVO, durch maschinenlesbare Compliance-Informationen und persistente Identifikatoren zu unterstützen.

+ **Grundlage für personalisierte Medizin und KI-Anwendungen:** Durch die standardisierte Verknüpfung und Strukturierung heterogener medizinischer Daten, wie beispielsweise detaillierte Tumorbefunde und histopathologische Bilder, wird eine essenzielle Basis für die Entwicklung KI-basierter Lösungen in der personalisierten Diagnostik und Therapie geschaffen.

+ **Granulare und dauerhafte Referenzierbarkeit von Daten:** Persistente Identifikatoren gewährleisten die lückenlose Nachverfolgbarkeit und die dauerhafte Referenzierbarkeit einzelner medizinischer Datenelemente über verschiedene Systeme und lange Zeiträume hinweg.

USE CASES_DIGITALE MEDIZIN

+ **Automatisierte Datenextraktion und -strukturierung:** Eine proof-of-concept-Konvertierungslösung wurde erfolgreich demonstriert, die unstrukturierte pathologische Befunde automatisiert in einheitliche, FDO-konforme Formate überführen kann.

+ **Erhöhung der Datenqualität und des Vertrauens:** Die FDO-Struktur ermöglicht die Integration umfassender Metadaten zur Datenherkunft, Qualitätskennzeichnung und zum Anonymisierungsgrad, wodurch die Vertrauenswürdigkeit und die reproduzierbare Verwendbarkeit medizinischer Daten gesteigert werden.

NOTWENDIGE WEITERE SCHRITTE

> *Es bedarf einer Validierung der entwickelten FDO-Lösungen mit realen Patientendaten, um die Praxistauglichkeit unter Berücksichtigung des Datenschutzes zu belegen.*

> *Eine DSGVO-konforme Anonymisierungspipeline sollte praktisch implementiert und validiert werden, um den Schutz sensibler Patientendaten in realen Anwendungen zu gewährleisten.*

> *Die vollständige technische Integration der FDO-Profile mit der LinkAhead-Datenbankinfrastruktur muss über die konzeptionelle Ebene hinaus realisiert werden, um eine robuste lokale Datenhaltung zu ermöglichen.*

> *Es ist notwendig, die FDO-Architektur umfassend in bestehende reale klinische Informationssysteme zu integrieren, um deren Nutzbarkeit und Akzeptanz in der klinischen Praxis zu demonstrieren.*

> *Der konzipierte dreistufige Qualitätssicherungsprozess für medizinische Daten, einschließlich struktureller, semantischer und Cross-Source-Validierung, muss vollständig implementiert und angewendet werden.*

Industrielle Fertigung

Beim Use Case „Industrielle Fertigung“ liegt der Fokus auf der Entwicklung eines digitalen Zwillings für die Batteriefertigung, wobei Informationen durch Tools automatisch extrahiert und in strukturierte Daten transformiert werden.

¬ Die europäische Batterieindustrie steht vor einer doppelten Transformationsherausforderung. Die Batterieverordnung (EU) 2023/1542 verlangt ab Februar 2027, dass alle Batterien über 2 kWh einen digitalen Produktpass mit lückenloser Dokumentation von Herkunft, Zusammensetzung und Leistungsdaten vorweisen müssen.

Diese Anforderungen treffen auf eine hoch fragmentierte Datenlandschaft: Produktdatenblätter als PDFs ohne standardisierte Struktur, bei denen identische Parameter je nach Hersteller an unterschiedlichen Positionen dokumentiert werden; proprietäre Layouts und Excel-Tabellen mit fehlerbehafteter manueller Datenextraktion; inkonsistente Parameterdefinitionen, bei denen Werte von anderen Größen abhängen. Das erschwert die automatisierte Extraktion und Harmonisierung der Daten erheblich.

Die gängige Praxis beinhaltet daher oft die manuelle Übertragung von Daten aus diesen PDFs in strukturierte Datenbanken oder Tabellen.

> »Durch FAIR Digital Objects können wir die Grundlage für Digitale Produktpässe und optimierte Fertigungsprozesse legen.«
>
> **Henrik Tom Wörden**
> Indiscale

Dieser Prozess ist nicht nur zeitaufwendig und kostenintensiv, sondern auch fehleranfällig und stark limitiert in seiner Skalierbarkeit, was die Effizienz und die Qualität der Produktentwicklung und Fertigungsoptimierung beeinträchtigt. Insbesondere bei komplexen technischen Produkten ist deswegen eine nahtlose Digitalisierung entlang der gesamten Lieferkette unerlässlich. Der Anwendungsfall „Industrielle Fertigung“ des Projekts FAIR Data Publisher widmet sich dieser Herausforderung. Er zeigt, wie FDOs die Bereitstellung von Datenblättern von Ausgangskomponenten wie beispielsweise von Operationsverstärkern und deren elektrische und mechanische Eigenschaften wie Leerlaufverstärkung, Ausgangsspannungshub, Ausgangswiderstand in einem industriellen Datenraum vereinheitlicht werden können.

DAS PROJEKT „BATCAT“ ALS DEMONSTRATOR

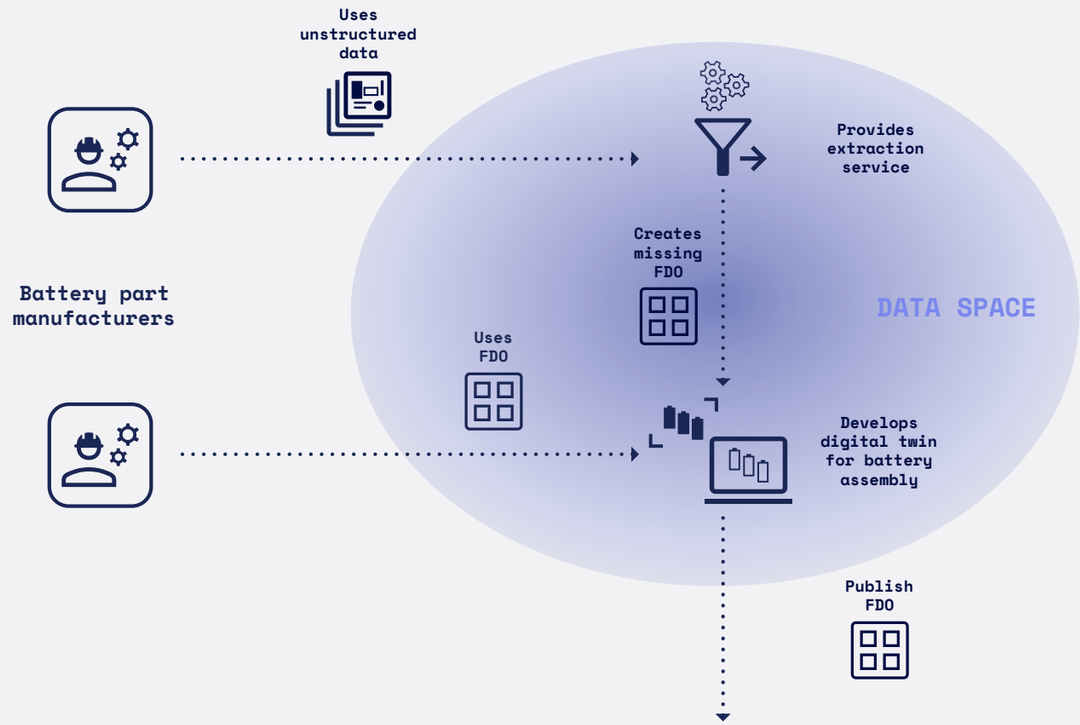
Für den Anwendungsfall dient das Projekt Battery Cell Assembly Twin („BatCAT“) als Beispiel für einen Datenraum mit Akteuren aus der Industrie und öffentlichen Forschung. BatCAT ist ein von Horizon Europe finanziertes Forschungs- und Innovationsprojekt, das einen Digitalen Zwilling zur Optimierung der industriellen Batteriezellproduktion entwickelt. Hierbei kommen statistische Verfahren, Multi-Criteria-Optimization und KI zum Einsatz. Der „BatCAT“-Datenraum basiert auf den Eclipse Dataspace Components (EDC) und verwendet Handle-basierte FDOs als Registrierungsinfrastruktur.

Der Fokus liegt darauf, zu demonstrieren, wie eine FDO basierter Konvertierungsservice in einem EDC-Datenraum bereitgestellt werden kann. Er soll die von Optimierungsalgorithmen geforderte Datenstruktur liefern und als FDO im Datenraum verfügbar machen. Gleichzeitig gilt es, die über FDOs strukturierten Daten zu Operationsverstärkern als AAS-Submodule (Asset Administration Shell) exportieren zu können, um die Erstellung von Digitalen Produktpässen zu vereinfachen. Die IndiScale GmbH ist für die Entwicklung und den Betrieb des „BatCAT“-Datenraums verantwortlich und stellt einen niedrigschwelligen Zugang als Datenanbieter zur Verfügung.

Der Prozessfluss (siehe Abbildung) verdeutlicht die Wertschöpfungskette: Hersteller von Batteriekomponenten nutzen unstrukturierte Daten, die über einen Extraktionsdienst in standardisierte FDOs überführt werden. Fehlende Informationen werden so ergänzt und in den Datenraum eingebracht, wo sie veröffentlicht und weitergenutzt werden können. Die so verbesserte Datenbasis kommt direkt der Erstellung eines digitalen Zwillings für die Batteriemontage zu Gute. So wird eine parametrische Planung ermöglicht, die Effizienz und Flexibilität stärkt.

DIE ROLLE DES FDO-BASIERTEN KONVERTIERUNGSSERVICES

Die Technologie begegnet der Herausforderung der unstrukturierten Daten mit einem FDO-basierten Konvertierungsservice, der auf der entwickelten Software in der FDO-Architektur aufbaut.



•FDO-PROZESS INDUSTRIELLE FERTIGUNG•

Die Infografik veranschaulicht den Anwendungsfall „BatCat“. Während manche Hersteller von Batteriekomponenten Informationen zu Bauteilen (zukünftig) direkt als strukturierte FDOs zur Verfügung stellen, können fehlende Informationen ergänzt werden, indem die unstrukturierten Daten anderer Hersteller aufbereitet und zu FDOs konvertiert werden.

Diese liefert automatisierte Methoden und einen Demonstrator, um aus verschiedenen unstrukturierten Daten strukturierte FDOs zu erzeugen. Dazu gehören Analyse, Aufbereitung und Bereitstellung von Daten aus Dokumenten wie PDFs als FDOs, einschließlich Metadaten.

Der Verarbeitungsprozess umfasst die Überführung der Daten in ein generisches Textformat, die Speicherung in gängigen strukturierten Formaten wie JSON oder CSV und die abschließende Umwandlung in die finale Struktur für das FDO. Hierbei kommen moderne Technologien wie Optical Character Recognition (OCR) sowie KI-basierte Verfahren und Large Language Models (LLMs) zum Einsatz, um relevante Inhalte zu extrahieren und als strukturierte Daten, wie etwa JSON-LD, zu speichern.

FDOS ALS ERMÖGLICHER VON INTEROPERABILITÄT UND MACHINE ACTIONABILITY

Die zentrale Rolle von FDOs in diesem Kontext ist die Ermöglichung von Machine Actionability und Interoperabilität. FDOs stellen dabei mit Regeln und Minimalstandards technologieunabhängig die Beschaffenheit von digitalen Objekten sicher. Und dies unabhängig von der zugrundeliegenden Implementierung wie beispielsweise Linked Data oder EDC.

Der Konvertierungsservice selbst wird als FDO-Operation modelliert, die eine Persistente ID (PID) und ein Profil besitzt, das sie als Operation auszeichnet. Dieses Operations-FDO definiert klar die erwarteten Input-Parameter (z. B. ein PDF-FDO

mit Nutzungsbedingungen) und die Output-Parameter (z. B. ein strukturiertes Datensatz-FDO, das einem spezifischen Profil folgt).

ANBINDUNG AN DIE ASSET ADMINISTRATION SHELL

Die Asset Administration Shell als Beitrag zur Industrie 4.0-Standardisierung zeigt praktische Einschränkungen bei der Komponentenauswahl in frühen Entwicklungsphasen. Für die notwendige Auswahl von Komponenten existieren keine dedizierten Templates. Das FAIR Data Publisher Projekt hat das generische Technical Data Submodel Template angepasst.

Die aus den PDF-Dokumenten extrahierten und strukturierten Daten lassen sich am Ende in ein AAS-Submodul nun transferieren und somit für die Verwendung im Digitalen Produktpass (DPP) vorbereiten. Dies schließt den Kreis des Anwendungsfalls von der Datenaufbereitung bis zur nutzbaren, standardisierten Information im Produktlebenszyklus.

Es ist wichtig hervorzuheben, dass im vorliegenden FAIR Data Publisher Projekt der Konvertierungsservice oder die extrahierten, strukturierten Daten als FDO abgebildet werden, nicht der Digitale Zwilling der Batterieherstellung selbst. Der Digitale Zwilling dient als übergeordneter Kontext und als Konsument dieser strukturierten Daten, um den Bedarf an automatisierten Datenprozessen zu verdeutlichen. Zukünftig könnte der Digitale Zwilling selbst als FDO modelliert werden, dies ist jedoch nicht der Fokus dieses spezifischen Projekts.

STRUKTUR UND TECHNISCHE UMSETZUNG DES DEMONSTRATORS

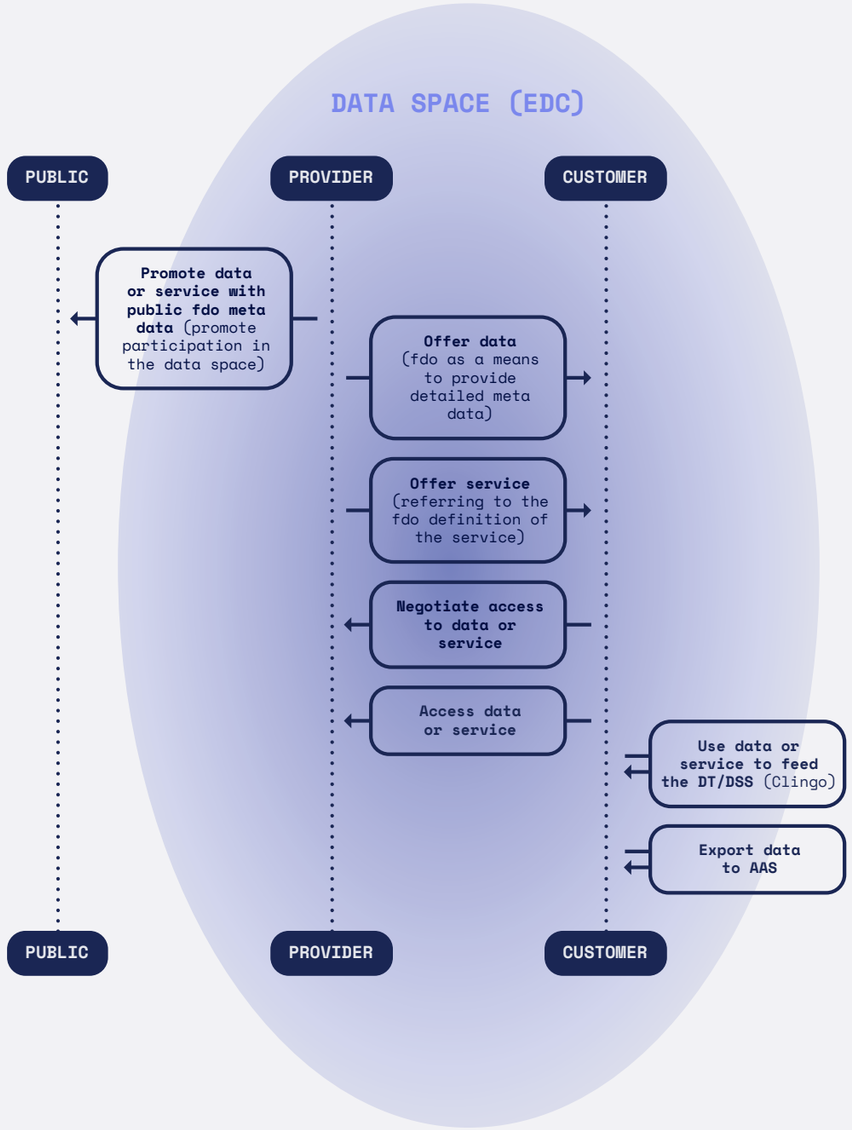
Der Demonstrator für den Anwendungsfall „Industrielle Fertigung“ zeigt prototypisch, wie ein FDO-basierter Service in einen EDC-Datenraum integriert und dort angeboten werden kann. Der BatCAT-Datenraum dient dabei als Beispiel-Implementierung. Im Demonstrator ermöglicht der BatCAT-Datenraum über den Eclipse Dataspace Connector (EDC) den standardisierten Austausch zwischen Datenraum-Teilnehmern, wobei Service-Provider Extraktionsdienste bereitstellen und andere Teilnehmer diese über DOIP-Endpoints nutzen können. Die Architektur des Demonstrators umfasst LinkAhead-Instanzen, die über EDC für den Datenaustausch verbunden sind.

Beim Ausführen des Service wird das Service-Asset angefordert, und EDC-Komponenten führen eine Vertragsverhandlung durch. Als Ergebnis erhält der Dienstnutzerknoten einen Endpunkt und einen Token, um den Dienst auszuführen. Dieser Ansatz simuliert, wie FDOs einen maschinenlesbaren Mechanismus zur Bereitstellung von Services in und über Datenräume hinaus bieten können. Die strategische Bedeutung liegt darin, dass der FDO-Ansatz als Blaupause oder sogar als Teil zukünftiger Datenraumprotokolle für Services dienen könnte.

Die Extraktionspipeline demonstriert Ansätze zur regulatorischen Compliance-Unterstützung in vier Phasen: Datenextraktion durch

→ Parametrische Planung in der Elektroindustrie

Die parametrische Planung hat sich in der Elektroindustrie als entscheidende Methode etabliert, um die wachsende Komplexität und Variantenvielfalt von Komponenten – von Widerständen bis hin zu komplexen Mikrocontrollern – effizient zu beherrschen. Der Schlüssel zur Effizienz liegt in der direkten Nutzung von Produktblättern und digitalen Herstellerkatalogen, um Fertigungsmodelle automatisiert zu generieren. Das führt zu einer massiven Beschleunigung des Designprozesses. Ingenieure können schnell zwischen Komponenten verschiedener Hersteller oder unterschiedlicher Leistungsklassen wechseln. Die aus den Modellen abgeleiteten Fertigungsdaten erleichtern beispielsweise die Programmierung der Bestückungsautomaten (Pick-and-Place-Maschinen) und die Einhaltung von Regulierung.



•BUSINESS CANVAS•

Die Infografik illustriert, wie FAIR Digital Objects (FDO) in einem industriellen Data Space auf Basis des Eclipse Dataspace Connector Datenaustausch unterstützen. Daten oder Services werden über öffentliche FDO-Metadaten sichtbar gemacht, detailliert beschrieben und als Angebot in den Data Space eingebracht. Auf dieser Grundlage lassen sich Zugänge aushandeln und schließlich Daten oder Services nutzen. Diese können in eine Asset Administration Shell (AAS) exportiert oder zur Versorgung von Digitalen Zwillingen und Entscheidungsunterstützungssystemen (DSS) verwendet werden.

KI-basierten Service, Standardisierung in FDO-konforme Strukturen, Validierung mit Human-in-the-Loop-Prozessen (direkter Bezug zu Artikel 14 des AI Act über menschliche Aufsicht), und Transformation zu AAS-Submodellen für digitale Produktpässe. In einer nächsten Entwicklungsstufe könnte die Extraktionsgenauigkeit durch spezialisierte Modelle erhöht werden.

KOMMERZIALISIERUNGSSTRATEGIEN UND WERTSCHÖPFUNG DURCH FDOS

Die Integration in EDC-basierte Datenräume (Eclipse Dataspace Components) ist für FDOs prädestiniert. Der entwickelte Service kann als Asset innerhalb eines Datenraums angeboten werden, wobei das Dataspace Protocol für die automatisierte Aushandlung von Verträgen und die Vergabe von Zugangsrechten (z. B. über Access Tokens nach Zahlung) genutzt wird. Dies eröffnet neue Möglichkeiten für die Monetarisierung von Daten und Services.

FAZIT UND EMPFEHLUNGEN

PROOF OF CONCEPT

- **Es wurde ein Service etabliert, der die automatisierte Extraktion und Strukturierung von Spezifikationsdaten aus unstrukturierten PDF-Datenblättern elektronischer Komponenten, wie Operationsverstärkern, ermöglicht. Diese extrahierten Informationen wurden anschließend in FDO-konforme Strukturen überführt.**

- **Zusätzlich wurde die Transformation dieser FDOs in AAS-Submodelle demonstriert, um die Integration in digitale Produktpässe zu erleichtern.**

- **Für die Qualitätssicherung wurden Human-in-the-Loop-Prozesse implementiert, um die Datenqualität der extrahierten Informationen zu validieren. Dies adressierte Anforderungen wie Artikel 14 des AI Act bezüglich menschlicher Aufsicht.**

- **Darüber hinaus wurde die Nutzung von FDO-Daten mit dem Clingo ASP-Framework für eine automatisierte Optimierung demonstriert.**

- **Die entwickelten FDOs bewiesen ihre Fähigkeit, als Container für EDC-Assets zu dienen und Daten innerhalb eines Datenraums zu referenzieren und zu transportieren.**

MEHRWERTE

+ **Erfüllung regulatorischer Anforderungen:** FDOs unterstützen die Einhaltung neuer EU-Regularien wie die Batterieverordnung (digitaler Produktpass ab 2027) und den AI Act (menschliche Aufsicht in sicherheitskritischen Anwendungen) durch standardisierte Dokumentation und integrierte Compliance-Mechanismen.

+ **Automatisierte Datenextraktion und -strukturierung:** Der entwickelte Service ermöglicht die effiziente automatisierte Extraktion von Spezifikationsdaten aus unstrukturierten PDF-Datenblättern und deren Überführung in FDO-konforme Strukturen, wodurch manuelle, fehleranfällige Prozesse ersetzt werden.

+ **Verbesserte Interoperabilität und Machine Actionability:** FDOs fungieren als technologieunabhängiges Regelwerk für digitale Objekte und Operationen, das den maschinenlesbaren Austausch und die automatisierte Verarbeitung von Daten und Diensten über heterogene Systeme und Industriestandards (z. B. AAS, EDC) hinweg ermöglicht.

+ **Optimierung von Design- und Fertigungsprozessen:** Durch die Bereitstellung strukturierter und interoperabler Daten unterstützen FDOs Decision Support Systeme bei der parametrischen Optimierung von Produktdesign (z. B. Komponentenauswahl) und Fertigungsprozessen (z. B. Reduzierung von Ausschuss).

+ **Nahtlose Integration in digitale Produktpässe (DPP) und Industrie 4.0-Infrastruktur:** FDOs können als AAS-Submodelle exportiert werden, was eine Kompatibilität mit der bestehenden deutschen Industrie 4.0-Infrastruktur sicherstellt, wodurch Vendor-Lock-in verhindert wird.

NOTWENDIGE WEITERE SCHRITTE

> *Vollständige Automatisierung der Datenextraktion und -validierung: Die automatische Extraktion erreichte noch keine ausreichende Genauigkeit für den vollautomatischen Einsatz ohne manuelle Nachkontrolle, was die fortgesetzte Notwendigkeit des Human-in-the-Loop-Ansatzes bestätigt und die Implementierung einer robusteren Validierungspipeline erfordert.*

INTERVIEW MIT HENRIK TOM WÖRDEN, INDISCALE

>>Warum ist dieses Projekt so wichtig für die industrielle Fertigung, gerade im Bereich Batterien?
<<Wir wollen die Batteriezellenproduktion digitalisieren, um Prozesse zu optimieren und die neuen EU-Anforderungen, wie den digitalen Produktpass ab 2027, zu erfüllen.

>>Was war bislang das Hemmnis?
<<Die größte Hürde war die manuelle und fehleranfällige Extraktion von Komponentendaten aus unstrukturierten PDF-Datenblättern, da jeder Hersteller sein eigenes Layout hat und diese Daten nicht maschinenlesbar sind.

>>Und was bringt das den Unternehmen am Ende des Tages konkret für Mehrwerte?
<<Durch FAIR Digital Objects können wir Daten automatisiert strukturieren, Dienste maschinenlesbar anbieten und so die Grundlage für Digitale Produktpässe und optimierte Fertigungsprozesse legen.

> *Entwicklung eines nachhaltigen und differenzierten Geschäftsmodells: Das aktuelle Geschäftsmodell leidet unter dem Fehlen „Unfairer Vorteile“ und der transienten Natur des Bedarfs, sobald Komponentenhersteller selbst FDO-konforme Daten bereitstellen, was eine stärkere Differenzierung, z. B. durch spezialisierte Modelle, erfordert.*

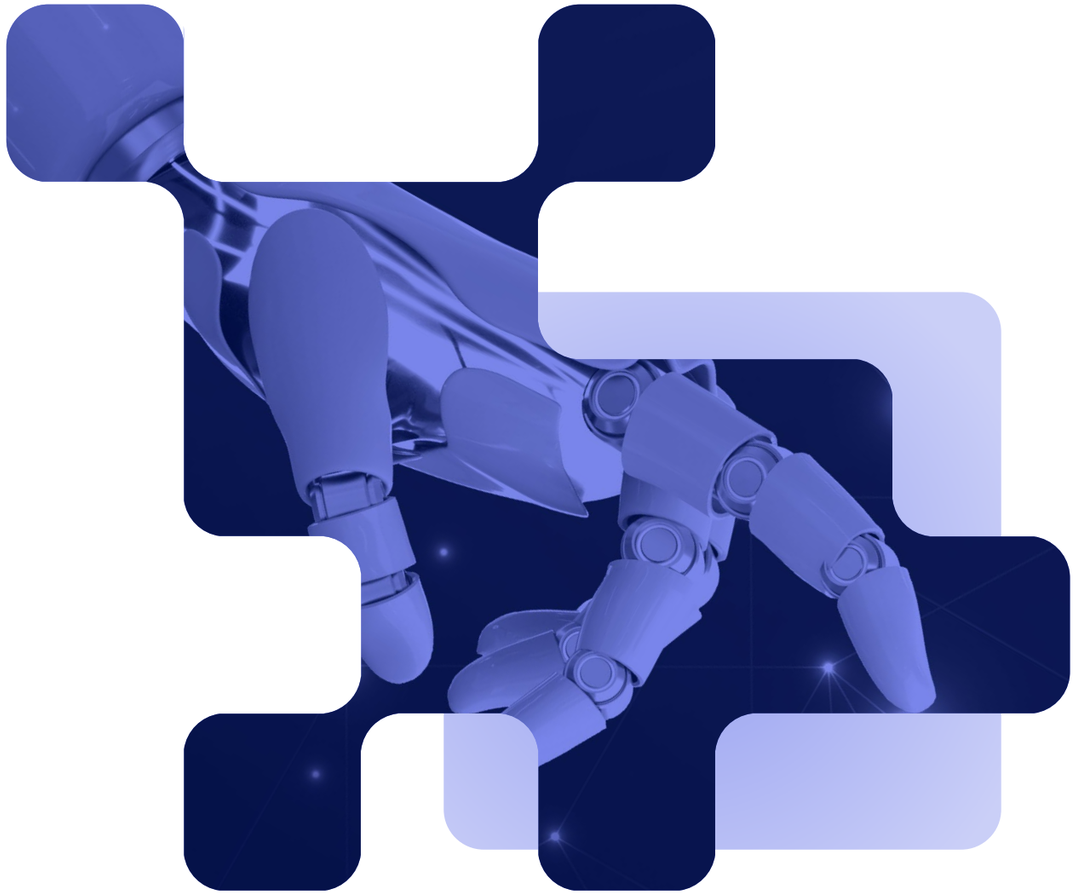
> *Umfassende Validierung für regulatorische Compliance und Produktivsysteme: Der Demonstrator ist ein Proof of Concept für Datenextraktion und -strukturierung und daher noch nicht für Zwecke der regulatorischen Compliance oder den Produktiveinsatz geeignet.*

> *Lösung der Datenfragmentierung auf Seiten der Komponentenhersteller: Die fundamentale Datenfragmentierung bleibt ein Hindernis, solange Komponentenhersteller nicht selbst standardisierte FDO-Daten bereitstellen, was eine breitere Akzeptanz und Standardisierung in der Lieferkette erfordert.*

> *Skalierung und Integration in umfassende Datenökosysteme: Skalierungsprobleme bei der Verarbeitung großer Datenmengen und die vollständige Integration in umfassende Datenökosysteme mit automatisierten Update- und Delete-Funktionen der FDO-Infrastruktur sind noch nicht vollständig gelöst.*

Verweise Industrielle Fertigung:
↗ Digikey Produktbeispiel (QAS254K600D150)
↗ Datenblatt Beispiel (QAS_Type.pdf)
↗ Regulatorische Erklärung Beispiel (CDE-Regulatory-Declaration.pdf)
↗ EC CDD Klasse für Operationsverstärker

Robotik



Der Use Case „Robotik“ befasst sich mit der Verwaltung und Bereitstellung von mehreren Terabyte umfassenden Legacy-Daten aus Forschungsprojekten im Agrarsektor, die semantisch mittels RDF-Logik atomisiert und interoperabel aufbereitet werden sollen. Unter der Leitung des Deutschen Forschungszentrums für Künstliche Intelligenz GmbH (DFKI) zielt dieser Anwendungsfall darauf ab, die Vorteile von FDOs als Datenstrukturstandard in realen Szenarien zu demonstrieren.

¬ Entscheidend für die Unterstützung KI-basierter Lösungen ist die Verbesserung syntaktischer und semantischer Interoperabilität durch granulare Adressierbarkeit von Datenelementen sowie kontrollierte Semantik über alle Ebenen des Datenmodells.

Die Robotik-Forschung ist mit einem erheblichen Bedarf an der Zusammenführung von Datensätzen aus vielfältigen Teildomänen konfrontiert, beispielsweise für die Entwicklung von „Robotics Foundation Models“. Diese Zusammenführung wird durch einen Mangel an Beschreibungsstandards für Datensätze erschwert. Eine weitere Herausforderung besteht in der Komplexität einzelner Datensätze, etwa aus Feldexperimenten mit verteilten Systemen, bei denen die kombinatorische Vielfalt möglicher Beziehungen zwischen Datenkomponenten und diversen Datenformaten Kausalanalysen erschwert. Eine standardisierte, maschinen-interpretierbare Spezifizierung von Komponenten kann diese Reibungsverluste mindern und umfassendere Analysen ermöglichen.

DAS ROBIVAL-PROJEKT ALS DATENGRUNDLAGE

Der Robotik-Anwendungsfall von FAIR Data Publisher basiert auf Daten des vorausgegangenen RoBivaL-Projekts, das sich mit dem Fahrverhalten und der Bodeninteraktion von vier mobilen Robotern in landwirtschaftlichen Anwendungen befasste. Der dabei generierte Datenkorpus umfasst hochaufgelöste Roboter-Positionen (RTK-GPS), Bewegungscharakteristiken (IMU), Bodenwiderstand (Digitales Penetrometer), 3D-Modelle der Bodenoberfläche (Laserscanner), Zugkraft (Kraftsensor), manuelle Messungen und Videodokumentationen. Ursprünglich auf Zenodo veröffentlicht, wies diese Datenbasis Einschränkungen hinsichtlich Interoperabilität und Nachnutzbarkeit auf, da sie als monolithisch strukturiert, mit nicht kodifizierter Semantik und ohne Vorkehrungen für zukünftige Erweiterungen vorlag.

Die wesentlichen Ziele im Kontext des RoBivaL-Datenkorpus sind die Transformation der Daten in ein Netzwerk aus Nanopublikation-basierten FDOs und die Bereitstellung von Software für den bidirektionalen Datentransfer zwischen einer domänenspezifischen Datenbank und der FDO-Infrastruktur. Dies beinhaltet die Aufteilung des Korpus in einzelne, granulare Datenelemente; die Etablierung kontrollierter Semantik über alle Ebenen des Datenmodells, um Mehrdeutigkeiten zu eliminieren und die Effizienz der Nachnutzung zu steigern; sowie die Entwicklung von Konzepten und Tools für eine fortlaufende Evolution des Datenkorpus in einem global verteilten Datenmanagement-Ökosystem.

Der Robotik-Anwendungsfall entwickelt ein mehrschichtiges Datenmodell für den RoBivaL-Korpus, das aus der Experimental Research Ontology (ERO), der RoBivaL Specification, den RoBivaL Payload Data und den FDO-Profilen für RoBivaL FDO-Records besteht. Jede Schicht repräsentiert unterschiedliche Abstraktions- und Allgemeinheitsebenen und adressiert spezifische Bedürfnisse und Nutzungsszenarien innerhalb der Forschungsgemeinschaft.

EXPERIMENTAL RESEARCH ONTOLOGY (ERO)

Die Experimental Research Ontology (ERO) bildet die abstrakte semantische Grundlage für den RoBivaL-Datenkorpus und ist ein Werkzeug für experimentelle Wissenschaftlerinnen und Wissenschaftler aller Disziplinen, um FAIR-konform zu spezifizieren, welche Datenelemente in welchen Teilen ihrer Arbeit erzeugt werden. Sie besteht aus 29 RDF/OWL-Entitäten, die ausgewählte Kernaspekte der Datenerzeugung in der experimentellen Forschung modellieren, wie die Charakterisierung von Experimenttypen durch Methoden und Parametersammlungen. ERO ist an etablierte Oberontologien wie BFO, IAO, OBI und SOSA angelehnt und in ERO Core (essenziell), ERO Util (optional, entwicklungsfähig) und ERO Alias (menschenlesbare Bezeichnungen) unterteilt.

Die RoBivaL Specification dient als Best-Practice-Beispiel für die Anwendung der ERO in einem Robotik-Kontext und für die FAIR-konforme Beschreibung von Forschungsergebnissen. Sie liefert Nachnutzerinnen und Nachnutzern der RoBivaL Payload Data den notwendigen Kontext für ein tiefgehendes Verständnis und nahtlose Interoperabilität. Die RoBivaL Payload Data selbst umfasst die Werte von Experimentparametern und hochauflösende Sensormessungen aus den Robotik-Experimenten, die für diverse Nachnutzungsszenarien, wie die Verbesserung von Fortbewegungssystemen oder als Trainingsdaten für KI-Algorithmen, von Bedeutung sind. Die Modellierung der hochauflösenden Sensordaten liegt jedoch aufgrund der Projektlaufzeit und Komplexität außerhalb des aktuellen Umfangs von FAIR Data Publisher.

FDO-PROFILE FÜR ROBIVAL FDO-RECORDS

Alle Datenobjekte der ersten drei Schichten des Modells werden als FDO-Records formatiert, deren Format explizit in FDO-Profilen spezifiziert

wird, um den FDO-Spezifikationen zu entsprechen. Diese Profile sind als RDF/SHACL-Graphen implementiert und kompatibel mit Nanopublikationen über die Nanodash-Plattform. Eine zentrale Herausforderung war die rekursive Definition von FDOs, bei der jedes FDO-Record auf ein FDO-Profil verweisen muss, das Profil selbst aber auch wieder als FDO-Record repräsentiert ist. Zur Beendigung dieser Rekursion wurde ein fundamentales, selbst-validierendes FDO-Profil implementiert, welches als kritische Basiskomponente die allgemeine FDO-Erstellung ermöglicht.

SOFTWARE UND AUTOMATISIERUNG

Der Robotik-Anwendungsfall umfasst die Entwicklung einer Software zur Validierung von FDO-Records gegen FDO-Profile und zur Veröffentlichung dieser Records als Nanopublikationen. Diese Software basiert auf der in der FDO-Architektur entwickelten Python-Bibliothek nanopub-py-fdo und nutzt Semantic Web Standards wie RDF-Graphen für Records und SHACL-Graphen für Profile. Da die Nanopublikations-Infrastruktur keine automatische Validierung der Konformität mit FDO-Profilen bietet, muss diese lokal auf der Client-Seite implementiert werden. Zudem ermöglicht die Python-API eine automatisierte Publikation der FDO-Records, um die Skalierung auf hunderte oder tausende von Datensätzen zu gewährleisten.

DATENTRANSFER UND DATENBANKINTEGRATION

Zur Demonstration der Interoperabilität wurden bidirektionale Datentransfer-Tools zwischen der FDO-Infrastruktur und einer domänenspezifischen SQLite-Datenbank entwickelt. Diese Tools ermöglichen das Schreiben und Lesen von Daten an beiden Enden der Transfer-Pipeline durch generische Clients und unterstützen die Simulation der Datenerfassung. Der Datentransfer konzentriert sich auf den Logbuch-Teil des RoBivaL-Korpus, da dieser kontinuierlich wächst (im Gegensatz zur Spezifikation) und bereits vollständig FDO-fähig ist (im Gegensatz zu den hochauflösenden Sensordaten).

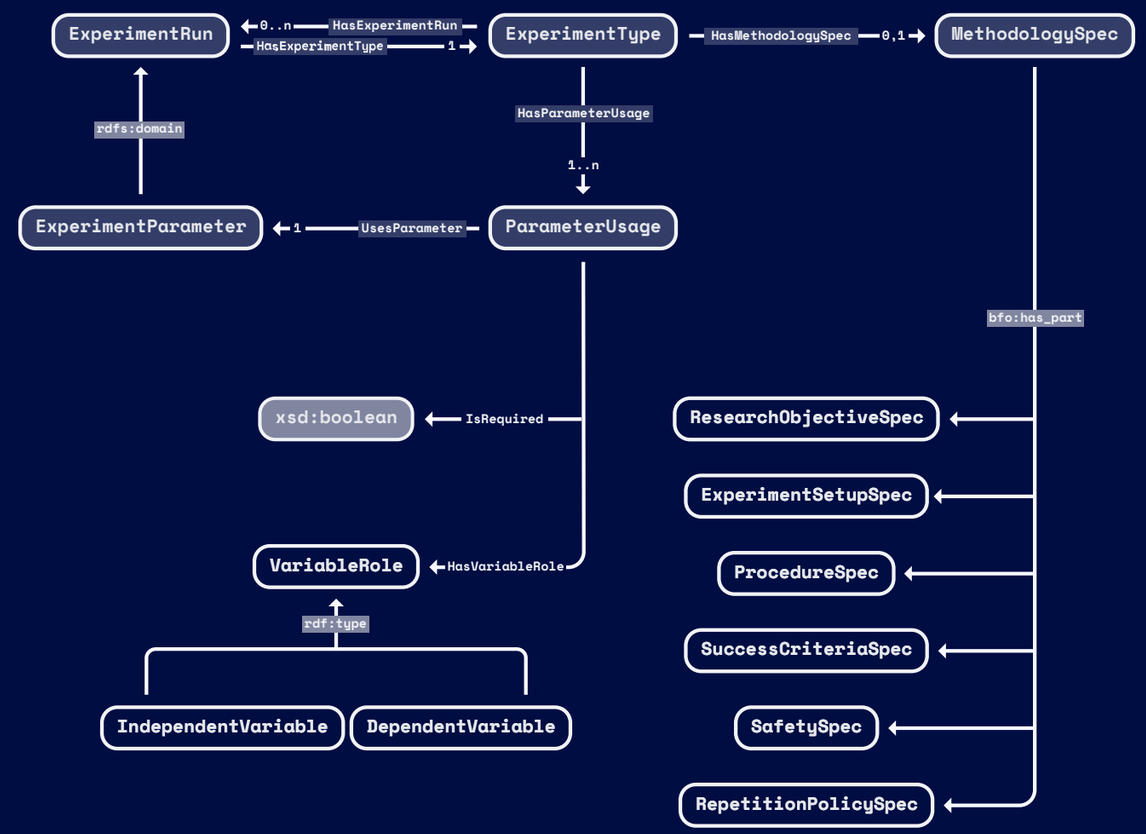
IMPLEMENTIERTE LIEFERPRODUKTE IM ANWENDUNGSFALL ROBOTIK

Im Rahmen des Robotik-Anwendungsfalls wurden alle angekündigten Lieferprodukte fertiggestellt: Die Spezifikation und Registrierung von drei grundlegenden FDO-Profilen wurde abgeschlossen. Darauf aufbauend wurden entsprechende FDO-Records als Nanopublikationen veröffentlicht. Dieses sogenannte RoBivaL FDO Network (RFN) nutzt die separat veröffentlichte Experimental Research Ontology (ERO) als semantisches Fundament. Die Dokumentation der Prozesse und Ergebnisse der Spezifikation, Registrierung und Veröffentlichung wurde abgeschlossen, ebenso wie die Dokumentation eines domänen-spezifischen relationalen Datenbankschemas und einer entsprechenden Datenbank-Instanz, welche die RoBivaL-Daten enthält. Software-Tools für den automatisierten, bidirektionalen Datentransfer wurden fertig entwickelt und getestet. Alle Ergebnisse sind in einer ausführlichen Dokumentation der Methoden, Ergebnisse und Erkenntnisse aus dem Anwendungsfall Robotik zusammengefasst.

> »Ein konkreter Vorteil von FAIR Digital Objects für Daten in der Robotik sind ihre weltweit eindeutigen, dauerhaften IDs.«
>
> **Christian Backe**
> DFKI Robotics Innovation Center, Bremen

GRENZEN UND AUSBLICK

Ein wesentlicher Aspekt ist, dass der aktuelle Anwendungsfall primär Nanopublikationen verwendet. Obwohl theoretisch kein Unterschied für Datennutzer zu Handle-basierten FDOs bestehen sollte, ist eine Replikation mit Handle-basierten FDOs nach Abschluss von FAIR Data Publisher geplant, um die Interoperabilität zu validieren und die Benutzerinnen und Benutzerfreundlichkeit aus der Perspektive der Datenproduktion zu vergleichen. Die Modellierung hochauflösender Sensordaten ist aufgrund der knappen Projektlaufzeit und der Komplexität der Modellierung vorerst nicht im Umfang enthalten. Es bestehen weiterhin Möglichkeiten zur Erweiterung der ERO



•SPEZIFIKATIONEN VON ERO•

Die Experimental Research Ontology (ERO) liefert eine abstrakte semantische Grundlage für den RoBivaL-Datenkorpus. ERO besteht aus 29 RDF/OWL-Entitäten. Sie modellieren ausgewählte Kernaspekte der Datenproduktion in experimenteller Forschung allgemein. (Ein Experimenttyp wird charakterisiert durch Parameter und Methoden. Parameterwerte werden beobachtet bei Experimentdurchläufen. Jeder Durchlauf hat genau einen Experimenttyp, während jeder Experimenttyp eine beliebige Anzahl von Durchläufen haben kann.) ERO ist ausgerichtet an übergeordneten Ontologien, darunter BFO, IAO, OBI und SOSA. Zur besseren Nutzung ist ERO in drei Teile unterteilt: ERO Core ist stabil und enthält alle nötigsten Entitäten. ERO Util enthält optionale Entitäten und ist auf Weiterentwicklung angelegt. ERO Alias enthält menschenlesbare Namen für Entitäten aus übergeordneten Ontologien, wenn der ursprüngliche (numerische) Name die Bedeutung nicht vermittelt.

INTERVIEW
CHRISTIAN BACKE
DFKI ROBOTICS INNOVATION CENTER, BREMEN

>>Warum ist die Nachnutzung von Daten in der Robotik bislang so schwierig?
<<Ein Problem ist, dass Datensätze oft als Monolithe vorliegen, also als große, starre, integrierte Einheiten. Auch wenn ich mich nur für ein paar Messwerte interessiere, muss ich zuerst mehrere Gigabytes herunterladen. Erscheint eine neue Version der Daten, ist nicht sofort klar, welche Teile überhaupt geändert wurden. Komme ich zu einer Erkenntnis, die auf einer bestimmten Messung beruht, kann ich als Referenz trotzdem nur den ganzen Datensatz angeben. Ein weiteres Problem betrifft Semantik und Kontext, also die Bedeutung und Umgebung der Daten. Welche Maßeinheit hat dieser Wert? Nach welcher Methode wurde gemessen? Was bezeichnet dieser Name? Verschiedene Leute haben denselben Begriff für verschiedene Konzepte und verschiedene Begriffe für dasselbe Konzept.

>>Welche konkreten Vorteile bieten FAIR Digital Objects für Daten in der Robotik?
<<Die einzelnen Datenelemente werden granular adressiert, so dass ich sie punktgenau auswählen, verarbeiten und darüber sprechen kann. Ihre IDs sind weltweit eindeutig und dauerhaft. Also können Leute, die an verschiedenen Orten mit denselben Daten arbeiten, einfach sicherstellen, dass ihre Ergebnisse zusammenpassen. Es werden formale Semantiken verwendet, ausgerichtet an etablierten Ontologien. Das verhindert Rätselraten, und es erleichtert die Auffindbarkeit und maschinelle Verarbeitung der Daten.

>>Was sind die nächsten Schritte, um die FDO-Implementierung in der Robotik noch weiter voranzubringen?
<<Um eine solide Basis zu schaffen und die neue Technologie gründlich zu testen, haben wir zuerst den Kontext der Daten sowie einzelne Messwerte in FDOs überführt. Jetzt sind wir bereit für die Skalierung auf große Mengen hochaufgelöster, mehrdimensionaler Messreihen. und zur Entwicklung weiterer spezifischer FDO-Profile. Eine Diskussion in der FDO-Gemeinschaft zur client-seitigen Validierung von FDO-Records und zur Erlaubnis mehrerer FDO-Profile pro Record wird als notwendig erachtet.

FAZIT UND EMPFEHLUNGEN

PROOF OF CONCEPT

- **Jedes Datenelement, von einzelnen Sensorwerten bis hin zu kompletten Experimentdurchläufen, wurde mittels persistenter Identifikatoren granular adressierbar gemacht.**
- **Die Experimental Research Ontology (ERO) mit 29 RDF/OWL-Entitäten wurde entwickelt, um die implizite Semantik des RoBivaL-Korpus zu formalisieren und semantische Interoperabilität sicherzustellen.**
- **Eine hierarchische Struktur von drei FDO-Profilen – FundamentalProfile, TrivialProfile und travelTimeProfile – wurde implementiert, um das Spektrum möglicher Restriktivität aufzuzeigen.**
- **Das RoBivaL FDO Network wurde durch die Implementierung von zwölf exemplarischen Nanopublikationen, bestehend aus drei FDO-Profilen und neun FDO-Records, praktisch umgesetzt.**
- **Eine bidirektionale Transferarchitektur zwischen lokalen SQLite-Datenbanken und der verteilten Nanopublikations-Infrastruktur wurde für Logbuchdaten erfolgreich demonstriert.**
- **Symmetrische Client-APIs wurden geschaffen, um die Nutzung des Datentransfers zu vereinfachen und die Transparenz zwischen Datenbank- und Nanopublikations-Enden zu erhöhen.**
- **Spezifische semantische Transformationsfunktionen wie push_experiment_run() und pull_experiment_run() wurden implementiert, um die Übersetzungslogik zwischen relationalen und RDF-Datenmodellen zu realisieren.**
- **Eine client-seitige SHACL-Validierung unter Verwendung von Apache Jena wurde als Proof of Concept entwickelt, um die Qualitätssicherung und Konformität von FDOs zu gewährleisten.**
- **Eine Simulator-Komponente (simulator.py) wurde implementiert, um synthetische Experimentdaten an beiden Enden der Transferpipeline zu generieren und das System zu testen.**
- **Die Python Nanopub API wurde aktiv genutzt, wodurch ihre Funktionalität und Eignung als integriertes Werkzeug für Datenproduzenten validiert wurde.**
- **Eine gemeinsame UID-Generierung (_mint_numeric_uid()) wurde prototypisch implementiert, um globale Eindeutigkeit der Identifikatoren auch bei paralleler Datengenerierung sicherzustellen.**

MEHRWERTE

+ **Granulare Adressierbarkeit:** Die FDOs ermöglichen eine präzise Verknüpfung wissenschaftlicher Aussagen mit spezifischen Beobachtungen und eine flexible Rekombination von Datensatzkomponenten, indem jedes Datenelement einen persistenten Identifikator erhält.

+ **Semantische Eindeutigkeit:** Durch die Entwicklung der Experimental Research Ontology (ERO) wird die implizite Semantik des Datenkorpus formalisiert und eine domänenübergreifende Interoperabilität gewährleistet.

+ **Verbesserte Reproduzierbarkeit:** Die systematische Dokumentation von Versuchsprotokollen und Metadaten durch FDOs trägt dazu bei, die Reproduzierbarkeit experimenteller Forschungsergebnisse zu erhöhen und der Replikationskrise entgegenzuwirken.

+ **Kontinuierliche Datenpflege und Evolution:** Die FDO-Struktur unterstützt eine fortlaufende Entwicklung und Anpassung des Datenkorpus, im Gegensatz zu starren, monolithischen Architekturen, die Änderungen blockierten.

+ **Erhöhte Interoperabilität:** Durch die Transformation in standardisierte Linked Data und die Ausrichtung an etablierten Ontologien wird eine nahtlose Integration in breitere FAIR-Datenlandschaften und industrielle Systeme ermöglicht.

+ **Antizipation regulatorischer Anforderungen:** Die geschaffenen technischen Strukturen antizipieren zukünftige Anforderungen des europäischen Datenraums, wie Interoperabilität für Robotersysteme und Dokumentationspflichten für KI-Trainingsdaten.

NOTWENDIGE WEITERE SCHRITTE

> *Die FDO-Modellierung hochauflösender Sensordaten muss angegangen werden, um den Kern robotischer Experimente vollständig abbilden zu können.*

> *Die Spezifikationsbereiche für Software, Sensoren und Datenformatierung müssen vollständig erarbeitet und in die FDO-Struktur integriert werden, um die Experimentwiederholung zu gewährleisten.*

> *Das Netzwerk hinter der Nanopublikations-Infrastruktur muss in Zukunft weiter wachsen, um die Kapazität für breite Anwendung von FDOs zu schaffen.*

> *Der Nanopublikations-Client muss um Update- und Delete-Methoden erweitert werden, um eine umfassende Datenverwaltung in Produktivsystemen zu ermöglichen.*

> *Die globale UID-Generierung muss für den Einsatz in Produktivsystemen durch robustere und zukunftssichere Mechanismen ersetzt werden.*

Verweise Robotik:
↗ w3id.org für Subject ID (Nanopublikation)
↗ w3id.org ERO Prefix Vorschlag
↗ w3id.org für URI Prefix
↗ dct:conformsTo (RO-Crate)
↗ https://w3id.org/fdoc/o/terms/hasFdoProfile (verworfen):
↗ ExperimentalResearchOntology Namespace @w3id.org
↗ RoBivaL Namespace @w3id.org
↗ RoBivaL Project (DFKI)
↗ Zenodo RoBivaL Corpus (initial)
↗ Nanodash Platform Home
↗ Resource Description Framework (RDF) Specification
↗ Shapes Constraint Language (SHACL) Specification
↗ Web Ontology Language (OWL) Specification
↗ SQLite Home
↗ Appropriate Uses for SQLite
↗ Python Standard Library
↗ Python Nanopub Package (PyPI)
↗ BFO Home (Basic Formal Ontology)
↗ IAO GitHub (Information Artifact Ontology)
↗ OBI Home (Ontology for Biomedical Investigations)
↗ SOSA Home (Sensor, Observation, Sample, Actuator)
↗ IEEE 1872-2015 (CORA)
↗ ISO 8373-2021
↗ ERO GitHub Repository (Mirror)
↗ RFN GitHub Repository (Mirror)
↗ w3id.org Service Home
↗ w3id.org GitHub Repository
↗ ORCID Home
↗ RSA Cryptosystem (Wikipedia)
↗ Zenodo RoBivaL Corpus Logbook Zip
↗ Python sqlite3 library documentation
↗ SQL Injection (Wikipedia)
↗ SQLite Interface Parameters
↗ SQLite Foreign Keys Documentation
↗ IRI Specification
↗ Nanopub Server Search Facilities
↗ Apache Jena (SHACL CLI Utility)
↗ Python pyshacl library
↗ Discover Mechanical Engineering Journal
↗ FAIR Principles
↗ FDO Forum FDO Requirement Specifications (Version 3.0)

Komplexe Systeme

Im Use Case „Komplexe Systeme“ wurde ein Python-basiertes Softwarepaket zur Kausalitätsanalyse entwickelt, das Zusammenhänge zwischen verschiedenen Datensätzen analysiert und für Anwendungsbereiche von der Marketinganalyse bis zur medizinischen Forschung ausgelegt ist.

¬ Der Anwendungsfall adressiert die Herausforderung, dass Institutionen und Forschungsfelder ihre Daten in unterschiedlichen Formaten und mit verschiedenen Metadatenstandards speichern, was die Auffindbarkeit, Integration und Nutzung erheblich erschwert. Herkömmliche Datenintegrationsmethoden sind häufig auf einfache Metadatenabgleiche beschränkt und scheiterten an der Erkennung tiefergehender kausaler Zusammenhänge zwischen den Daten.

DIE LÖSUNGSSTRATEGIE: FDOS UND KAUSALE INFERENZ

Im Projekt FAIR Data Publisher wurde als Lösung eine Kombination von FDOs und maschinellem Lernen (ML), insbesondere die Methode der kausalen Inferenz, eingesetzt und erfolgreich umgesetzt. Dies eröffnet neue Möglichkeiten, um Zusammenhänge und Datenflüsse besser zu verstehen und automatisch verknüpfte Datenindizes zu erstellen. Es wurde dabei ein System entwickelt, das FDO-Daten im lokalen Kontext als Input aufnimmt und kausale Zusammenhänge identifiziert, anstatt sich nur auf Korrelationen zu konzentrieren. Die Ergebnisse sind nun in einem verlinkten Index zusammengefasst, der die Datenauffindbarkeit verbessert und weitere nützliche Datenquellen aufzeigt.

Dazu wurde ein Modell zur Analyse kausaler Zusammenhänge entwickelt, eine Validierungsumgebung zur Überprüfung des Modells anhand synthetischer und kleiner exemplarischer Realdaten im FDO-Format implementiert sowie eine umfassende Dokumentation erstellt. Um Unabhängigkeit und vielseitige Einsetzbarkeit des Moduls zu gewährleisten, wurden ausschließlich eigens erzeugte Daten und selbst entwickelte FDO-Datenstrukturen genutzt. Allerdings wurden keine Tools aus der FAIR Data Publisher-Architektur für die Analyse oder Verknüpfung verteilter FDOs verwendet; der Fokus lag auf der Kausalanalyse von Daten und nicht auf der Verarbeitung geografisch verteilter FDO-Daten.

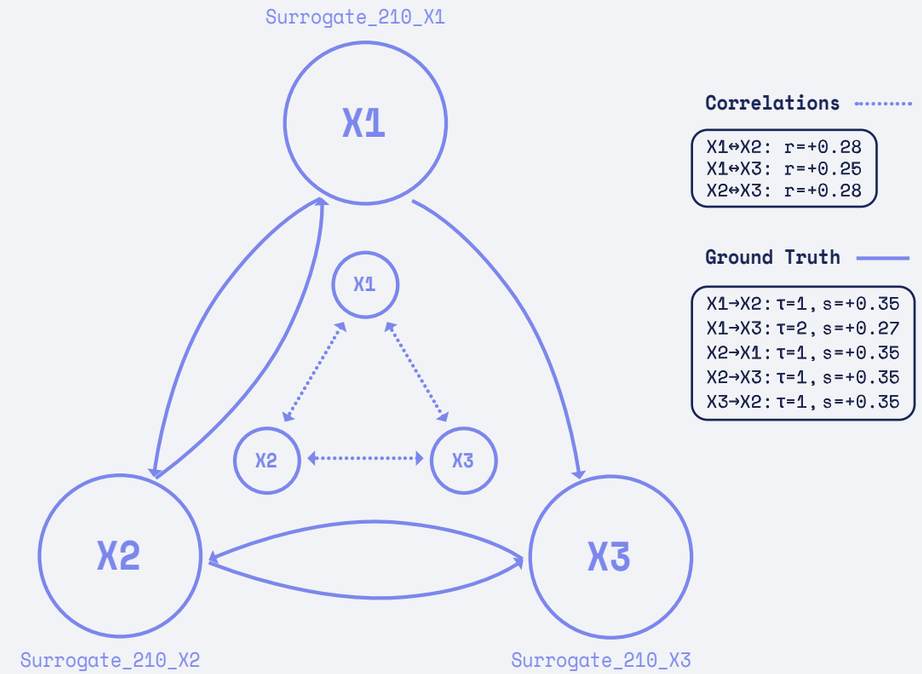
→ Warum ist Kausalitätsforschung mit Daten wichtig?

Das Wissen zur Kausalität wird benötigt, um Ursache-Wirkung-Beziehungen (Inferenz) zu verstehen. Im Vergleich zur Korrelation, also zunächst scheinbare Zusammenhänge, ermöglicht die Kenntnis von Kausalität, gezielt in Prozesse einzugreifen, um erwünschte Ergebnisse zu erzielen oder unerwünschte zu verhindern. In komplexen Systemen wie in der Medizin oder Soziologie kann die Korrelation zu trügerischen Rückschlüssen verleiten, wogegen die Kausalitätsforschung hilft, den tatsächlichen Ursache-Wirkung-Beziehungen auf den Grund zu gehen. Die Abbildung klarer Kausalitäten hilft der Künstlichen Intelligenz, über die Korrelation hinaus Zusammenhänge klarer zu verstehen und somit zuverlässigere und effizientere Ergebnisse zu erzielen.

TECHNISCHE REALISIERUNG VON FDO-KAPSELN UND PCMCI-ALGORITHMUS

Die Demonstration erfolgte primär mit Surrogat-Daten, die vordefinierte Kausalhierarchien enthielten, um die prinzipielle Anwendbarkeit des Frameworks zu evaluieren. FDOs mit symbolischen persistenten Identifikatoren (PIDs) und maschinenlesbaren Metadaten wurden als Input in die Analyse integriert. Die Ergebnisse der Kausalanalyse wurden als FDO gespeichert und über PIDs mit deren Datenquellen verknüpft, wodurch diese für weitere Analysen auffindbar gemacht wurden.

Auch wurde ein prototypischer Python-Code entwickelt, der lokale FDOs erzeugte und verarbeitete. Beispielsweise im JSON-Format mit symbolischen PIDs. Der Austausch mit anderen Arbeitspaketen im Rahmen des Projektes war im aktuellen Prototyp nicht vorgesehen, jedoch wurde die Software modular aufgebaut und kann zukünftig für Schnittstellen erweitert werden.



•GROUND TRUTH (SURROGATE_210)•

Wahre Kausalstruktur des Surrogat-Datensatzes SURROGATE_210: Die drei autoregressiven Prozesse (X1, X2, X3) sind in einem gerichteten Zyklus verbunden, der mit dem FDO-Surrogatgenerator (fdo_create_surrogate.py, Modell „210“) erzeugt wurde. Das Modell definiert einen linearen multivariaten AR-Prozess mit folgenden nicht-null Kanten: X1→X2 (τ=1, s=+0,35), X1→X3 (τ=2, s=+0,27), X2→X1 (τ=1, s=+0,35), X2→X3 (τ=1, s=+0,35) und X3→X1 (τ=1, s=+0,35). Die Knoten sind in einem gleichseitigen Dreieck angeordnet; Pfeile verlaufen knapp außerhalb der Knoten, ihre Krümmung ergibt sich geometrisch. Zeitgleiche paarweise Korrelationen (r≈0,25–0,28) sind separat angegeben und verdeutlichen, dass einfache Korrelationen den zugrunde liegenden zeitverzögerten Kausalzyklus nicht erfassen.

Das zentrale Werkzeug war der PCMCI-Algorithmus aus Tigramite, der auf triadischen Zeitreihen angewendet wurde und verzögerte, bedingte Abhängigkeiten inferierte.

VALIDIERUNG MITTELS SURROGATDATEN

Für die Kausalanalyse komplexer Systeme wurden reproduzierbare I/O-Kapseln im FDO-Format genutzt, die Daten, Parameter, Code-Umgebung und Ergebnisse bündeln. Die Demonstration erfolgte mit Surrogatdatensätzen, die bekannte kausale Strukturen (z. B. Common Cause, Chain, Cycle) enthielten. Die Ergebnisse zeigten eine perfekte Wiederfindung: Precision = 100 % und Recall = 100 % (bei exakten Lags) über alle getesteten Surrogates.

Zusätzlich wurden kleine exemplarische Realdatensätze (z. B. aus Physiologie, Klima, Finanzen) eingesetzt, um die Pipeline Ende-zu-Ende zu demonstrieren. Hier lagen keine Ground-Truth-Daten vor, sodass diese Anwendungen als Plausibilitäts-Checks dienen, nicht als Validierung der Inferenzqualität.

ERGEBNISSE DES ANWENDUNGSFALLS „KOMPLEXE SYSTEME“

Der Schwerpunkt dieses Projekts lag auf der prototypischen Implementierung des PCMCI-Frameworks (zur kausalen Entdeckung in Zeitreihen) in FAIR Digital Objects sowie in selbstbeschreibenden FDO-Kapseln. Um die Inferenzqualität objektiv bewerten zu können, wurde eine Bibliothek

INTERVIEW DR. JAN NAGLER, FSOFM

>>Was war denn die Hauptmotivation hinter diesem Projekt zur automatisierten Kausalanalyse?
<<Wir wollten endlich das Problem lösen, dass Forscher zwar komplexe Systeme untersuchen, aber ihre Kausalanalysen meist nicht reproduzierbar sind und in irgendwelchen Black Boxes verschwinden.

>>Was war die größte technische Herausforderung bei der Umsetzung?
<<Die kniffligste Sache war, die PCMCI-Methode so robust anzuwenden, dass sie über verschiedene Parameter-Konfigurationen stabile Ergebnisse liefert, ohne dass dabei die Transparenz verloren geht.

>>Welchen Mehrwert bringt das System für die wissenschaftliche Praxis?
<<Forscher können jetzt mit einem Klick ihre Zeitreihendaten analysieren und bekommen alle Daten und Parameter so verpackt, dass andere das Experiment exakt nachvollziehen können.

von Surrogatdatensätzen mit bekannter Ground Truth (GT) erstellt. Die Validierungsumgebung erzielte 100 % Precision/Recall auf die Surrogates-Daten.

Zusätzlich zur Validierung wurde der Ansatz durch erste Tests mit kleinen Realdatensätzen demonstriert, die aus fünf Domänen stammten, darunter Physiologie, Finanzen, Klima, Hydrologie und Energie.

FAZIT UND EMPFEHLUNGEN

PROOF OF CONCEPT

- **Das Proof-of-Concept wurde erfolgreich demonstriert: Surrogates mit Ground Truth wurden in FDO-Kapseln integriert, durch PCMCI analysiert und die Ergebnisse als FDO gespeichert. Ein Stability-Dashboard (Grid-Sweep über α und τ_max) wurde genutzt, um die Robustheit zu prüfen.**

MEHRWERTE

+ **Hohe Reproduzierbarkeit durch FDO-Kapseln** (Kombination von Daten, Parametern, Code, Ergebnissen)

+ **Perfekte Validierung auf Surrogates** (100 % Precision/Recall)

+ **Erhöhte FAIRness** durch strukturierte, portable Artefakte

+ **Übertragbarkeit auf verschiedene Anwendungsdomänen** (Klima, Finanzen, Neuro, Gesundheit) mit Anpassungsaufwand

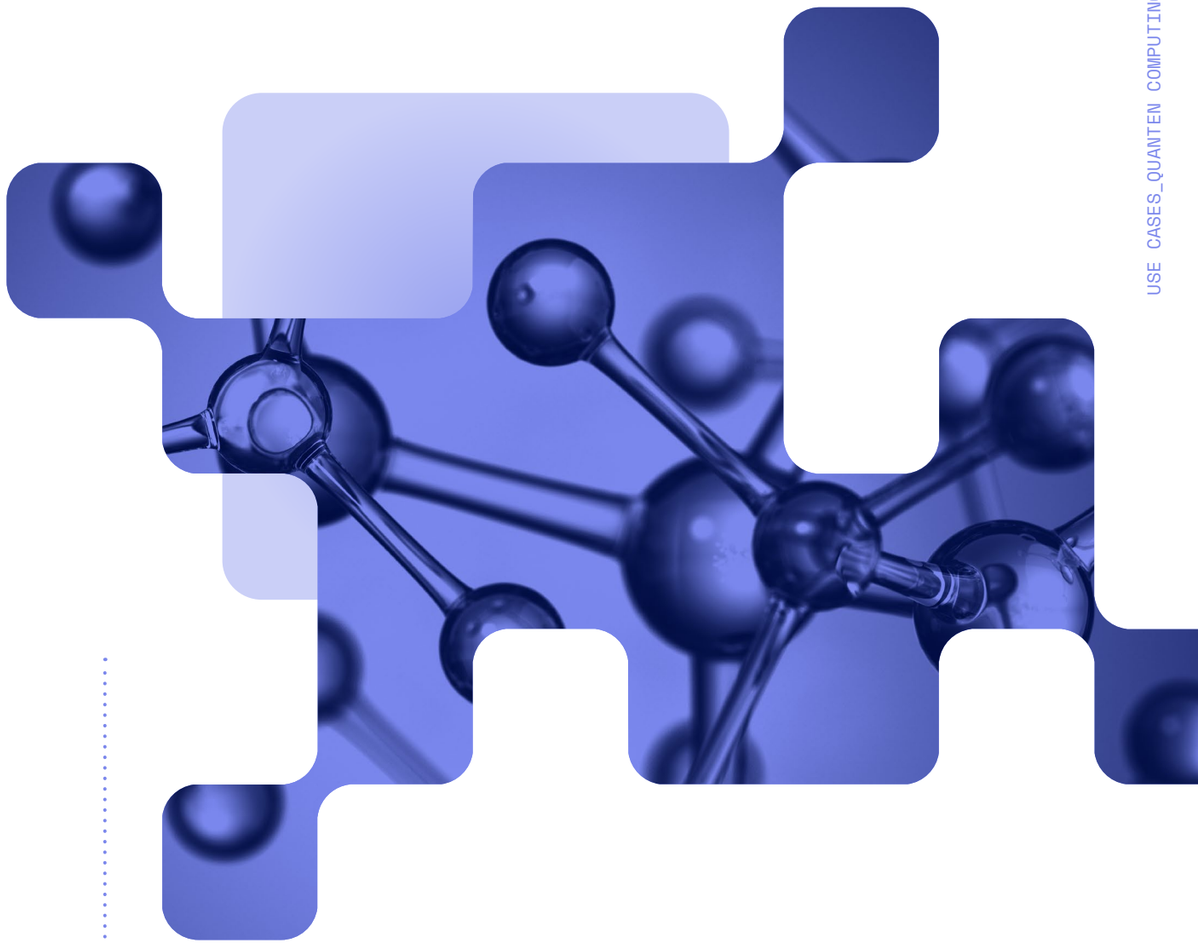
NOTWENDIGE WEITERE SCHRITTE

> *Der aktuelle Prototyp war limitiert auf triadische Systeme und lineare Zusammenhänge. Vorverarbeitungsschritte erfolgten noch manuell. Für die nächste Phase wurden Erweiterungen vorgesehen, u. a. Integration nichtlinearer Tests, τ=0 Fälle, sowie Anbindung an Nanopublikations-Adapter und AP1-Profile.*

Verweise Komplexe Systeme:
↗ Komplexe Systeme bei Wikipedia
↗ Nanopublikation zu einer kausalen Relation („povertyCausesCrime“)

Quanten Computing

Der Anwendungsfall Quantencomputing für den FAIR Data Publisher verfolgt das Ziel, erstmals konkrete Daten aus Experimenten des Quantenmaschinellen Lernens (Quantum Machine Learning, QML) als spezifische FDOs bereitzustellen. Damit soll ein erster Schritt hin zu einer tragfähigen Grundlage für wissenschaftliche Vergleichbarkeit und Wiederverwendbarkeit im Bereich des Quantencomputings gemacht werden.



¬ Der Vergleich von Quantencomputing-Algorithmen (QC-Algorithmen) ist bislang nur eingeschränkt möglich. Unterschiedliche Hardwareplattformen, divergente Simulatoren, heterogene QML-Modelle sowie uneinheitliche Evaluationsverfahren führen dazu, dass wissenschaftliche Ergebnisse oftmals isoliert nebeneinanderstehen.

> »Der Mehrwert liegt in der Reproduzierbarkeit und Transparenz, weil jedes Q-FDO eindeutig auf die Artefakte und Konfigurationen verweist.«
>
> **Patrick Draheim**
> DFKI

Damit fehlt ein konsistenter Rahmen, um Aussagen über die Leistungsfähigkeit, Anwendungsbreite und die praktischen Beschränkungen von Quantencomputing zuverlässig treffen zu können. Durch die Bereitstellung von QML-Experimentdaten im FDO-Format ermöglicht die FDO-Architektur einen strukturierten, maschinenlesbaren und verständlichen Zugriff auf diese Ergebnisse.

Zudem ist die Durchführung systematischer Evaluationen unabdingbar, um den praktischen Nutzen und die Potenziale des Quantencomputings einordnen zu können. Nur durch den Vergleich vieler Experimente wird erkennbar, in welchen Anwendungsfeldern, Domänen und Problemstellungen tatsächlich ein Quantenvorteil erzielt werden kann – also eine Überlegenheit quantenbasierter Methoden gegenüber klassischen Ansätzen. Ohne eine gemeinsame Datenbasis bleibt die Diskussion über Quantenvorteile spekulativ und fragmentiert. Die im Rahmen des Use Cases Quantencomputing im Projekt bereitgestellten QuantenFDOs (Q-FDOs) stellen daher einen entscheidenden ersten Schritt hin zu einer ganzheitlichen, transparenten und offenen wissenschaftlichen Vergleichbarkeit dar. Durch sie wird es möglich, strukturierte Erkenntnisse zu generieren, die nicht nur innerhalb einzelner Forschungsgruppen, sondern domänenübergreifend nachvollzogen, reproduziert und weiterverarbeitet werden können.

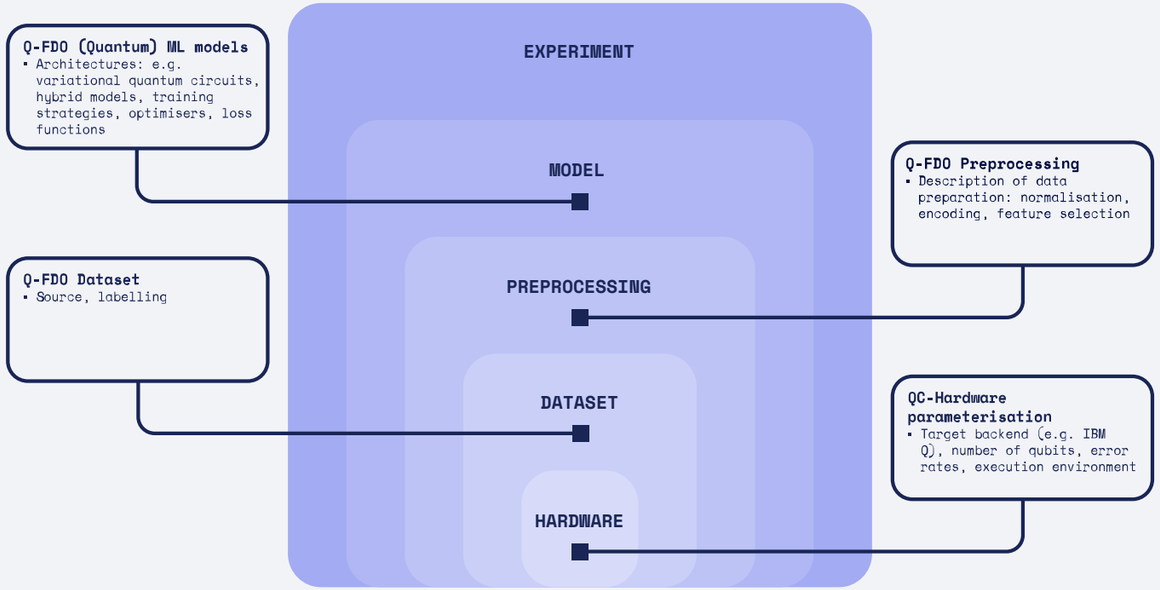
Außerdem eröffnet die Bereitstellung von Q-FDOs für QML einen unmittelbaren Mehrwert über das Quantencomputing hinaus. Da QML methodisch eng mit klassischen Ansätzen des Maschinellen Lernens (ML) und des Deep Learnings (DL) verwandt ist, können die entwickelten Strukturen und Modelle auf diese Bereiche übertragen werden. Der gleiche Daten- und Metadatenrahmen – bestehend aus Hardwarebeschreibungen, Vorverarbeitungsmethoden, Datensätzen, Modellarchitekturen, Evaluationsparametern und Hyperparametern – ist auch im klassischen ML relevant. Damit leistet das Arbeitspaket nicht nur einen Beitrag zur Etablierung von Standards im Quantenbereich, sondern trägt auch zu einer breiteren Interoperabilität zwischen Quanten- und klassischen Lernverfahren bei. Damit schafft das Projekt die Voraussetzungen für ein Ökosystem, in dem wissenschaftliche Erkenntnisse im Bereich des Quantenmaschinellen Lernens nicht nur erzeugt, sondern auch langfristig zugänglich, überprüfbar und nutzbar gemacht werden.

Der Use Case zielte im Projekt darauf ab, eine praxisnahe Umsetzung von Q-FDOs zu entwickeln und deren Mehrwert exemplarisch darzustellen. Um die Komplexität des Feldes zu begrenzen und eine valide, nachvollziehbare Evaluationsgrundlage im Rahmen der Projektlaufzeit zu schaffen, wird der Fokus auf den Anwendungsfall der Binärklassifikation von Bilddaten gelegt.

DAS Q-FDO KONZEPT

Ein Q-FDO soll detailliert Informationen zu den verwendeten Komponenten eines Quantenexperiments festhalten: die genutzte Hardware, den Datensatz inklusive Vorverarbeitung zur Anpassung an Quantencomputer, das spezifische Quanten-Lernmodell (z. B. Quantenneuronale Netze) und die erzielten Ergebnisse des Experiments.

Dazu wurde die Entscheidung getroffen, Q-FDOs als domänenspezifische Spezialisierung von Nanopublikationen einzuführen. Hintergrund ist, dass Nanopublikationen zwar ein allgemein anerkanntes, standardisiertes Format darstellen, um Forschungsdaten und -ergebnisse FAIR-konform zu publizieren, sie legen jedoch nicht fest, welche inhaltlichen Strukturen für ein bestimmtes Fachgebiet zwingend zu berücksichtigen sind. Nanopublikationen bieten damit einen stabilen Container, der Referenzierbarkeit, Provenienz-Informationen und die langfristige



•Q-FDO IM SCHICHTENMODELL•

Das hierarchische Schichtenmodell umfasst fünf zentrale Komponenten des Q-FDO und verdeutlicht die Komplexität von QML-Experimenten sowie die Notwendigkeit einer strukturierten, standardisierten Dokumentation zur Vergleichbarkeit von Experimenten für diesen aufstrebenden Technologiebereich.

Nachvollziehbarkeit sicherstellt, aber keine fachspezifische Granularität erzwingt.

Q-FDOs greifen genau an diesem Punkt: Sie definieren eine hierarchische und modulare Struktur für Quantencomputing-Experimente, die sechs zentrale Teildefinitionen umfasst. Damit wird sichergestellt, dass wesentliche Aspekte von Quanten-ML-Experimenten systematisch beschrieben werden. So können beispielsweise Vorverarbeitungen (z. B. PCA, Standardisierung), Modellarchitekturen (z. B. Data Reuploading Classifier, VQC), Hyperparameter oder Hardware-Backends in klaren, wiederverwendbaren Einheiten dargestellt werden. Die einzelnen JSON-Dateien, die operativ in die Evaluations-Pipeline eingesetzt werden, bilden dafür die technische Grundlage. Sie werden in Nanopublikationen eingebettet, die auf Plattformen wie B2Share archiviert und mit persistenten Identifikatoren (DOI) versehen werden. Die Q-FDOs wiederum referenzieren diese Nanopublikationen und abstrahieren sie in einer vereinheitlichten, stringbasierten Struktur. Hierbei wurden die vorliegenden Q-FDOs und Nanopublikationen in enger Zusammenarbeit mit den Partnern im Projekt erarbeitet. So ist das genutzte Repository über den in der FDO-Architektur erstellten B2Share-Adapter nutzbar, und die Nanopublikationen werden in dem Nanopub-Network publiziert.

Der entscheidende Vorteil dieses hierarchischen Vorgehens liegt in der Kombination: JSON-Dateien gewährleisten die technische Ausführbarkeit der Pipeline, Nanopublikationen sorgen für die FAIR-konforme Publikation und Nachvollziehbarkeit und Q-FDOs stellen die domänenspezifische Vergleichbarkeit sicher. Durch dieses Zusammenspiel entsteht eine dreistufige Architektur (JSON, Nanopublikation, Q-FDO), die sowohl operative Nutzung, FAIR-Publikation als auch domänenspezifische Vergleichbarkeit sicherstellt.

Die Q-FDO-Struktur gliedert sich zudem in sechs zentrale Definitionen, welche die

verschiedenen Aspekte eines Quantenexperiments abbilden. Diese Teildefinitionen sind: Q-FDO-Hardware, Q-FDO-Datensatz, Q-FDO-Vorverarbeitung, Q-FDO-(Quantum)ML-Modellarchitektur, Q-FDO-Experiment und Q-FDO-Ergebnis. JSON-Dateien bilden dabei die technische Grundlage, die operativ in der Evaluations-Pipeline eingesetzt werden.

INTEGRATIONSASPEKTE UND SYNERGIEN

Die entwickelten FDO-Blaupausen für Q-FDOs sind anpassbar und können auch für klassisches maschinelles Lernen genutzt werden, da viele Prozesse, wie die Datenvorverarbeitung, zu einem großen Teil identisch sind. Dies schafft Synergien und erleichtert die Übertragbarkeit zwischen Quanten- und klassischen Lernverfahren. Die Q-FDOs sind dabei integrativ gedacht und sollen direkt in der Verarbeitungspipeline verwendet werden, nach dem Prinzip „FDO in, FDO out“, um Brüche in der Datenverarbeitung zu vermeiden.

Im Rahmen des Projekts wurde außerdem eine Proof-of-Concept-Pipeline-Software, die Quantum Experiment Evaluation Pipeline (QEEP), entwickelt. Diese Pipeline nutzt die in den Q-FDOs definierten Konfigurationen, um Experimente auf simulierter Quantenhardware durchzuführen. Die JSON-Dateien, welche die Grundlage für die Q-FDOs bilden, werden dabei direkt von der Pipeline eingelesen und parametrisieren die Vorverarbeitung, Modelle und Evaluation. Dies demonstriert die direkte operative Relevanz der Q-FDOs.

VORTEILE DER Q-FDO NUTZUNG

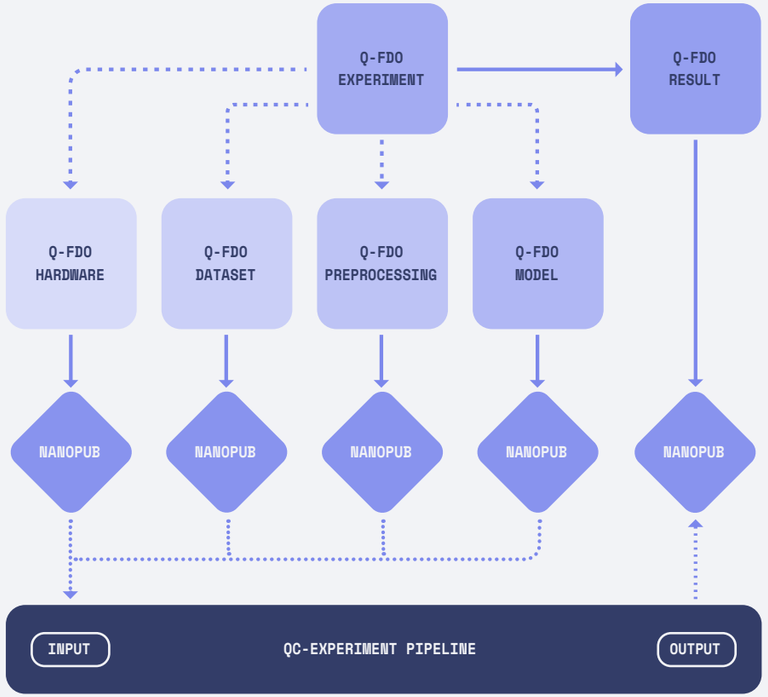
Die modulare Struktur der Q-FDOs ermöglicht eine gezielte Analyse einzelner Experimentaspekte, wie Hardwarebeschränkungen oder Modellarchitekturen, und damit eine verbesserte strukturelle Vergleichbarkeit. Zudem besteht durch die Nähe zu klassischen ML-Strukturen eine domänenübergreifende Nutzungsmöglichkeit, die zur Harmonisierung der Forschungspraktiken beiträgt. Schließlich besitzen Q-FDOs einen operativen Mehrwert, da sie direkt in experimentelle Pipelines eingebunden werden können, was die Effizienz und Nachhaltigkeit von Forschungsprozessen steigert.

Die Einführung von Q-FDOs im Rahmen von FAIR Data Publisher adressiert eine zentrale Herausforderung der Quantencomputing-Forschung: die bislang eingeschränkte Vergleichbarkeit und Wiederverwendbarkeit von Experimentdaten. Aufbauend auf der in bereits dargestellten Motivation sowie den konkret formulierten Zielen, ermöglichen Q-FDOs eine strukturierte und FAIR-konforme Repräsentation von QML-Experimenten, die einen erheblichen Mehrwert für Forschung, Entwicklung und Anwendung generiert. Ein erster Vorteil liegt in der Standardisierung und Interoperabilität. Durch die klare Definition von Kernfeldern (z. B. id, description, author, nanopub) sowie spezifischen Erweiterungen je nach Typ (Hardware, Preprocessing, Modell, Dataset, Experiment, Result) wird ein konsistenter Rahmen geschaffen, der die heterogene Landschaft an QML-Ansätzen harmonisiert. Damit wird es möglich, Experimente unterschiedlicher Plattformen, Simulatoren oder Modellvarianten in ein gemeinsames Schema zu überführen, ohne deren inhaltliche Spezifik zu verlieren.

Darüber hinaus fördern Q-FDOs die Reproduzierbarkeit und Transparenz: jedes Q-FDO referenziert über Nanopublikationen eindeutig identifizierbare Artefakte. Dies stellt sicher, dass ein Experiment nicht nur einmalig nachvollzogen werden kann, sondern langfristig überprüfbar und in anderen Kontexten wiederholbar bleibt. Gerade im Quantencomputing, wo geringe Änderungen in Hardware oder Software große Auswirkungen auf Ergebnisse haben können, schafft diese Rückverfolgbarkeit eine bislang fehlende Grundlage für belastbare wissenschaftliche Vergleiche. Ein weiterer Vorteil ist die strukturelle Vergleichbarkeit. Die Modularisierung in verschiedene Q-FDO-Typen erlaubt es, einzelne Aspekte eines Experiments gezielt zu analysieren: Hardware-Beschränkungen können separat von

→ Architektur-Diversität der Quantencomputer

Anders als klassische Computer, die auf Silizium-Transistoren basieren, existiert keine Standardbauweise für Quantencomputer. Stattdessen konkurrieren fundamentale Architekturen wie supraleitende Qubits (extrem gekühlt) und Ionenfallen-Qubits (elektromagnetisch gefangene Atome). Diese Systeme unterscheiden sich nicht nur im Aufbau, sondern auch in ihren inhärenten Fehlerprofilen (Rauschlevel), der Konnektivität zwischen den Qubits und der Güte der Quantengatter. Ein Algorithmus, der für die Gitterstruktur eines supraleitenden Prozessors optimiert ist, kann auf einem vollständig verbundenen Ionenfallen-System ineffizient sein. Somit sind die Ergebnisse stark an die spezifische Hardware und deren Limitierungen gebunden. Deswegen war der Vergleich von Quantencomputing-Algorithmen bislang nur eingeschränkt möglich.



•PROZESS EINES Q-FDO•

Der Workflow der Q-FDOs veläuft in einer dreistufigen Architektur. Die Q-FDO-Komponenten Hardware, Dataset, Preprocessing, Model und Result werden durch ein übergeordnetes „Q-FDO Experiment“ aggregiert und die Ergebnisse des Quantenexperiments in „Q-FDO Result“ dokumentiert. Die mittlere Ebene bilden die Nanopublikationen, die als FAIR-konforme Container fungieren. Jede Q-FDO-Komponente ist mit einer entsprechenden Nanopublikation verbunden und im B2Share Repository archiviert und mit persistenten Identifikatoren versehen, wodurch sie dauerhaft referenzierbar und nachvollziehbar werden. Die QC-Experiment Pipeline nutzt die in den Nanopublikationen gespeicherten Konfigurationsdaten für die tatsächliche Durchführung der Quantencomputing-Experimente.

USE CASES_QUANTEN COMPUTING

INTERVIEW
PATRICK DRAHEIM,
DFKI

>>Was ist eigentlich die größte Hürde in der Quantencomputing-Forschung, die Sie mit den FDOs angehen wollten?
<<Im Quantencomputing ist die Hardwarelandschaft total dynamisch und heterogen. Das führt dazu, dass die Ergebnisse von Experimenten oft isoliert nebeneinanderstehen und kaum vergleichbar sind. Ohne einen konsistenten Rahmen ist es schwer zu beurteilen, ob in bestimmten Anwendungsbereichen überhaupt ein Quantenvorteil existiert.

>>Wie genau lösen Sie dieses Problem der fehlenden Vergleichbarkeit mit den neuen Quanten-FDOs (Q-FDOs)?
<<Wir haben die Q-FDOs als domänenspezifische Spezialisierung von Nanopublikationen konzipiert, die eine hierarchische und modulare Struktur für Quantenexperimente schaffen. Diese Q-FDOs erfassen systematisch alle relevanten Aspekte, wie die verwendete Hardware, die Datenvorverarbeitung und das Quanten-Modell. So sorgen wir dafür, dass die Experimente strukturiert und FAIR-konform dokumentiert sind.

>>Und was ist der konkrete Nutzen, den die Forschung jetzt durch diese standardisierten Q-FDOs hat?
<<Der Mehrwert liegt in der Reproduzierbarkeit und Transparenz, weil jedes Q-FDO eindeutig auf die Artefakte und Konfigurationen verweist. Durch die Modularisierung in Typen wie Hardware, Modell und Datensatz können Forschende gezielt Hypothesen über Ursache-Wirkungs-Zusammenhänge systematisch prüfen. Das Beste ist, dass die Q-FDOs auch einen operativen Nutzen haben: Sie steuern direkt die Quantum Computing Experiment Pipeline (QCEP).

>>Wenn Sie vorausschauen: Was sind die nächsten großen Schritte, um das volle Potenzial der Q-FDOs in der Praxis auszuschöpfen?

Modellarchitekturen oder Preprocessing-Strategien betrachtet werden. Dadurch lassen sich nicht nur Ergebnisse zwischen Experimenten vergleichen, sondern auch Hypothesen über Ursache-Wirkungs-Zusammenhänge (z. B. Einfluss der Hardware vs. Einfluss der Datenaufbereitung) systematisch überprüfen. Überdies eröffnen Q-FDOs durch ihre Nähe zu bestehenden Strukturen im klassischen Machine Learning die Möglichkeit einer domänenübergreifenden Nutzung. Da viele Metadatenfelder (z. B. für Modelle, Datensätze oder Preprocessing) auch im klassischen ML relevant sind, entsteht ein direkt anschlussfähiger Rahmen. Dies erleichtert nicht nur die Übertragbarkeit zwischen Quanten- und klassischen Methoden, sondern trägt zu einer generellen Harmonisierung der Forschungspraktiken in KI-gestützten Wissenschaftsbereichen bei.

Schließlich besitzen Q-FDOs einen operativen Mehrwert. Sie sind nicht nur eine abstrakte Metadatenebene, sondern können unmittelbar in experimentelle Pipelines (wie QCEP) eingebunden werden. Dies erlaubt es, aus den in Q-FDOs dokumentierten Konfigurationen automatisiert reale Experimente zu generieren und deren Ergebnisse wiederum in strukturierter Form zurückzuführen. Damit fungieren Q-FDOs als Brücke zwischen Dokumentation und praktischer Ausführung – ein entscheidender Vorteil für die Effizienz und Nachhaltigkeit von Forschungsprozessen.

Zusammenfassend bieten Q-FDOs im Zusammenhang mit QC-Experimenten also eine Vielzahl von Vorteilen: Sie schaffen Vergleichbarkeit, sichern Reproduzierbarkeit, ermöglichen Transparenz, fördern Interoperabilität und tragen durch ihre operative Einbindung zur praktischen Effizienz bei.

FAZIT UND EMPFEHLUNGEN

PROOF OF CONCEPT

Q-FDO-Struktur und -Profile:
- **Entwicklung einer hierarchischen Q-FDO-Architektur mit sechs Komponenten: Hardware, Dataset, Preprocessing, Model, Experiment und Result**

- **Definition und Publikation von drei FDO-Profilen als Nanopublikationen unter persistenten URIs**

- **13 exemplarische Q-FDOs wurden erstellt (1 Hardware, 2 Datasets, 2 Preprocessing, 4 Modelle, 2 Experimente, 2 Resultate)**

Technische Implementierung:
- **Quantum Computing Experiment Pipeline (QCEP): Vollständig automatisierter Workflow von Datenvorbereitung bis Ergebnisdokumentation**

- **Erfolgreiche Integration mit AP1-Infrastruktur (B2Share Repository, Nanopub Network, Handle-System)**

- **Reproduzierbare Experimente durch Seed-basierte Determinierung**

- **Strukturierte Dokumentation aller Experimentkomponenten in maschinenlesbarer Form**

Kritische Limitationen:
- **Fehlende Automatisierung: Manuelle Konversion zwischen JSON, Nanopublikationen und Q-FDOs**

- **Begrenzte Modellvielfalt: Nur Data Reuploading Classifier implementiert. Folgeaktivitäten sollen dann unüberwachte Lernverfahren ebenfalls integrieren.**

MEHRWERTE

+ Die klare Definition von Kernfeldern und domänenspezifischen Erweiterungen schafft einen konsistenten Rahmen zur Harmonisierung der heterogenen Landschaft an Quanten-Machine-Learning-Ansätzen.

+ Jedes Quanten-FDO (Q-FDO) referenziert über Nanopublikationen eindeutig identifizierbare Artefakte, was die langfristige Überprüfbarkeit und Wiederholbarkeit von Quantenexperimenten sicherstellt.

+ Die Modularisierung der Q-FDOs in Typen wie Hardware, Modell und Vorverarbeitung erlaubt eine gezielte, systematische Analyse einzelner Experimentaspekte zur Hypothesenprüfung über Ursache-Wirkungs-Zusammenhänge.

+ Durch die Übertragbarkeit der Q-FDO-Strukturen auf klassische Ansätze des Maschinellen Lernens entsteht eine breitere Interoperabilität zwischen Quanten- und klassischen Lernverfahren.

USE CASES_QUANTEN COMPUTING

<<Absolut notwendig ist die Automatisierung der Prozesse, denn aktuell müssen die JSON-Dateien noch manuell in Nanopublikationen überführt werden. Langfristig ist die große Idee, eine Software zur automatischen Korrelationserkennung und systemischen Auswertung des Quantenvorteils zu implementieren, da die Q-FDO-Struktur dafür jetzt die Grundlage bietet.

+ Q-FDOs besitzen einen operativen Mehrwert, da sie unmittelbar in experimentelle Pipelines wie die QCEP eingebunden werden können, um Konfigurationen zu steuern und Ergebnisse strukturiert zurückzuführen.

+ Die strukturierte Datenerfassung schafft die notwendige Grundlage, um den praktischen Nutzen und die Potenziale des Quantencomputings systematisch zu bewerten und einen Quantenvorteil einzuordnen.

NOTWENDIGE WEITERE SCHRITTE

> *Vollautomatisierte Konvertierungspipeline entwickeln, die den manuellen Transfer zwischen JSON-Konfigurationen, Nanopublikationen und Q-FDO-Abstraktionen eliminiert und damit produktive Nutzung für iterative Quantenforschung ermöglicht.*

> *Erweiterte Modellvielfalt implementieren, die über den Data Reuploading Classifier hinausgeht und weitere QML-Architekturen wie Variational Quantum Circuits, Quantum Neural Networks und hybride Ansätze standardisiert abbildet.*

> *Zusammenspiel von Nanopublikationen und Q-FDOs sollte weiter präzisiert und operationalisiert werden.*

> *Unüberwachte Lernverfahren integrieren, um neben Klassifikationsproblemen auch Clustering, Dimensionsreduktion und andere wichtige Zweige des Quantum Machine Learning systematisch zu erfassen.*

> *Automatische Korrelationsanalyse implementieren, die Muster zwischen Experimenten erkennt und systematische Vergleiche verschiedener Hardware-Backends, Algorithmus-Parameter und Datensätze ermöglicht.*

> *Produktive Validierung mit realen Quantencomputern durchführen, die über synthetische MNIST-Daten hinausgeht und die Praxistauglichkeit der Q-FDO-Struktur in echten Forschungsumgebungen mit IBM Q, Google Sycamore oder anderen Plattformen demonstriert.*

Verweise Quanten Computing:
↗ MNIST Datensatz: Quelle des Datensatzes

STRATEGISCHE BETRACHTUNGEN

Rechtliche Einordnung von FDOs

FDO im allgemeinen Rechtsrahmen

FDOs stehen an der Schnittstelle zwischen technologischer Innovation und regulatorischer Wirklichkeit. Die im Rahmen des Projekts FAIR Data Publisher erstellte Analyse untersucht systematisch, wie diese neuartige Datenarchitektur mit den zentralen Regelwerken der europäischen Datenregulierung interagiert – von der Datenschutz-Grundverordnung über den Digital Services Act bis hin zur KI-Verordnung: als Datenobjekte unterliegen sie datenschutz- und datenrechtlichen Regelungen, als technische Infrastrukturkomponenten produkt- und sicherheitsbezogenen Vorschriften, bei algorithmischer Verarbeitung den Anforderungen der KI-Verordnung und bei plattformbasierter Bereitstellung den Vorgaben des Digital Services Act.

Die Mehrdimensionalität der FDOs spiegelt sich in der technischen Architektur wider: Die hierarchisch-rekursive Struktur von FDOs mit Persistent Identifiers, FDO Records und selbstreferenziellen Attribute-Definitions schafft eine in sich geschlossene, dauerhaft interpretierbare Datenstruktur. Diese Datenobjekte bilden den Kern der datenschutz- und datenrechtlichen Bewertung.

Die sie verwaltenden FDO-Infrastrukturen – Softwarekomponenten und Systeme zur Verarbeitung und Zugänglichmachung – fallen demgegenüber primär unter produkt- und sicherheitsbezogene Vorschriften. Werden FDOs algorithmisch ausgewertet oder über Plattformen distribuiert, greifen zusätzlich die spezifischen Anforderungen der KI-Verordnung beziehungsweise des Digital Services Act. Diese rechtliche Schichtung korrespondiert mit der technischen Trennung zwischen Datenebene und Infrastrukturebene, wobei jede Ebene eigenständigen, teils überlappenden Regelungsregimen unterliegt.

INTEROPERABILITÄT DURCH TECHNISCHE STANDARDISIERUNG

FDOs realisieren die europäische Interoperabilitätsvision durch eine durchgängig standardisierte Architektur. Ihre selbstbeschreibende Struktur mit Persistent Identifiers, maschinenlesbaren Metadaten und dem universellen DOIP-Protokoll schafft eine technische Basis, die heterogene Datenquellen ohne proprietäre Abhängigkeiten verbindet. Jede Interpretationsanweisung existiert als auffindbares FDO in offenen Registries – ein sich selbst tragendes System, das dauerhafte Kompatibilität ohne externe Spezifikationen gewährleistet.

Diese technische Standardisierung adressiert unmittelbar die zentralen Anforderungen des europäischen Datenrechts. Der Data Act fordert strukturierte, gängige und maschinenlesbare Formate sowie die Vermeidung von Vendor-Lock-in-Effekten – Anforderungen, die FDOs durch ihre offenen Standards und einheitlichen Schnittstellen erfüllen. Die im Data Act verankerte Datenportabilität wird durch das DOIP-Protokoll technisch implementiert, während die geforderte Echtzeitverfügbarkeit durch die performante FDO-Infrastruktur sichergestellt wird.

Für den Data Governance Act bilden FDOs eine Grundlage vertrauenswürdiger Datenräume. Granulare Zugriffskontrollen auf Objektebene ermöglichen differenzierte Datenteilung ohne Kontrollverlust. Neutrale Datenvermittlungsdienste können auf FDO-Infrastrukturen aufbauen und dabei technische Neutralität mit rechtlicher Compliance verbinden. Nutzungsbedingungen, Lizenzen und Zugriffsrechte können in strukturierten Metadaten maschinenlesbar dokumentiert und rechtssicher verwaltet werden.

Die einheitliche technische Basis ermöglicht die simultane Erfüllung unterschiedlicher regulatorischer Regime. Die FAIR-konforme Strukturierung unterstützt sowohl die Weiterverwendung öffentlicher Informationen nach PSI-Richtlinie als auch private Datenzugangsrechte nach Data Act. Die persistente Identifizierung gewährleistet dauerhafte Referenzierbarkeit für beide Anwendungsfälle. Hochwertige Datensätze nach PSI-Richtlinie und Echtzeitdaten nach Data Act werden durch dieselben technischen Mechanismen bereitgestellt – ein kohärentes System, das regulatorische Fragmentierung durch technische Konvergenz überwindet.

DATENSCHUTZRECHTLICHE POTENZIALE UND HERAUSFORDERUNGEN

FDOs bieten technische Mechanismen zur Umsetzung datenschutzrechtlicher Anforderungen, werfen jedoch gleichzeitig neue Compliance-Fragen auf. Die Architektur ermöglicht Privacy by Design durch maschinenlesbare Dokumentation von Rechtsgrundlagen, Verarbeitungszwecken und Einwilligungen in strukturierten Metadaten. Die Rechenschaftspflicht nach Artikel 5 Absatz 2 DSGVO erhält eine technische Grundlage: Verarbeitungsvorgänge hinterlassen auditierbare Spuren, Zweckbindungen werden persistent dokumentiert, und die transparente Datenflussstrukturierung unterstützt Datenschutz-Folgenabschätzungen nach Artikel 35 DSGVO.

Die Umsetzung von Betroffenenrechten profitiert von der FDO-Strukturierung. Auskunftsansprüche nach Artikel 15 DSGVO können durch die zentrale Identifizierung und Metadatenverwaltung effizienter bearbeitet werden. Die eindeutige Referenzierbarkeit über Persistent Identifiers erleichtert die Identifikation aller relevanten Datenobjekte einer betroffenen Person. Die Granularität der FDO-Metadaten ermöglicht präzise Eingrenzungen des Verarbeitungskontexts und der beteiligten Systeme, was die Transparenzanforderungen der DSGVO unterstützt.

Die technische Persistenz von FDOs kollidiert jedoch mit datenschutzrechtlichen Löschpflichten und dem Grundsatz der Speicherbegrenzung. Artikel 17 DSGVO fordert die Löschung personenbezogener Daten bei Wegfall der Rechtsgrundlage, Artikel 5 Absatz 1 lit. e DSGVO begrenzt die Speicherdauer auf das erforderliche Maß – die dauerhafte Verfügbarkeit als FDO-Kernprinzip steht beiden Anforderungen entgegen. Die Metadatenreichhaltigkeit, die FDOs auszeichnet, kann dem Grundsatz der Datenminimierung nach Artikel 5 Absatz 1 lit. c DSGVO widersprechen, wenn mehr Kontextinformationen gespeichert werden als für den konkreten Verarbeitungszweck erforderlich.

Grenzüberschreitende Datenflüsse in föderierten FDO-Infrastrukturen erfordern besondere Aufmerksamkeit. Die dezentrale Architektur kann zu Datenübermittlungen in Drittländer führen, die den Anforderungen der Artikel 44 ff. DSGVO genügen müssen. Angemessenheitsbeschlüsse, Standarddatenschutzklauseln oder verbindliche interne Datenschutzvorschriften müssen in die technische Infrastruktur integriert werden. Die Verteilung der Verantwortlichkeiten zwischen Betreibern von FDO-Repositorien, Datenbereitstellern und -nutzern bedarf klarer vertraglicher Regelungen zur gemeinsamen Verantwortlichkeit nach Artikel 26 DSGVO oder zur Auftragsverarbeitung nach Artikel 28 DSGVO.

CYBERSICHERHEIT IN DER MEHRSTUFIGEN REGULIERUNGSARCHITEKTUR

FDO-Infrastrukturen unterliegen einer mehrstufigen normativen Sicherheitsarchitektur, die von der Produktentwicklung über den Betrieb bis zur Diensterbringung reicht. Diese Regulierungsschichten greifen unabhängig davon, ob personenbezogene Daten verarbeitet werden.

STRATEGISCHE BETRACHTUNGEN_RECHTLICHE EINORDNUNG VON FDOS

Auf Produktebene erfasst der Cyber Resilience Act (CRA) FDO-Software als Produkt mit digitalen Elementen. Hersteller müssen Security-by-Design implementieren, Schwachstellen durch systematische Risikobewertungen identifizieren und über die gesamte Supportdauer Sicherheitsupdates bereitstellen. Die Mindestanforderung von fünf Jahren Produktunterstützung entspricht dabei nicht den Anforderungen langfristig angelegter Forschungsdateninfrastrukturen – FDO-Systeme zielen auf Jahrzehnte dauernde Verfügbarkeit. Die CRA-Dokumentationspflichten für Konformitätserklärungen und technische Spezifikationen können teilweise durch die selbstdokumentierende FDO-Metadatenstruktur erfüllt werden, ersetzen diese jedoch nicht vollständig.

Die Betreiberebene wird durch die NIS-2-Richtlinie reguliert, sofern FDO-Infrastrukturbetreiber die Schwellenwerte für wesentliche oder wichtige Einrichtungen erreichen. Cloud-Provider und Rechenzentren, die FDO-Services hosten, müssen umfassende Risikomanagementmaßnahmen implementieren, Incident-Response-Prozesse etablieren und behördliche Meldepflichten erfüllen. Die Richtlinie schafft einheitliche Mindeststandards für die Betriebssicherheit kritischer digitaler Infrastrukturen.

Auf Dienstebene greift § 19 TDDDG für FDO-Infrastrukturen, die als digitale Dienste öffentlich angeboten werden. Die Norm fordert angemessene technische und organisatorische Vorkehrungen zum Schutz der Verfügbarkeit, Integrität, Authentizität und Vertraulichkeit. Diese Schutzpflichten konkretisieren sich bei FDO-Diensten in Zugriffskontrollen, Verschlüsselung der Datenübertragung und Absicherung der DOIP-Endpunkte gegen unbefugte Zugriffe. Die Angemessenheit der Maßnahmen bemisst sich nach Stand der Technik und Schutzbedarf der verarbeiteten Daten.

KI-GOVERNANCE DURCH STRUKTURIERTE DATEN

FDOs können zur Erfüllung der Datenqualitätsanforderungen beitragen, die die KI-Verordnung für Hochrisiko-Systeme statuiert. Artikel 10 KI-VO verlangt von KI-Anbietern die Verwendung repräsentativer, fehlerfreier und vollständiger Datensätze. Die strukturierten Metadaten von FDOs ermöglichen die systematische Dokumentation dieser Qualitätsmerkmale, ersetzen jedoch nicht die erforderlichen Validierungsprozesse.

Die persistente Speicherung von Provenienz-Informationen und Datenaufbereitungsschritten in FDO-Metadaten unterstützt die geforderte Nachvollziehbarkeit. Versionierung und Änderungshistorien bleiben dauerhaft referenzierbar. Diese technische Transparenz erleichtert die Erstellung der nach Artikel 11 KI-VO erforderlichen technischen Dokumentation, wobei die rechtliche Verantwortung für deren Vollständigkeit beim KI-Anbieter verbleibt.

Bias-Analysen und Qualitätsprüfungen nach Artikel 10 Absatz 2 lit. f KI-VO bleiben externe Prozesse, die jedoch von der FDO-Strukturierung profitieren. Die einheitlichen Metadatenschemata ermöglichen automatisierte Vorprüfungen auf Vollständigkeit und formale Konsistenz. Inhaltliche Verzerrungen müssen durch spezialisierte Analysewerkzeuge identifiziert werden, deren Ergebnisse wiederum als FDO-Metadaten persistiert werden können. Die FAIR-Prinzipien überschneiden sich mit den Qualitätszielen der KI-VO, ohne diese vollständig abzudecken – Interoperabilität erleichtert die Zusammenstellung domänenübergreifender Trainingsdatensätze, Wiederverwendbarkeit reduziert redundante Datenaufbereitung.

KI-INTEGRATION IN FDO-INFRASTRUKTUREN

FDO-Infrastrukturen nutzen zunehmend KI-Komponenten zur Automatisierung der Metadaten-Generierung und Datenklassifizierung. Large Language Models extrahieren strukturierte Informationen aus unstrukturierten Quellen, Natural Language Processing identifiziert Entitäten und Beziehungen, Machine Learning klassifiziert Daten nach domänenspezifischen Taxonomien. Diese KI-gestützte Verarbeitung reduziert manuelle Annotationsfehler und erhöht die Konsistenz der Metadatenqualität.

Der Einsatz solcher KI-Systeme in FDO-Komponenten löst die Anwendbarkeit der KI-Verordnung aus. Die Systemabgrenzung folgt dabei der funktionalen Betrachtung: Eigenständige technische Einheiten mit abgrenzbarer KI-Funktionalität werden jeweils als separate KI-Systeme im Sinne der Verordnung qualifiziert, unabhängig von ihrer Integration in übergreifende Infrastrukturen. Die konkrete Risikoeinstufung hängt vom Verwendungszweck ab: Automatisierte Metadaten-Generierung für wissenschaftliche Datensätze fällt regelmäßig nicht in den Hochrisikobereich, unterliegt aber dennoch den allgemeinen Transparenzpflichten nach Artikel 50 KI-VO. Bei Einsatz in kritischen Bereichen – etwa bei der Verarbeitung gesundheitsbezogener Forschungsdaten – können strengere Anforderungen an Datenqualität (Art. 10 KI-VO) und menschliche Aufsicht (Art. 14 KI-VO) greifen.

Die strukturierte FDO-Architektur erleichtert dabei die Compliance: Die transparente Datenorganisation unterstützt die geforderte Nachvollziehbarkeit, versionierte Metadaten dokumentieren KI-generierte Annotationen, und standardisierte Schnittstellen ermöglichen die Integration menschlicher Überprüfungsprozesse.

> »FDOs unterstützen KI-Governance durch strukturierte Dokumentation von Trainingsdaten und transparente Provenienz-Nachweise, was besonders für Hochrisiko-Systeme relevant ist.«
>
> **David Wagner**
> Spirit Legal

Die regulatorischen Anforderungen erhöhen zwar den Implementierungsaufwand, werden aber durch die Effizienzgewinne automatisierter Verarbeitung in großskaligen Dateninfrastrukturen kompensiert.

REGULATORISCHE ANFORDERUNGEN ALS ENTWICKLUNGSPFAD – AUSGEWÄHLTE BEISPIELE

Die praktische Relevanz von FDOs zeigt sich besonders in hochregulierten Sektoren, wo spezifische Compliance-Anforderungen die technische Weiterentwicklung vorantreiben. Zwei exemplarische Anwendungsfelder illustrieren dieses Wechselspiel zwischen regulatorischen Vorgaben und technischer Innovation.

Der European Health Data Space (EHDS) definiert mit seiner gestaffelten Einführung konkrete Harmonisierungsanforderungen für Gesundheitsdaten. Die Verordnung verpflichtet zur Implementierung standardisierter Austauschformate, wobei Kapitel IV zur Sekundärnutzung ab 2028 gilt und die sechs prioritären Datenkategorien – von Patientenkurzakten über medizinische Bildgebung bis zu Entlassungsberichten – gestaffelt ab 2029 und 2031 zur Anwendung kommen.

FDO-Strukturen müssen hier nicht nur als passive Datenspeicher fungieren, sondern aktiv die geforderten harmonisierten Softwarekomponenten integrieren: Die Interoperabilitätssoftware muss Daten im europäischen Austauschformat bereitstellen, während die Protokollierungskomponente jeden Zugriff mit präzisen Metadaten dokumentiert. Die bis März 2027 zu erlassenden Durchführungsrechtsakte bieten der FDO-Community ein konkretes Zeitfenster zur Mitgestaltung: Die technischen Spezifikationen für Datensätze, Kodiersysteme und sichere Verarbeitungsumgebungen können durch FDO-Prinzipien – maschinenlesbare Metadaten, persistente Identifikatoren, FAIR-konforme Qualitätsmetriken – maßgeblich geprägt werden.

Die Batterieverordnung stellt ab Februar 2027 mit dem digitalen Batteriepass andere, aber ebenso präzise Anforderungen. Für Industriebatterien über 2 kWh und Elektrofahrzeugbatterien müssen detaillierte Informationen zur stofflichen Zusammensetzung, zu gefährlichen Stoffen und zur Lieferkette dauerhaft – bis zum Recycling – verfügbar bleiben. Die entwickelten FDO-Asset-Administration-Shell-Transformationen schaffen hier strukturelle Voraussetzungen, stoßen aber an Grenzen: Die bloße Datentransformation genügt nicht den Anforderungen an Authentizität und Integrität nach Art. 78 BattVO.

Erforderlich sind verlustfreies Parsing der in Anhang XIII spezifizierten Pflichtangaben, manipulationssichere Integritätsnachweise und die Gewährleistung dauerhafter Verfügbarkeit auch bei Geschäftsaufgabe des Wirtschaftsakteurs. Diese regulatorischen Vorgaben treiben die Entwicklung kryptografisch gesicherter, dauerhaft verfügbarer FDO-Strukturen voran.

Beide Beispiele demonstrieren eindrücklich, wie regulatorische Anforderungen als Innovationsmotor wirken: Sie transformieren FAIR Digital Objects von einem abstrakten technischen Konzept zu konkreten Compliance-Instrumenten mit definierten Entwicklungspfaden. Die gestaffelten Einführungszeitpunkte – 2027 für die Batterieverordnung, 2028-2031 für EHDS-Kernbestimmungen – ermöglichen dabei eine iterative Weiterentwicklung, bei der technische Innovation und regulatorische Mitgestaltung Hand in Hand gehen.

INTERVIEW DAVID WAGNER, SPIRIT LEGAL

>>Welchen zentralen Mehrwert bieten FDOs für die Compliance mit europäischen Datenvorschriften?
<<FDOs schaffen eine einheitliche technische Basis, die gleichzeitig DSGVO-Rechenschaftspflichten, Data Act-Portabilitätsanforderungen und KI-VO-Dokumentationsvorgaben bedient. Diese Konvergenz reduziert den Implementierungsaufwand erheblich, da ein technischer Standard multiple regulatorische Anforderungen durch maschinenlesbare Metadaten und standardisierte Schnittstellen erfüllt.

>>Wo sehen Sie die größten rechtlichen Herausforderungen beim Einsatz von FDOs?
<<Die kryptografische Unveränderlichkeit von Nanopublications kollidiert fundamental mit DSGVO-Löschpflichten und DSA-Entfernungsanforderungen für rechtswidrige Inhalte. Diese technische Persistenz, die für Forschungsintegrität essentiell ist, steht im direkten Konflikt mit dem Recht auf Vergessenwerden und erfordert innovative Lösungsansätze oder regulatorische Klarstellungen.

>>Wie bewerten Sie FDOs im Kontext der KI-Verordnung?
<<FDOs unterstützen KI-Governance durch strukturierte Dokumentation von Trainingsdaten und transparente Provenienz-Nachweise, was besonders für Hochrisiko-Systeme relevant ist. Die Herausforderung liegt in der präzisen Systemabgrenzung einzelner KI-Komponenten innerhalb der FDO-Infrastruktur – ein Problem, das allerdings alle modularen Systeme betrifft, nicht nur FDOs.

>>Welche Empfehlung geben Sie Organisationen, die FDO-Infrastrukturen aufbauen wollen?
<<Organisationen sollten frühzeitig Legal Requirements Engineering mit interdisziplinären Teams aus Juristen, Technikern und Domänenexperten durchführen, um die spezifischen Use-Cases rechtssicher zu gestalten.

DIGITAL SERVICES ACT: ARCHITEKTURBEDINGTE REGULIERUNGSPFADE

Die Anwendbarkeit des Digital Services Act (DSA) auf FDO-Infrastrukturen hängt von deren technischer Ausgestaltung ab. Die Verordnung differenziert zwischen Hostingdiensten nach Artikel 16 ff. DSA und Online-Plattformen nach Artikel 19 ff. DSA mit unterschiedlichen Pflichtenkatalogen. Die gewählte Architektur determiniert die regulatorische Einordnung und damit den Umfang der Compliance-Anforderungen.

Nanopublikationen werfen derzeit noch spezifische Regulierungsfragen auf. Ihre kryptografisch gesicherte Unveränderlichkeit steht in Konflikt mit den Entfernungspflichten des DSA für rechtswidrige Inhalte. Technische Retraction-Mechanismen, die lediglich nachträgliche Widerrufe dokumentieren, genügen den rechtlichen Anforderungen nicht. Der DSA verlangt die tatsächliche Unzugänglichmachung, nicht bloß deren Kennzeichnung als ungültig.

Drei regulatorische Konstellationen sind zu unterscheiden: Geschlossene Systeme wie industrielle Handle-System-Konsortien können durch restriktive Zugangskontrolle, begrenzte Mitgliederzahl und qualifizierte Aufnahmeverfahren als reine Hostingdienste operieren. Die Pflichten beschränken sich auf die Entfernung rechtswidriger Inhalte nach Kenntniserlangung und die Einrichtung von Meldewegen.

Offene Wissenschaftsnetzwerke mit Nanopublikationen erfüllen hingegen regelmäßig die Kriterien einer Online-Plattform. Die dezentrale Replikation macht Inhalte öffentlich zugänglich, die automatische Verbreitung konstituiert eine aktive Rolle bei der Informationsvermittlung. Der erweiterte Pflichtenkatalog mit Beschwerdemanagementsystem, Begründungspflichten und Transparenzberichten ist bei unveränderlichen Datenstrukturen technisch noch nicht vollständig umsetzbar.

Mischformen aus Handle-System und Nanopublikationen in derselben Infrastruktur unterliegen der strengeren Qualifikation als Online-Plattform. Eine differenzierte regulatorische Behandlung ließe sich durch vollständige organisatorische und technische Trennung rechtssicher erreichen – separate juristische Personen, getrennte Betriebsteams und unterschiedliche technische Systeme.

Das Ergebnis ist ambivalent und aufschlussreich zugleich: FDOs bieten in vielen Bereichen innovative und solide technische Lösungen für komplexe regulatorische Anforderungen. Ihre strukturierten Profile, persistenten Identifikatoren und standardisierten Schnittstellen ermöglichen es, rechtliche Anforderungen wie Privacy by Design, Datenportabilität und transparente KI-Governance präzise und maschinenlesbar umzusetzen.

Gleichzeitig kollidieren zentrale Architekturmerkmale – insbesondere die derzeitige, grundsätzliche Unveränderlichkeit dezentral replizierter Nanopublikationen – mit rechtlichen Anforderungen, die darauf ausgerichtet sind, dass Daten nachträglich korrigiert oder gelöscht werden können.

Diese Reibungspunkte sind nicht per se ein Makel, sondern Beleg dafür, dass FDOs technologisches Neuland erschließen. Sie stellen Fragen, für die das Recht noch keine fertigen Antworten bereithält, und treiben damit die notwendige Weiterentwicklung sowohl der Technologie als auch ihrer rechtlichen Rahmenbedingungen voran. Wo bestehende Regelwerke nahtlos greifen, zeigen FDOs ihr Potenzial als Infrastruktur der nächsten Generation. Wo Konflikte entstehen, markieren sie den Raum für künftige Innovation.

FAZIT UND EMPFEHLUNGEN

RECHTLICHE MEHRWERTE VON FDOS

- **Compliance-Unterstützung:** Maschinenlesbare Metadaten unterstützen DSGVO-Rechenschaftspflichten, KI-VO-Dokumentationsanforderungen und DSA-Transparenzberichte durch einheitliche technische Mechanismen.

- **Native Interoperabilität:** Ein technischer Standard bedient simultan DSGVO-Datenportabilität, Data Act-Anforderungen, DGA-Datenräume und PSI-Weiterverwendung ohne regulatorische Fragmentierung.

REGULATORISCHE HERAUSFORDERUNGEN UND NOTWENDIGE WEITERE SCHRITTE

> *Unveränderlichkeit vs. Löschpflichten: Nanopublikationen kollidieren mit DSA-Entfernungspflichten für rechtswidrige Inhalte und DSGVO-Speicherbegrenzung, da ihre Konzeption derzeit keine nachträgliche Löschung ermöglicht. Technische Adaptionen sind hier nötig und möglich.*

Die konkrete Architekturentscheidung – etwa zwischen Handle-System oder Nanopublications – muss aus den funktionalen Anforderungen und regulatorischen Rahmenbedingungen des jeweiligen Anwendungsfalls abgeleitet werden, nicht umgekehrt.

> *KI-Systemabgrenzung: Die funktionale Qualifikation einzelner KI-Komponenten als eigenständige Systeme unter der KI-VO erfordert präzise technische Abgrenzung – eine generelle Herausforderung modularer Infrastrukturen, nicht FDO-spezifisch.*

> *Entwicklungen nachgelagerter Rechtsakte, bei deren Ausgestaltung man mitwirken kann und muss, sollten aufmerksam beobachtet und adressiert werden. Beispielsweise bieten die bis März 2027 zu erlassenden Durchführungsrechtsakte der FDO-Community ein konkretes Zeitfenster zur Mitgestaltung: Die technischen Spezifikationen für Datensätze, Kodiersysteme und sichere Verarbeitungsumgebungen können durch FDO-Prinzipien – maschinenlesbare Metadaten, persistente Identifikatoren, FAIR-konforme Qualitätsmetriken – maßgeblich geprägt werden.*

> *Ungeachtet dieser Rechtsakte bedarf es für jeden Anwendungsfall ein Legal Requirement Engineering.*

→ EU-Regulierungsrahmen

Das EU-Regulierungsrahmenwerk bildet das rechtliche Fundament für den gemeinsamen Binnenmarkt und umfasst eine Vielzahl von Verordnungen, Richtlinien und Standards. Es soll einheitliche Wettbewerbsbedingungen schaffen, Verbraucher schützen und Innovation fördern. Zentrale Säulen sind die Datenschutz-Grundverordnung (DSGVO), die europaweit den Umgang mit personenbezogenen Daten regelt, sowie umfangreiche Vorschriften in den Bereichen Produktsicherheit, Umweltschutz und Finanzmarktregulierung. In jüngster Zeit hat die EU ihren Fokus verstärkt auf die Regulierung digitaler Märkte gelegt, etwa durch den Digital Markets Act und den Digital Services Act, die großen Technologiekonzernen klare Regeln auferlegen. Auch im Bereich der künstlichen Intelligenz setzt die EU mit dem AI Act weltweit Maßstäbe für eine risikobasierte Regulierung.

FDO-Integration in die öffentliche Registerlandschaft

Da sich FDOs gut zum Management in Registerlandschaften eignen, wurde auch deren Potenzial für die Registermodernisierung der öffentlichen Verwaltung aus rechtlicher Sicht betrachtet, insbesondere unter Berücksichtigung des Staatsvertrages NOOTS, des Registermodernisierungsgesetzes (RegModG) und des Identifikationsnummerngesetzes (IDNrG).

¬ Die Untersuchung gliedert sich in drei Kernaspekte: die rechtliche Bewertung der FDO-Nutzung im öffentlichen Kontext hinsichtlich Datensicherheit, Interoperabilität und Datenschutz; die Prüfung der Rechtskonformität von FDOs bei registerübergreifender Datenverknüpfung und -nutzung, insbesondere unter Einbeziehung von Identifikationsnummern; sowie die Entwicklung konkreter Handlungsempfehlungen zur rechtssicheren Implementierung von FDOs in der Registerlandschaft, inklusive gesetzgeberischer Anpassungen.

HERAUSFORDERUNGEN DER DATENLANDSCHAFT

Insbesondere die öffentliche Verwaltung ist mit einer heterogenen und fragmentierten Datenlandschaft konfrontiert, die zu zeitaufwendigen und ineffizienten Suchprozessen führt. Unterschiedliche strukturelle und architektonische Beschaffenheiten der Datensätze verursachen Kompatibilitätsprobleme beim Austausch zwischen verschiedenen Datenräumen. Die föderale Organisation bedingt zudem eine dezentrale Führung öffentlicher Register, was die Abweichungen in Aufbau und Struktur der Datensätze vergrößert. Diese Defizite behindern eine einheitliche Digitalisierung und die effektive Umsetzung des Once-Only-Prinzips, da Datenbestände oft in abgeschotteten Silos liegen und interoperable Schnittstellen fehlen.

DAS POTENZIAL VON FAIR DIGITAL OBJECTS

FDOs bieten eine technische Grundlage für ein strukturiertes und verknüpftes Datenmanagement und unterstützen den Aufbau eines interoperablen Datenökosystems. Sie können in bestehende Systeme integriert werden, um Registereinträge strukturiert und eindeutig referenzierbar zu machen. FDOs sind von Haus aus interoperabel, wobei ihre Komplexität durch einfache Nutzungsoberflächen in der Software abgefangen werden muss. Aus vorhandenen Datensätzen lassen sich nach vordefinierten Mustern automatisch FDOs erzeugen, deren eindeutig zugewiesene Persistent Identifier eine registerübergreifende Weiterverarbeitung ermöglichen.

TECHNISCHE ASPEKTE UND SICHERUNGSMECHANISMEN

Die Architektur der FDOs erlaubt es, individuelle Zugriffs- und Nutzungsrechte in den Metadaten zu hinterlegen, wodurch die Zugriffsberechtigung automatisch überprüft und jeder Zugriff technisch protokolliert werden kann. Änderungen in den Metadaten können durch versionierte Ergänzungen dargestellt werden, ohne dass die ursprünglichen Daten überschrieben werden, was eine verbesserte ex-post-Kontrolle durch die Betroffenen gewährleistet. Die Datennutzlast verbleibt stets auf behördlichen Servern, und Metadaten werden ausschließlich in staatlich kontrollierten Infrastrukturen verarbeitet.

RECHTLICHER RAHMEN UND HERAUSFORDERUNGEN

FDOs fügen sich grundsätzlich in den bestehenden Rechtsrahmen ein. Die Integration der Personenidentifikationsnummer (PKZ) und anderer Identifikationsnummern in FDO-Profile ist technisch möglich, jedoch bedarf es einer Diskussion über den besten Weg auf Grundlage der rechtlichen Anforderungen. Das Registermodernisierungsgesetz (RegModG) schafft hierfür die gesetzliche Grundlage zur Verknüpfung räumlich getrennter Datenbestände über individuelle Registergrenzen hinweg, ohne alle Daten in einem zentralen Register zusammenzuführen.

DATENSCHUTZRECHTLICHE ASPEKTE UND PERSÖNLICHKEITSPROFILE

Die Zusammenführung von Datensätzen aus öffentlichen Registern birgt die Gefahr der Bildung umfassender Persönlichkeitsprofile. Das Bundesverfassungsgericht hat in seinem Volkszählungsurteil diesbezüglich bereits das Recht auf informationelle Selbstbestimmung (Art. 2 Abs. 1 i.V.m. Art 1 Abs. 1 GG) abgeleitet, das Schutz vor

INTERVIEW
PROF. DR. MARIO MARTINI,
Universität der Bundeswehr München

>>Welche konkrete Rolle spielt die Identifikationsnummer, wenn es um die registerübergreifende Verknüpfung von Daten geht?
<<Die IDNr ist der absolut zentrale Baustein der Registermodernisierung, da sie Personen und die dazugehörigen Daten schnell und vor allem eindeutig identifizieren soll. Der Gesetzgeber baut hierbei auf der bestehenden Steuer-identifikationsnummer (§139b AO) auf und nutzt diese als inhaltliche Erweiterung und zusätzliches Ordnungsmerkmal.

>>Wenn FAIR Digital Objects (FDOs) voraussichtlich eine tragende Komponente der Datenorganisation werden, warum wird dann im Gesetz über eine Verordnungsermächtigung nachgedacht, anstatt alles direkt festzulegen?

unbegrenzter Erhebung, Speicherung, Verwendung und Weitergabe personenbezogener Daten durch staatliche Stellen gewährt. Eine Überschreitung der Grenze zulässiger Datenzusammenführung liegt vor, wenn aus bestehenden Daten neue Rückschlüsse möglich werden und eine virtuelle Abbildung der Persönlichkeit entsteht.

MASSNAHMEN ZUR RISIKOBEGRENZUNG

Zum Schutz vor unzulässigen Datenzusammenführungen und der Bildung umfassender Persönlichkeitsprofile wird das 4-Corner-Model (§ 7 IDNrG) eingesetzt. Dieses Modell bindet zusätzlich zur empfangenden und übermittelnden Behörde zwei neutrale Intermediäre ein, die Ver- und Entschlüsselung sowie die Berechtigungsüberprüfung übernehmen. Die Identifikationsnummer spielt hierbei eine Schlüsselrolle als Verbindungselement. Zudem dient das bundesweite Datenschutzcockpit (§ 10 OZG) als eigenständiges Prüfinstrument zur Sicherstellung verfassungsrechtlicher Transparenz- und Sicherheitsvorgaben während des Datenaustauschs.

ANFORDERUNGEN AN EINE RECHTSSICHERE IMPLEMENTIERUNG

Für eine rechtssichere Implementierung von FDOs sind klare Rahmenbedingungen erforderlich, insbesondere bezüglich der Datenzusammenführung und der Einhaltung rechtlicher Schrankenbestimmungen. Dies erfordert eine bereichsspezifische gesetzliche Grundlage, die Anlass, Zweck und Grenzen der Datenverarbeitung klar normiert. Die Anforderungen an die Bestimmtheit der Ermächtigungsgrundlage steigen mit der Intensität des Eingriffs. Angesichts neuartiger technologischer Eingriffsformen wie FDOs bedarf es spezieller gesetzlicher Grundlagen, die Art, Umfang und Grenzen der Datenzusammenführung eindeutig regeln.

VERORDNUNGSERMÄCHTIGUNG UND DATENARCHITEKTUR

Weil FDOs wahrscheinlich eine tragende Komponente der Datenorganisation spielen werden, sollte in den hierfür maßgeblichen Gesetzen eine Verordnungsermächtigung vorgesehen werden, um technische Fragen der Implementierung zu klären. Die gesetzlich festzulegende FDO-Architektur sollte dem Prinzip der Datensparsamkeit folgen und sich durch bereichsspezifische Verarbeitung und technische Isolierung sensibler Daten auszeichnen. Hierfür ist eine Metadaten-

STRATEGISCHE BETRACHTUNGEN_FDO-INTEGRATION IN DIE ÖFFENTLICHE REGISTERLANDSCHAFT

<<Weil FDOs – den Wunsch ihrer Implementierung vorausgesetzt – wahrscheinlich eine tragende Komponente in dem „Wie“ der Datenorganisation spielen. Deshalb sollte in den hierfür maßgeblichen Gesetzen eine Verordnungsermächtigung vorgesehen werden, innerhalb derer alle rein technischen Fragen der Implementierung geklärt werden können.

>>Das 4-Corner-Modell soll die Datenzusammenführung kontrollieren; wie schützt dieses Konstrukt uns Bürger konkret vor der Bildung umfassender Persönlichkeitsprofile?
<<Das Modell ist ein zentraler Lösungsansatz, um unzulässige Zusammenführungen von Daten und die Entstehung umfassender Persönlichkeitsprofile zu verhindern. Es werden hierzu zwei neutrale Intermediäre in den Datenaustausch eingebunden, die sowohl die Berechtigungsprüfung der anfragenden Stelle

architektur zu schaffen, die eine klare Trennung von Stamm- und Fachdaten vorsieht und behördenübergreifende Beziehungsanalysen verhindert.

Zusammenfassend spiegelt die Implementierung von FDOs zur registerübergreifenden Verknüpfung von Daten den Zielkonflikt zwischen Verwaltungseffizienz und Rechtskonformität wider. Während FDOs eine leistungsfähigere und bürgerfreundlichere Verwaltung ermöglichen, droht gleichzeitig eine Aushöhlung des informationellen Selbstbestimmungsrechts und des datenschutzrechtlichen Zweckbindungsgrundsatzes. Bei restriktiver Ausgestaltung können FDOs jedoch als im Grundsatz rechtlich tragfähiger Kompromiss zwischen den Bedürfnissen der Verwaltungsmodernisierung und der Wahrung informationeller Selbstbestimmung angesehen werden.

FAZIT UND EMPFEHLUNGEN

POTENZIAL FÜR REGISTERMODERNISIERUNG

- **FDOs können in bestehende Systeme integriert werden, um Registereinträge strukturiert und eindeutig referenzierbar zu machen.**

- **Die Architektur der FDOs erlaubt es, individuelle Zugriffs- und Nutzungsrechte in den Metadaten zu hinterlegen und Zugriffsberechtigung automatisch zu überprüfen und protokollieren.**

- **Metadaten können durch versionierte Ergänzungen dargestellt werden, ohne dass die ursprünglichen Daten überschrieben werden, was eine verbesserte ex-post-Kontrolle durch die Betroffenen gewährleistet.**

- **Die Datennutzlast kann auf den jeweiligen behördlichen Servern verbleiben und Metadaten werden ausschließlich in staatlich kontrollierten Infrastrukturen verarbeitet.**

NOTWENDIGE WEITERE SCHRITTE

> *Die notwendige Software zur FDO-Erstellung sollte auf staatlichen Infrastrukturen laufen und von staatlichen Betreibern angeboten werden.*

> *Technische und organisatorische Maßnahmen müssen Spannungen mit dem Zweckbindungsgrundsatz einhegen und äquivalent zu denen bei der Einführung einer Personenidentifikationsnummer sein.*

als auch die Ver- und Entschlüsselung der Datensätze übernehmen.

>>FDOs sollen die Fragmentierung im Datenraum der öffentlichen Verwaltung überwinden; was ist die technische Vision dahinter?

<<FDOs legen eine technische Grundlage für ein strukturiertes und verknüpftes Datenmanagement sowie den Aufbau eines interoperablen Datenökosystems, da sie Registereinträge strukturiert und eindeutig referenzierbar machen. Durch ihre Interoperabilitäts-offenheit und die eindeutige PID können sie die Registermodernisierung technisch umsetzen und die behördenübergreifende Weiterverarbeitung der Daten weitgehend fehlerfrei ermöglichen.

> *Eine unabhängige Kontrollstelle sollte regelmäßig Protokolldaten evaluieren und auf unbefugte Zugriffe überprüfen.*

> *FDOs müssen durch eindeutige und dauerhafte Referenzen eine lückenlose Dokumentation (Datenherkunft, Verwendungszweck, beteiligte Stellen) ermöglichen.*

> *Eine Verordnungsermächtigung in den maßgeblichen Gesetzen ist nötig, um technische Fragen der Implementierung, wie die der Datenorganisation durch FDOs, nachträglich klären zu können.*

XFDO

Für einen sicheren und medienbruchfreien Datentransfer ist die Festlegung verbindlicher IT-Standards (Protokolle und Datenformate) zwischen Bund, Ländern und Kommunen unerlässlich. Die XÖV-Standards (XML in der öffentlichen Verwaltung) bieten hierbei eine einheitliche Entwicklungsumgebung und eine methodische Grundlage für den maschinenlesbaren Austausch von Daten, wobei der NOOTS-Standard (National Once-Only Technical System, DE-EDM-Standard) den technischen Transport der Fachinformationen regelt.

Das Konzept der Extended FAIR Digital Objects (XFDO) adressiert die Herausforderung, nationale deutsche Digitalisierungsstandards (wie XDatenfelder und BOB-Elemente) mit internationalen Interoperabilitätsprinzipien zu verbinden. Deutsche Verwaltungsformulare basieren auf bewährten XMLStrukturen des FIM-Portals(Föderales Informationsmanagement), während internationale Organisationen zunehmend auf JSON-basierte FDO-Objekte mit persistenten Identifikatoren setzen. XFDO schließt die Lücke zwischen den in deutschen Behörden verwendeten XML-Strukturen des FIM-Portals und den zunehmend global genutzten JSON-basierten FDO-Objekten.

Das XFDO-Transformations-Framework ist darauf ausgelegt, bestehende Verwaltungsdaten semantisch aufzuwerten, indem es automatisch „Baukasten-optimierte Bausteinelemente“ (BOB-Elemente) in Formularen erkennt und diese verlustfrei mit strukturierten Metadaten im Sinne der FAIR-Prinzipien anreichert. Diese Transformation führt von monolithischen Datenfeldern zu semantisch strukturierten Objekthierarchien, bspw. durch die Aufteilung von „Vor- und Nachname“ in separate, ontologisch validierte XFDOs.

Jedes transformierte XFDO-Objekt erhält einen persistenten Identifikator (PID) über das internationale Handle-System, was eine langfristige Referenzierbarkeit und eine präzise Versionskontrolle ermöglicht. Die FDO-Kompatibilität, die durch XFDO erreicht wird, könnte Deutschland zu einem bevorzugten Partner für grenzüberschreitende Verwaltungsverfahren machen und die Anbindung des deutschen NOOTS an das European-Once-Only-Technical-System (EU-OOTS) erleichtern.

Integration und Datensouveränität

FDOs können in die bestehenden Systeme integriert werden, um die Registermodernisierung technisch umzusetzen, indem sie Registereinträge strukturiert und eindeutig referenzierbar machen. Ein zentraler Vorteil von XFDO liegt in der Vermeidung multipler Definitionen identischer Datenstrukturen. Statt dass jede Behörde eigene Varianten von „Adresse“ oder „Personendaten“ entwickelt, entstehen zentrale FDO-Referenzen, die automatisch synchronisiert und validiert werden können.

XFDO-Objekte beinhalten explizite Referenzen auf die zugrundeliegenden rechtlichen Grundlagen. Änderungen in der Rechtslage können automatisch erkannt und entsprechende Anpassungen in abhängigen Formularen vorgeschlagen werden. Die theoretischen Grundlagen von XFDO sind mittlerweile in erste Tests überführt worden. Ein konkreter Workflow zur FDO-Generierung wurde entwickelt und erfolgreich getestet.

Weitere Informationen unter
https://www.lindner-consult.de//files/2025/08/xVermerk_XDFO_Marvin-Spyrka.pdf

STRATEGISCHE BETRACHTUNGEN_FDO-INTEGRATION IN DIE ÖFFENTLICHE REGISTERLANDSCHAFT

FDO-Anwendungsfall Nutzungsrechte-Management

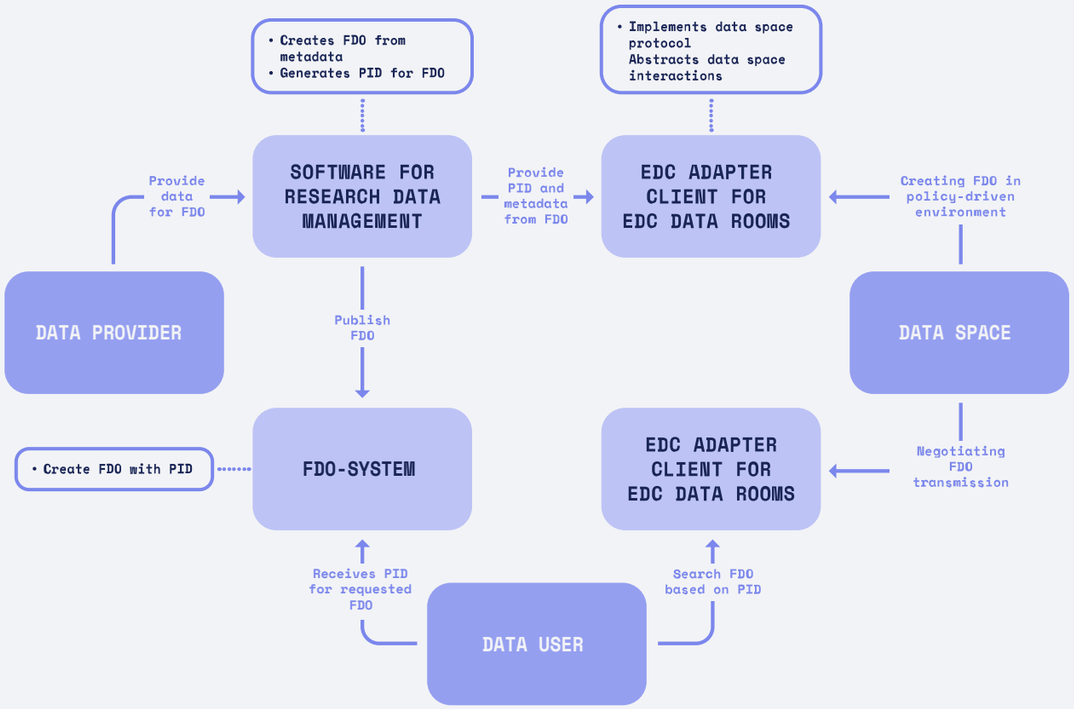
Die Verwaltung von Nutzungsrechten ist für den erfolgreichen und rechtskonformen Einsatz von FAIR Digital Objects (FDOs) in Datenökosystemen von entscheidender Bedeutung. Im Rahmen des Projekts FAIR Data Publisher wurden auch die rechtlichen und technischen Konzepte für die Verwaltung allgemeiner Nutzungsrechte und die Abbildung von Zuordnungsrechten (insbesondere Urheberrechte) beleuchtet. Dazu wurden Handlungsempfehlungen für den flächendeckenden Einsatz abgeleitet sowie die technische Abbildung des rechtlichen Rahmens ermöglicht.

Die rechtliche Zuordnung von FDO-Inhalten ist heterogen, wobei kein allgemeines Datenausschließlichkeitsrecht existiert. Ein FDO-Record enthält Metadaten, die auf die eigentlichen Datenobjekte referenzieren. Faktendaten (wie Messwerte, Fakten, wissenschaftliche Erkenntnisse oder Forschungsergebnisse als solche) sind frei von Urheberrecht. Das schränkt die Steuerungsmöglichkeiten offener Datenbestände ein, da nur die konkrete Darstellung und Form, nicht aber die Idee oder das Faktenwissen geschützt ist. Standardlizenzen wie Creative Commons (CC) entfalten ihre verbindliche Wirkung nur bei urheberrechtlich geschützten Inhalten und sind auf dieses Rechtsgebiet beschränkt. Für urheberrechtsfreie Inhalte gehen sie also ins Leere.

ENTWICKLUNG EINES KONZEPTS ZUR VERWALTUNG ALLGEMEINER NUTZUNGSRECHTE

Angesichts der begrenzten Reichweite des Urheberrechts ist die vertragliche Regelung von Nutzungsrechten das einzig tragfähige Instrument zur Steuerung der Zugänglichkeit und Nutzung von FDOs in geschlossenen Datenräumen. Durch die faktische Zugriffskontrolle – indem der Zugang zu den Inhaltsdaten von einem Vertragsabschluss abhängig gemacht wird – kann eine präzise Steuerung unabhängig vom gesetzlichen Schutzstatus erreicht werden. Dies ermöglicht eine souveräne Zugangskontrolle durch einen Vertrag. Die im Expert Group Report der EU vorgeschlagenen Mustervertragsklauseln (MCTs) können hierbei als konkrete juristische Spezifikation dienen.

Das technische Konzept nutzt die Open Digital Rights Language (ODRL) als maschinenlesbare Policy-Sprache, um vertragliche Vereinbarungen als technische Beschränkungen (Constraints) abzubilden. ODRL ermöglicht die Spezifikation von Richtlinien (Policies), die sowohl den gesetzlichen Schutzstatus als auch vertragliche Regelungen abbilden können, indem die grundlegenden Konzepte „Permission“, „Prohibition“ und „Duty“verwendet werden. ODRL-Profile müssen



•FDO-ÜBERTRAGUNGSVERHANDLUNG•

Interaktion zwischen den drei Hauptakteuren (Data Provider, FDO-System, Data User): Der EDC-Adapter fungiert als technische Abstraktionsschicht für Data Space-Protokolle. Die Policy-getriebene Umgebung stellt sicher, dass Nutzungsrechte bereits bei der FDO-Erstellung berücksichtigt werden, während die bidirektionale Verhandlung zwischen den Parteien eine flexible, aber kontrollierte Datennutzung ermöglicht. Der bidirektionale Workflow umfasst die FDO-Übertragungsverhandlung zwischen Data User und Data Provider sowie die Datenbereitstellung für FDOs. Diese Architektur gewährleistet eine standardisierte, Policy-konforme Verwaltung von Nutzungsrechten, indem sie technische Interoperabilität mit rechtlicher Compliance in dezentralen Datenräumen kombiniert und dabei die FAIR-Prinzipien durchgängig implementiert.

für spezifische Anwendungsfälle erweitert werden, um deren Anforderungen gerecht zu werden. Der FDO-Record fungiert dabei als zentraler Rechte-Hub für sämtliche Metainformationen zu den Nutzungsrechten.

ABBILDUNG VON URHEBERRECHTEN UND GESETZLICHEM STATUS

Neben der vertraglichen Regelung im Closedness-Konzept wurde beim Projekt FAIR Data Publisher auch die Abbildung der Zuordnungsrechte selbst in den Blick genommen. Zur Beschreibung des gesetzlichen Status nicht-lizenzierbarer Inhalte dienen Standards, die als deskriptive Attribute im FDO-Record hinterlegt werden. Die Public Domain Mark (PDM) kennzeichnet Inhalte, die frei von Urheberrechten und damit gemeinfrei sind. Sie dient als Attributions-Vokabular und Statusindikator ohne vertragliche Bindungswirkung. Ferner sind Rights Statements standardisierte Statusindikatoren, die primär für Kulturerbe-Institutionen entwickelt wurden, um den Urheberrechtsstatus digitaler Objekte abzubilden. Sie sind besonders für Situationen konzipiert, in denen der Rechtsstatus unklar oder komplex ist. Sie dienen als maschinenlesbare FDO-Attribute und ergänzen die Policy-Attribute wie ODRL.

Über das Urheberrecht hinaus müssen weitere Zuordnungsrechte wie Datenschutzrecht (DSGVO, relevant z. B. im Anwendungsfall Digitale Medizin), Patentrechte, Geschäftsgeheimnisse und arbeitsrechtliche Zuordnungen berücksichtigt werden. Diese nicht-absoluten Rechte wie Geschäftsgeheimnisse oder arbeitsrechtliche Zuordnungen entfalten nur zwischen den beteiligten Parteien Wirksamkeit und erfordern ebenfalls eine Steuerung über vertragliche Regelungen in geschlossenen Datenräumen, da sie keine Außenwirkung gegenüber Dritten entfalten.

Die technische Abbildung der Zuordnungsrechte im FDO-Record erfordert eine integrative Mapping-Strategie und die klare Unterscheidung zwischen deskriptiven Status-Attributen (PDM, Rights Statements) und operativen Policy-Attributen (ODRL).

TECHNISCHE IMPLEMENTIERUNG DES POLICY FRAMEWORKS

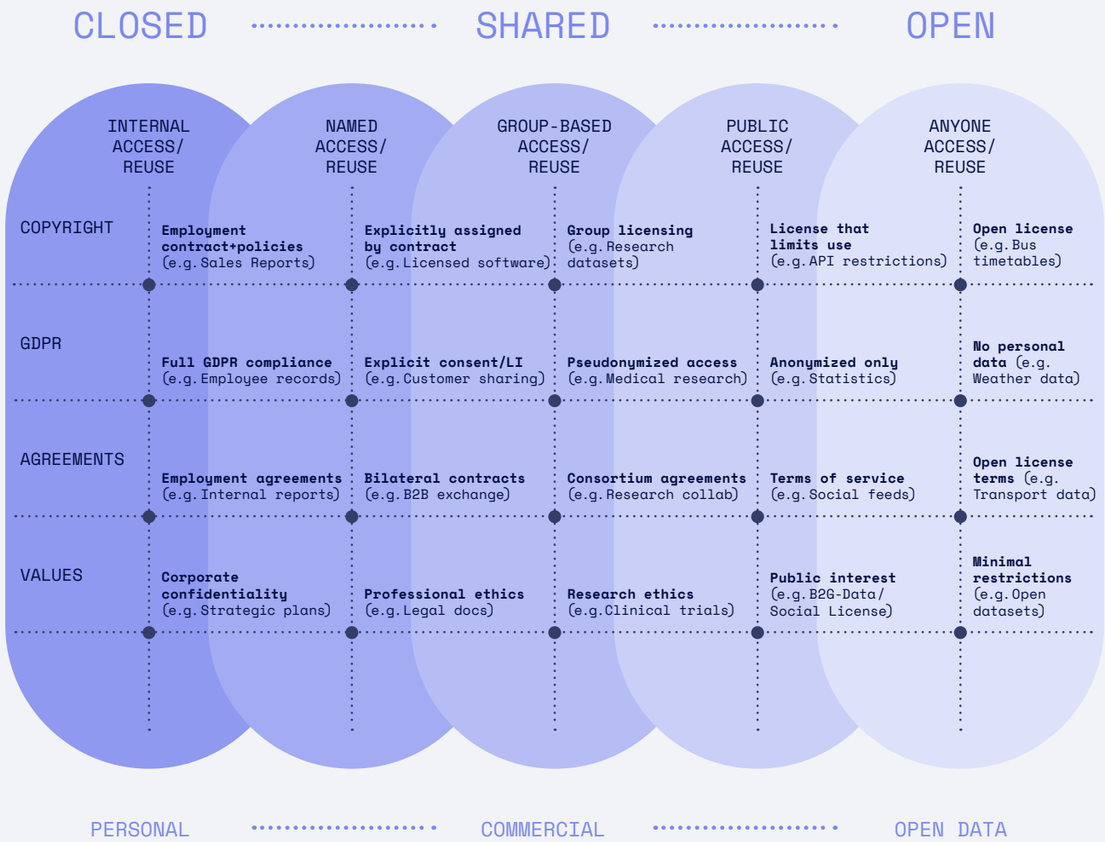
Die technische Durchsetzung der FDO-Richtlinien erfolgt über den Eclipse Dataspace Connector (EDC), der als Infrastruktur für Datenräume fungiert. Der EDC implementiert zentrale Datenraumprotokolle und bietet native Unterstützung für die Evaluierung von ODRL-Richtlinien. Der EDC überprüft vor jeder Datenübertragung – und bindet die Bewertung an die Vertragsverhandlungsphase – ob die ODRL-Policy-Bedingungen, die als Constraints im FDO-Record hinterlegt wurden, erfüllt sind. Nachdem übertragene Daten den Einflussbereich des Connectors verlassen haben, erfolgt keine technische Durchsetzung durch den EDC; die Einhaltung der Nutzungsbedingungen basiert dann auf vertraglichen und organisatorischen Maßnahmen. Dieser Ansatz nutzt die Service-Erweiterungsarchitektur des EDC, um ein eigenständiges Richtliniensystem zu vermeiden und die ausgereiften Funktionen des EDC zur Richtlinienbewertung zu nutzen.

Es wurden dazu vier verschiedene, konsistent implementierte Richtlinientypen entwickelt, um die wichtigsten Anforderungen aus den FDO-Anwendungsfällen abzudecken.

Erstens eine Richtlinie zur Datenzugriffskontrolle. Sie operationalisiert die grundlegende Unterscheidung zwischen offenem und geschlossenem FDO-Zugriff. Sie bewertet die Zugriffsebenen der User im Rahmen der Vertragsverhandlungen anhand der Datenzugriffsanforderungen. Die Implementierung umfasst eine Zugriffskontrollmatrix, in der User mit geschlossenem Zugriff volle Berechtigungen für offene und geschlossene Daten erhalten, während User offenem Zugriff nur auf offene Daten beschränkt sind.

→ Standard-Klauseln

Mustervertragsklauseln (Model Contractual Clauses, MCTs) sind standardisierte Vertragsvorlagen, die in verschiedenen Rechtsbereichen zur Vereinfachung und Vereinheitlichung von Vertragsbeziehungen eingesetzt werden. Sie bieten vordefinierte rechtliche Regelungen für häufig wiederkehrende Geschäftsvorfälle und reduzieren dadurch Verhandlungsaufwand, Rechtsunsicherheit und Transaktionskosten. Ein Beispiel sind die EU Standard Contractual Clauses (SCCs) für den rechtssicheren Transfer personenbezogener Daten aus der EU in Drittländer. Zu frühen und bewährten Mustervertragsklauseln zählen die bekannten Creative-Commons-Lizenzen. Auch sie bestehen aus standardisierten, modular aufgebauten Vertragsmustern, die rechtliche Komplexität reduzieren und Transaktionskosten senken. Im digitalen Zeitalter gewinnen MCTs zusätzliche Relevanz durch die Möglichkeit ihrer Digitalisierung.



•POLICY SPEKTRUM•

Die Freigabe von Daten – das Datasharing – ist kein einheitlicher Prozess, sondern bewegt sich auf einem Spektrum von strenger interner Kontrolle bis hin zu offener, uneingeschränkter Nutzung. Die Positionen eines Datensatzes auf diesem Spektrum werden durch eine Kombination aus rechtlichen Mechanismen und organisatorischen Werten bestimmt. Ein klares Verständnis dieser Mechanismen ist essenziell für die Implementierung in modernen Datenräumen (z. B. durch ODRL-Regelwerke oder FDO-Metadaten).

STRATEGISCHE BETRACHTUNGEN_FDO-ANWENDUNGSFALL NUTZUNGSRECHTE-MANAGEMENT

INTERVIEW
FABIAN RACK,
FIZ KARLSRUHE
DAHAM MOHAMMED MUSTAFA,
FRAUNHOFER FIT

>>Inwieweit spielen vertragliche Nutzungsrechte in geschlossenen Datenräumen eine Rolle?
<<Verträge sind das zentrale Steuerungsinstrument für die Nutzung von FDOs. Da die meisten industriellen und Forschungsdaten urheberrechtsfrei sind (Faktenwissen), ist die faktische Zugriffskontrolle durch einen vorangehenden Vertrag der einzig tragfähige Weg, Nutzungsbedingungen festzulegen. Der spannende Teil ist die technische Abbildung dieser Verträge mittels des maschinenlesbaren ODRL-Vokabulars (Open Digital Rights Language).

>>Hilft das Urheberrecht, die Nutzung von FDOs zu steuern, und welche Rolle spielen CC-Lizenzen?
<<Das Urheberrecht hilft nur sehr bedingt, da die meisten relevanten Inhalte, wie Messdaten oder Faktenwissen, urheberrechtsfrei sind. Die Gestaltungsmöglichkeit durch Standardlizenzen wie Creative Commons (CC) ist folglich stark eingeschränkt. CC-Lizenzen sind reine Urheberrechtslizenzen, deren Bindungswirkung vom tatsächlichen Bestehen eines Urheberrechts abhängt.

>>Welche weiteren Instrumente gibt es außerhalb von Verträgen, um Nutzungen, insbesondere im Kontext des maschinellen Lernens, zu regeln?
<<Derzeit werden im Kontext von machine reuse und KI-Training sogenannte Preference Signals entwickelt, wie die IETF AI Preferences und die CC Preference Signals. Diese dienen als maschinenlesbare technische Richtlinien, um Präferenzen für oder gegen KI-Training zu setzen. Die juristische Bewertung ihrer Bindungswirkung bei frei zugänglichen, urheberrechtsfreien Objekten ist Gegenstand zukünftiger Diskussionen – ebenfalls ihr Einsatz in einem geschlossenen Datenraum.

Zweitens eine Richtlinie zur geografischen Zugriffskontrolle, die sich mit standortbasierten Zugriffsbeschränkungen zur Einhaltung gesetzlicher Vorschriften, insbesondere der DSGVO-Anforderungen im Anwendungsfall Digitale Medizin befasst. Die Richtlinie wertet die Standortangaben der Teilnehmerinnen und Teilnehmer anhand festgelegter geografischer Anforderungen aus und unterstützt Operatoren wie exakte Übereinstimmung, Ausschlussregeln und Berechtigungen für mehrere Regionen.

Eine rollenbasierte Datenzugriffsrichtlinie unterscheidet als dritte Steuerung zwischen verschiedenen Datentypen und Benutzerrollen. Sie erstellt Zugriffsmatrizen, die definieren, welche Benutzerrollen auf welche Datentypen zugreifen können, und schützt beispielsweise kommerzielle Daten vor unbefugter Nutzung oder personenbezogene Daten aus Datenschutzgründen.

Eine vierte, zeitliche, Zugriffskontrollrichtlinie ermöglicht begrenzte Datenzugriffsvereinbarungen für Testzugriffszeiträume, abonnementbasierten Datenaustausch und zeitlich begrenzte Forschungskooperationen. Die Implementierung umfasst Funktionen zum Vertragslebenszyklusmanagement, die Ablaufdaten überwachen und nach Ablauf entsprechende Maßnahmen auslösen können.

Alle implementierten Richtlinien nutzen standardisierte ODRL-Constraint-Vokabulare und folgen einer konsistenten Service-Erweiterungsarchitektur, die das systematische Hinzufügen neuer Richtlinientypen ermöglicht und somit die Erweiterbarkeit des Frameworks sicherstellt.

VALIDIERUNG UND ZUKÜNFTIGE ARBEITEN

Die Validierung des Frameworks erfolgte durch umfassende Tests aller implementierten Richtlinientypen in einer Multi-Connector-Testumgebung, die realistische Datenrauminteraktionen simulierte. Die Testmethodik umfasste die funktionale Verifizierung der Richtlinienlogik und die operative Validierung des gesamten Systems über den gesamten Lebenszyklus der Datenfreigabe von der Vertragsverhandlung bis zum Abschluss der Datenübertragung hinweg. Das Test-Framework validierte alle möglichen Kombinationen von Nutzerberechtigungsstufen und Datenzugriffsanforderungen, inklusive richtlinienübergreifender Interaktion (z. B. geografische und rollenbasierte Richtlinien gleichzeitig). Detaillierte Protokollierung und Live-Verfolgung der Richtlinienbewertung sorgten für Transparenz bei Zugriffsentscheidungen und Compliance-Audits.

Zukünftige Arbeiten konzentrieren sich auf die Stärkung der Robustheit. So könnte eine Infrastruktur zur Geolokalisierungsverifizierung zur Ablösung der aktuellen Selbstangaben der Teilnehmerinnen und Teilnehmer durch die Integration mit externen Geolokalisierungsdiensten entwickelt werden.

Die entwickelten Konzepte und deren Implementierung im FAIR Data Publisher Projekt tragen maßgeblich dazu bei, die Verwaltung von Nutzungsrechten und die Abbildung von Urheberrechten in einer zunehmend komplexen und vernetzten Datenökonomie zu standardisieren und zu automatisieren.

FAZIT UND EMPFEHLUNGEN

MEHRWERTE

+ **Ermöglichung der automatisierten Rechtsverwaltung:** Rechtliche Aspekte (wie Nutzungsrichtlinien, Regeln und Pflichten) können als maschineninterpretierbare Metadaten im FDO ausgedrückt werden, was eine automatisierte Richtliniendurchsetzung und Lizenzierung (z. B. für KI-Training) ermöglicht.

+ **Sichere Handhabung des gesamten Spektrums der Datennutzung:** Das Konzept unterstützt die Verwaltung von Daten über das gesamte Spektrum hinweg – von sehr offenen/gemeinfreien Daten bis hin zu sehr restriktiven und proprietären Daten – und hilft somit, die Abstufungen der Offenheit und Geschlossenheit zu verwalten.

+ **Gestärkte Datensouveränität durch vertragliche Regelungen:** Da Standardlizenzen (wie CC) oft nicht auf reines Faktenwissen (das nicht urheberrechtlich geschützt ist) anwendbar sind, werden individuelle Nutzungsvereinbarungen (Data Usage Agreements, DUAs) zwischen Parteien ermöglicht und durch Sprachen wie ODRL technisch umgesetzt, um Datensouveränität zu gewährleisten.

+ **Gewährleistung der Interoperabilität in Datenräumen:** Durch die Verwendung standardbasierter, maschinenlesbarer Policy-Sprachen wie ODRL trägt das Konzept zur Interoperabilität über verschiedene Jurisdiktionen hinweg bei und unterstützt die Anforderungen des EU Data Act.

STRATEGISCHE BETRACHTUNGEN_FDO-ANWENDUNGSFALL NUTZUNGSRECHTE-MANAGEMENT

>>Warum ist der Eclipse Dataspace Connector (EDC) für die Durchsetzung der FDO-Richtlinien unverzichtbar und welche Rolle spielt dabei ODRL?
<<Der Eclipse Dataspace Connector (EDC) ist die technische Infrastruktur für den sicheren, souveränen Datenaustausch und somit der zentrale Durchsetzer (Enforcer) der Richtlinien. Der EDC nutzt die ODRL-Policy-Sprache, um die im FDO-Record hinterlegten vertraglichen Vereinbarungen als maschinenlesbare Constraints (Beschränkungen) zu interpretieren. Entscheidend ist, dass der EDC diese Policy-Bedingungen vor jeder Datenübertragung evaluiert und dadurch die Einhaltung der vereinbarten Nutzungsrechte automatisiert und gewährleistet – bis zu dem Zeitpunkt, an dem die Daten den Connector verlassen.

+ **Technische Erleichterung des Rechtsprozesses für Datenkonsumenten:** Durch die Darstellung rechtlicher Informationen als maschineninterpretierbare Metadaten im FDO kann der Empfänger der Daten die rechtlichen Bedingungen sofort verstehen und weiterverwenden, ohne den rechtlichen Prozess manuell durchlaufen zu müssen.

+ **Umfassende Berücksichtigung verschiedener Schutzrechte:** Das Konzept geht über das reine Urheberrecht hinaus und integriert weitere Zuteilungsrechte (wie Patentrechte, Geschäftsgeheimnisse und Datenschutzanforderungen) durch die Einbeziehung anderer Vokabulare wie dem Data Protection Vocabulary (DPV), um alle relevanten Constraints (COPYRIGHT, GDPR, AGREEMENTS, VALUES) abzudecken.

NOTWENDIGE WEITERE SCHRITTE

> *Stärkung der Robustheit: Eine Infrastruktur zur Geolokalisierungsverifizierung zur Ablösung der aktuellen Selbstangaben der Teilnehmerinnen und Teilnehmer könnten durch die Integration mit externen Geolokalisierungsdiensten entwickelt werden.*

Verweise Nutzungsrechte:
↗ ODRL-Modell Version 2.2 (W3C Recommendation): Open Digital Rights Language
↗ Data Protection Vocabulary Version 2.2 (DPV): Für Datenschutz-Aspekte
↗ SAP/OData Vokabulare für Personendaten:
↗ GitHub-Repository des DRK Information Model (Visual Models): Beispiele für visuelle Modelle im Kulturdatenraum
↗ IDS Knowledge Base – Usage Control/ Contract: Zum Datenraumprotokoll
↗ Fraunhofer IESE Blog – Data Sovereignty: Technologieunabhängige Spezifikation mit ODRL
↗ Fraunhofer IESE – Informationelle Selbstbestimmung mit MyData: Zum Datenschutz

Ökonomische Betrachtungen

Ökonomische Auswirkungen einer FDO-basierten Datenökonomie

¬ FAIR Digital Objects (FDO) können als standardisierte Architektur wesentlich zum wirtschaftlichen Wachstum beitragen. Durch die Nutzung persistenter Identifikatoren und standardisierter Metadaten entsteht eine robuste Infrastruktur, die den Austausch von Daten und Wissen deutlich effizienter macht. Einheitliche Standards senken zugleich die Transaktionskosten, da Zugriffs- und Lizenzierungsprozesse transparent, automatisiert und rechtssicher erfolgen können.

Die dadurch gewährleistete Interoperabilität erleichtert die Integration neuer Innovationen in bestehende Systeme und unterstützt die Entstehung von Folgeinnovationen. Auf dieser Basis können Unternehmen neue datengetriebene Geschäftsmodelle entwickeln und ihre Markteintrittsbarrieren senken. Schließlich kann man argumentieren, dass die internationale Anschlussfähigkeit und die entstehenden Netzwerkeffekte der FDO-Standards zu einer nachhaltigen Stärkung des langfristigen Innovations- und Wirtschaftswachstums führen.

Vor diesem Hintergrund wird deutlich, dass auch standardessentielle Patente (SEPs) eine zentrale Rolle für die Wachstumsdynamik moderner Volkswirtschaften spielen. Die zunehmende Bedeutung von SEPs prägt maßgeblich das langfristige wirtschaftliche Wachstum. Insbesondere in technologieintensiven Sektoren wie Telekommunikation, Elektronik und IT-Systemen gewinnen SEPs an Bedeutung, verstärkt durch die zunehmende Komplexität und Modularisierung technischer Produkte. Unternehmen setzen auf standardisierte Technologien, um Innovationen effizient zu integrieren und Marktkompatibilität sicherzustellen. Standardisierung schafft somit Plattformen für Folgeinnovationen, welche langfristig das Wirtschaftswachstum unterstützen können.

Das makroökonomische Modell der Studie analysiert die Wechselwirkung zwischen patentierten Innovationen und SEPs sowie deren Einfluss auf die Innovations- und Wachstumsdynamik einer Volkswirtschaft. Nicht jedes Patent entwickelt sich zwangsläufig zum Standard; vielmehr erfolgt Standardisierung zufällig und ist ex ante nicht präzise vorhersehbar. Langfristig wachsen Patente und SEPs parallel, wodurch ein stabiles Gleichgewicht ermöglicht wird. Zwar können standardisierte Patente Folgeinnovationen fördern, dies geschieht jedoch nur unter spezifischen Bedingungen.

Im vorliegenden Projekt wird ein makroökonomisches, dynamisches allgemeines Gleichgewichtsmodell herangezogen, um die dynamischen

INTERVIEW
TORBEN KLARL,
UNIVERSITÄT BREMEN

>>Erzählen Sie uns kurz, was genau das makroökonomische Modell in Ihrer Studie eigentlich untersucht.
<<Wir nutzen ein dynamisches allgemeines Gleichgewichtsmodell, um zu sehen, wie patentierte Innovationen und sogenannte standardessenzielle Patente (SEPs) die Innovations- und Wachstumsdynamik einer Volkswirtschaft beeinflussen. Dieses Modell bildet detailliert ab, wie der Forschungs- und Entwicklungssektor mit den Zwischen- und Endgutsektoren interagiert und welche mittel- und langfristigen Wachstumseffekte sich daraus ergeben.

>>Wie können Regulierungen wie FRAND helfen, wirklich kooperative Strukturen zu schaffen?
<<FRAND steht für faire, angemessene und nicht-diskriminierende Bedingungen und sorgt bei freiwilligen Lizenzierungen für deutlich mehr Planbarkeit und Transparenz. Genau diese verbesserten Bedingungen fördern die Innovationsbereitschaft und schaffen die nötigen kooperativen Strukturen, was im Gegensatz zu verpflichtenden Lizenzierungen steht, die oft innovationshemmend wirken. Unser Modell zeigt insbesondere, dass FRAND so ausgestaltet sein muss, dass die Produktivität von Standards erhöht wird. Nur so wächst der Forschungs- und Entwicklungssektor, was wiederum die Wachstumsraten im Übergang steigen lässt. Es kommt daher auf eine zielgerichtete FRAND Regulierung an.

>>Welchen konkreten Mehrwert sehen Sie bei FDOs, besonders in Verbindung mit Lizenzierungen?
<<Die Kombination von FDOs und FRAND-Bedingungen bringt zusätzliche Vorteile, weil FDOs einen effizienten Zugang zu Daten und Lizenzen ermöglichen. Das passiert dadurch, dass sie mit persistenten Identifikatoren und standardisierten Metadaten arbeiten, was die Prozesse stark vereinfacht. Dieser Zugang reduziert volkswirtschaftliche Kosten und wirkt produktivitätssteigernd. Dies ist vermutlich auch

Interdependenzen zwischen dem Forschungs- und Entwicklungssektor (F&E), in dem neue Designs entstehen (die über Patentierungen neue Industriestandards werden können), und dem Zwischen- und Endgutsektor einerseits sowie die sich daraus ergebenden mittel- und langfristigen Wachstumseffekte von Standardisierung und möglicher Regulierung andererseits abzubilden. Der Forschungs- und Entwicklungssektor produziert mit Wissen und Humankapital (Forschern) neue Designs, welche patentiert werden können. Im Lebenszyklus eines Patentes können nun zwei Ereignisse eintreten: Ein Patent kann durch die Wahl einer Behörde zu einem neuen Standard erhoben werden, was andere, konkurrierende Standards verdrängt, oder es wird selbst vom Markt verdrängt. Durch die Verdrängung von konkurrierenden Standards kommt dem Standardinhaber eine Monopolmacht zu. Diese kann nun über Zwangslizenzierung oder FRAND (Fair, Reasonable, and Non-Discriminatory) seitens des Staates reguliert werden.

Ein wesentlicher Aspekt des Modells betrifft den inhärenten Trade-off der Standardisierung: Unternehmen, die Patente entwickeln, tragen im Voraus das Risiko, den Wettbewerb um den Standard zu verlieren. Die Möglichkeit, die Monopolstellung zu verlieren, reduziert ex ante die Investitionsbereitschaft in Innovationsaktivitäten. Selbst wenn Standards positive Produktivitätseffekte erzielen, werden diese Vorteile häufig durch die stärkere Diskontierung zukünftiger Monopolgewinne im Vergleich zu unmittelbaren Produktivitätsgewinnen abgeschwächt. Folglich profitieren oft weniger innovative Sektoren stärker von der Standardisierung als forschungsintensive Unternehmen, was langfristig Innovationsanreize reduziert.

Die ökonomischen Charakteristika von Standards liefern weitere zentrale Einsichten. Einerseits erhöhen Standards durch positive Netzwerkexternalitäten, Skaleneffekte und verbesserte Kompatibilität die Marktattraktivität und den Wert der beteiligten Technologien. Andererseits können sie den Wettbewerb verschärfen und deutliche Gewinner-Verlierer-Strukturen erzeugen. Der letztliche wirtschaftliche Effekt hängt maßgeblich davon ab, ob die Vorteile der Standardisierung die Risiken der Wettbewerbsverdrängung überwiegen.

Diese Erkenntnisse verdeutlichen erhebliche politische Herausforderungen, insbesondere bei der Regulierung standardessentieller Patente. Verpflichtende Lizenzierungen (Compulsory Licensing) können kurzfristig Marktversagen und Monopolmissbrauch adressieren, wirken jedoch langfristig innovationshemmend. Unternehmen reduzieren ihre Innovationsaktivitäten, wenn regulatorische Maßnahmen potenzielle Gewinne einschränken. Solche restriktiven Maßnahmen wirken daher langfristig wachstumshemmend.

Im Gegensatz dazu bieten freiwillige Lizenzierungen auf Basis von FRAND-Bedingungen eine effektivere Lösung. FRAND-Regelungen verbessern Planbarkeit und Transparenz der Lizenzierung und fördern dadurch kooperative Strukturen und Innovationsbereitschaft. Eine Integration von FRAND und FAIR Digital Objects (FDOs) bringt zusätzliche Vorteile, da FDOs mittels persistenter Identifikatoren (PIDs) und standardisierter Metadaten effizienten Zugang zu Daten und Lizenzen ermöglichen.

Jedoch ist die Einbindung rechtlicher Rahmenbedingungen, insbesondere im Bereich des Datenschutzes, nicht zwangsläufig wachstumsfördernd. Während die Datenschutzgrundverordnung (DSGVO), der Data Act und der Data Governance Act Rechtssicherheit und Vorhersehbarkeit erhöhen, bergen sie zugleich Risiken einer Verringerung von Innovationsanreizen. Datenschutzregulierungen können den Zugang zu relevanten Daten erschweren, Compliance-Kosten erhöhen und damit datengetriebene Innovationen potenziell bremsen. Ihre wachstumsfördernde Wirkung hängt maßgeblich von einer ausgewogenen Implementierung ab, die regulatorische Anforderungen und Innovationsfreiheit angemessen verbindet.

Für politische Entscheidungsträger folgt daraus, dass ausschließlich restriktive Maßnahmen wie verpflichtende Lizenzierungen oder strikte Datenschutzanforderungen nicht ausreichend sind, um langfristiges Wachstum zu gewährleisten. Ergänzende innovationsfördernde Maßnahmen, beispielsweise steuerliche Anreize oder gezielte Forschungsförderungen für Unternehmen, die offene Standards unterstützen, sind von zentraler Bedeutung. Initiativen der Europäischen Union, wie der Digital Markets Act und die SEP-Verordnung, verfolgen bereits diesen integrierten Ansatz.

Zusammenfassend zeigt die Analyse, dass die Regulierung von SEPs und Standards nicht nur technologische, sondern insbesondere auch wirtschaftspolitische Herausforderungen beinhaltet. Die angemessene Balance zwischen regulatorischen Maßnahmen, einschließlich Datenschutz, und Innovationsförderung bestimmt entscheidend das langfristige Innovations- und Wachstumspotenzial Europas.

STRATEGISCHE BETRACHTUNGEN_ÖKONOMISCHE BETRACHTUNGEN

für den Forschungs- und Entwicklungssektor relevant, da man davon ausgehen kann, dass künftig mehr Daten in diesem Bereich eingesetzt werden. Dies wiederum kann allerdings zu einer Verdrängung von F&E Personal führen, was ohne weitere Regulierung wiederum wohlfahrtsreduzierend wirken kann. Ohne eine effiziente und zielgerichtete Regulierung geht es also nicht.

>>Was ist die größte Herausforderung für Entscheidungsträger, wenn sie langfristig das Innovationspotenzial Europas sichern wollen?
<<Der Knackpunkt ist, die angemessene Balance zwischen strenger Regulierung und der notwendigen Innovationsförderung zu finden. Nur restriktive Maßnahmen wie verpflichtende Lizenzierungen oder strikte Datenschutzanforderungen reichen nicht aus, um langfristiges Wachstum zu gewährleisten, weshalb ergänzende Anreize wichtig sind.

FAZIT UND EMPFEHLUNGEN

NOTWENDIGE WEITERE SCHRITTE

> *Der Modellrahmen liefert die Grundlage für eine tiefergehende Analyse von FDOs. Für eine fundierte wirtschaftspolitische Bewertung wären empirische Komponenten, konkretere Quantifizierungen und eine stärkere Integration der FDO-spezifischen Eigenschaften erforderlich. Dies erfordert allerdings eine fundierte Datenlage um kausale Effekte von FDOs auf Wachstum ableiten zu können.*

→ Catena-X

Catena-X ist ein kollaborativer Datenraum für die Automobilindustrie, um Daten unter Wahrung der Datensouveränität auszutauschen. Das Projekt wurde 2021 als Reaktion auf die Anforderungen der Kreislaufwirtschaft ins Leben gerufen und setzt technisch auf offene Standards und Schnittstellen, die Interoperabilität gewährleisten und gleichzeitig auf bestehenden Konzepten wie GAIA-X und dem International Data Spaces Association (IDSA) Connector aufbauen. Herausforderungen stellen sich jedoch bei der Überzeugung kleinerer und mittlerer Zulieferer zur Teilnahme. Catena-X repräsentiert einen Paradigmenwechsel hin zu kollaborativen Datenökonomien und könnte als Blaupause für andere Industrien dienen.

Betrieb einer FDO-Service-Architektur

¬ Betriebs- und Geschäftsmodelle bilden das Fundament für die wirtschaftliche Tragfähigkeit von Organisationen. Damit Organisationen in digitalen Ökosystemen zusammenarbeiten und Daten souverän austauschen können, benötigen Datenräume klare „Spielregeln“ und Rahmenwerke. Diese ermöglichen vertrauenswürdige Kollaborationen zwischen Unternehmen und bilden das organisatorische und technische Fundament für den Datenaustausch. Besonders bei Datenräumen erweist sich die Entwicklung resilienter Betriebskonzepte als relevant, da es sich um einen föderierten Ansatz mit heterogener Teilnehmerschaft und vielfältigen Interessen handelt.

Im Gegensatz zu traditionellen Ansätzen, wie beispielsweise zentralen Datenplattformen, müssen Datenräume die Interessen verschiedener Stakeholder ausbalancieren und gleichzeitig einen Mehrwert für alle generieren. Branchenspezifische Unterschiede können dabei ebenfalls eine Rolle spielen. Diese neue Form der kollaborativen Wertschöpfung ermöglicht zudem neue Anwendungsszenarien, die einzelne Akteure nicht realisieren könnten.

SPEZIALISIERTE GESCHÄFTSMODELLE FÜR JEDEN AKTEUR

Das Betriebsmodell eines Datenraums zielt auf die langfristige Gewährleistung des operativen Betriebs ab. Die Wahl der Betreiberorganisation kann je nach Datenraum erheblich variieren. Praxisbeispiele zeigen verschiedene Ansätze: Der Mobility Data Space (MDS) wird beispielsweise von einer gemeinnützigen GmbH betrieben, während Catena-X mit Cofinity-X eine kommerzielle Betreibergesellschaft etabliert hat. Alternativ können Vereine, öffentliche Akteure oder Konsortien als Betreiber fungieren. Die gewählte Rechtsform beeinflusst dabei sowohl die Governance-Struktur als auch die Vertrauenswürdigkeit des Datenraums.

Datenräume können unterschiedliche Servicestrukturen aufweisen, von zentralisiert bis hin zu dezentralisiert, wobei die technische Architektur die Verteilung von Verantwortlichkeiten und Kosten bestimmt. Zentralisierte Ansätze ermöglichen Skaleneffekte durch die Betreibergesellschaft, während dezentrale Architekturen den Teilnehmenden eine höhere Autonomie bieten und den Datenraum robuster und unabhängiger gestalten. Betreiber von Datenräumen stehen zudem vor einer „Make or Buy“-Entscheidung hinsichtlich der Bereitstellung der technischen Infrastruktur. Viele Datenräume basieren auf den Standards der International Data Spaces Association (IDSA) und nutzen bereits bestehende, IDSA-konforme Komponenten wie die Eclipse Dataspace Components. Die Wiederverwendung etablierter Komponenten reduziert nicht nur Entwicklungskosten, sondern erhöht auch die Interoperabilität.

Die Kostenstruktur eines Datenraums umfasst initiale Aufbaukosten und laufende Betriebskosten. Zu den Aufbaukosten zählen die Entwicklung der technischen Infrastruktur, die Etablierung von Governance-Strukturen und der Ausbau des Netzwerks. Laufende Kosten entstehen durch den Betrieb der Infrastruktur, diverse Support-Services und die Weiterentwicklung. Die initiale Finanzierung erfolgt häufig durch öffentliche Förderungen oder die Betreibergesellschaft, sollte jedoch langfristig durch ein sich selbst finanzierendes Erlösmodell gesichert werden. Zur Deckung der Betriebskosten existieren verschiedene Erlösmodelle, wie Mitgliedsbeiträge, „Pay per Use“-Modelle, Premium-Services, Zertifizierungsgebühren oder Transaktionsgebühren. Es ist essentiell, dass das Erlösmodell eine Balance zwischen finanzieller Nachhaltigkeit und niedrigen Eintrittsbarrieren für neue Teilnehmende wahrt.

Datenräume weisen eine klare Rollenverteilung auf. Datenanbieter stellen die Daten zur Verfügung, Datennutzerinnen und -nutzer beziehen diese für ihre Anwendungsfälle, Service Provider bieten technische Dienste wie Konnektoren an, und die Data Space Authority übernimmt Governance-Funktionen und stellt Kernservices wie das Identitätsmanagement bereit. Diese klare Rollendifferenzierung ermöglicht spezialisierte Geschäftsmodelle für jeden Akteur und eine auf den Datenraum zugeschnittene Vergütungsstruktur.

CHARAKTERISTIKA ETABLIERTER DATENRÄUME

Catena-X stellt das erste offene und kollaborative Datenökosystem für die Automobilindustrie dar und schafft einen vertrauenswürdigen, kollaborativen und sicheren Datenraum zur Ermöglichung einer datengetriebenen Wertschöpfungskette. Das Ökosystem basiert auf den Konzepten von Gaia-X und der IDSA und erweitert diese um automobilspezifische Anforderungen für Compliance, Interoperabilität und Sicherheit. Alle Teilnehmenden können in geschäftsprozesszentrierten Wertschöpfungsketten verbunden werden, wobei jeder auf einem gleichberechtigten

Spielfeld operiert, souveräne Kontrolle über seine Daten behält und keine Lock-in-Effekte auftreten.

Der Mobility Data Space (MDS) entwickelt sich als offener Datenraum für die souveräne und plattformübergreifende Bewirtschaftung von Mobilitätsdaten und basiert auf der dezentralen Systemarchitektur der IDSA. Das Konzept ermöglicht es Datengebern festzulegen und zu kontrollieren, unter welchen Bedingungen ihre Daten durch andere Akteure genutzt werden, wodurch Datensouveränität und Vertrauen geschaffen werden. Die Verknüpfung von öffentlichen und privatwirtschaftlichen Daten über regionale und nationale Datenplattformen transformiert den MDS zu einem digitalen Vertriebskanal für datengetriebene Geschäftsmodelle.

Aus den untersuchten unterschiedlichen Anwendungsszenarien im Projekt FAIR Data Publisher ergeben sich spezifische Anforderungen an die FDO-Architektur. Eine zentrale Anforderung aller Fälle ist die Durchsuchbarkeit und Auffindbarkeit in der FDO-Infrastruktur, da fehlende Suchfunktionalitäten in Handle-Systemen als wesentliches Hindernis für die praktische Umsetzbarkeit identifiziert wurden. Weiterhin muss die Infrastruktur skalierbar sein, um sowohl einfache Metadaten-Einträge als auch hochvolumige Datentransfers zu unterstützen. Komplexe Anfragen und große Datenmengen erfordern dabei effiziente APIs und eine entsprechend ausgelegte Infrastruktur.

Zudem sollte eine gewisse Strukturflexibilität gegeben sein, um verschiedene Datenformate und Verarbeitungsprozesse zu unterstützen, von proprietären PDF-Dokumenten über JSON-strukturierte Daten bis hin zu selbstkonstruierten FDO-Profilen, die von Communitys eigenständig entwickelt werden können. Die rechtskonforme Transformation ist dabei essenziell, insbesondere bei der Extraktion von Informationen aus urheberrechtlich geschützten Materialien. Im medizinischen Anwendungsfall zeigt sich die Notwendigkeit robuster Compliance-Mechanismen und Datenschutz-Frameworks, die regulatorische Anforderungen erfüllen und Zulassungsprozesse vereinfachen. Schließlich müssen nachhaltige Betriebsmodelle konzipiert werden, die verschiedene Finanzierungsansätze wie Mitgliedschaftsmodelle, Pay-per-View-Systeme oder institutionelle Förderungen integrieren, um die Strukturen langfristig und idealerweise kostendeckend betreiben zu können.

Das FAIR Data Publisher Projekt adressiert zentrale Herausforderungen der digitalen Datenverwaltung, indem es FAIR Digital Objects in Datenräume integriert, um eine sichere, souveräne und interoperable Datenlandschaft zu schaffen. Die detaillierte Analyse der Anwendungsfälle und der Anforderungen an die FDO-Architektur sowie die Betrachtung etablierter Datenräume wie Catena-X und des Mobility Data Space unterstreichen die Komplexität und das Potenzial dieses Vorhabens. Der Erfolg hängt maßgeblich von der Entwicklung nachhaltiger Betriebs- und Geschäftsmodelle ab, die eine ausgewogene Balance zwischen technischer Leistungsfähigkeit, rechtlicher Konformität und wirtschaftlicher Tragfähigkeit gewährleisten.

{make or buy}

Ein Unternehmen benötigt regelmäßige Analysen und steht vor der Wahl: Soll es ein eigenes Data-Analytics-Team aufbauen (Make) oder die Analysen von einer externen Agentur einkaufen (Buy)? „Make“ würde erforderliche Investitionen in Lizenzen, Schulungen und Gehälter bedeuten, bietet aber volle Kontrolle über sensible Daten und ermöglicht ad-hoc Analysen. „Buy“ ist schneller einsatzbereit und kostet initial weniger, jedoch entstehen Abhängigkeiten und Datenschutzrisiken.

FDOS ALS TEIL EINER ÖFFENTLICHEN DATENINFRASTRUKTUR

Es ergeben sich unter anderem drei zentrale Empfehlungen, die notwendig sind, um das FDO-Konzept im FAIR Data Publisher zum Erfolg zu bringen und dessen Betrieb zur Verfügung zu stellen. Diese Empfehlungen basieren auf Interviews mit Use Cases und Learnings aus anderen Datenraumprojekten.

Die erste Empfehlung betrifft die Weiterentwicklung und Erweiterung der Anwendungsfälle (Use Cases). Die bisherigen Use Cases des Projekts stellen zwar einen guten Anfang dar, sind jedoch nur prototypisch oder befinden sich auf Proof-of-Concept-Level. Um die Mehrwerte der FDOs hinsichtlich Auffindbarkeit und Datennutzung zu verdeutlichen, ist es notwendig, weitere (öffentliche) Förderung zu erhalten. Ziel ist es, die entwickelte Software marktreifer zu machen, d. h. ein höheres TRL (Technology Readiness Level) zu erreichen. Es wird empfohlen, die bestehenden Anwendungsfälle mit weiteren Daten und Partnern weiterzuentwickeln. Zudem sollen neue Anwendungsfälle gesucht werden, die insbesondere den domänenübergreifenden Aspekt der FDOs demonstrieren, da dieser bisher relativ wenig gezeigt wurde.

Die zweite Empfehlung schlägt vor, das FDO-Konzept und die entwickelte Software in bestehende Datenraumprojekte zu integrieren. Dies könnte Projekte umfassen, die von Organisationen wie acatech oder dem Bundesministerium für Verkehr (BMV) gefördert werden, wie beispielsweise den Datenraum Mobilität. Man könnte argumentieren, dass die Betriebsgesellschaft solcher Datenräume gewisse FDO-Software betreiben sollte, um die Kosten zu decken und FDOs dort zu integrieren, wodurch deren Mehrwerte sowie die operative Betreibbarkeit gezeigt werden könnten.

Die dritte Empfehlung ist die Anerkennung der FDO-Lösung als Teil einer öffentlichen Dateninfrastruktur. Aufgrund der infrastrukturellen Aspekte des FDO-Konzepts und der benötigten Software sollte es in Betracht gezogen werden, bestimmte Services mittels öffentlicher Gelder oder von öffentlichen Stellen betreiben zu lassen. Solche Dienste stiften einen Mehrwert für alle und bilden die Grundlage für Wertschöpfung und Innovation in Deutschland. Hierfür müssten Gelder bei Stellen wie der Bundesregierung gefunden werden, um diese Dienste zu betreiben. Konkreter Vorschlag für die Verankerung dieser Dienste könnte die NFDI (Nationale Forschungsdateninfrastruktur) sein.

Allgemein ist festzuhalten, dass es entscheidend sein wird, „Henne-Ei-Probleme“ zu überwinden und die Finanzierung langfristig zu sichern. Dazu muss die Technologie auf höheres „Technology Readiness Level“ (TLR) gebracht werden und den „Unique Selling Point“ (USP) durch weitere Anwendungsfälle, bestenfalls mit Industriepartnern, untermauern. Ein weiterer Ansatz, der sich aus der zweiten Handlungsempfehlung ergibt, eignet sich zur Ergänzung des Vorgehens. So kann die Technologie parallel in bestehende Datenräume eingebracht werden, was Finanzierung und Weiterentwicklung vereinfacht und sichert.

Insgesamt schafft das Projekt FAIR Data Publisher die Grundlage zur Weiterentwicklung

INTERVIEW JOSHUA GELHAAR UND MAIK MANNSFELD,

Fraunhofer ISST

>>Wie stellen Datenräume denn sicher, dass die Teilnehmer wirklich die Kontrolle über ihre sensiblen Informationen behalten?

<<Das ist fundamental: Datenanbieter behalten die vollständige Kontrolle über ihre Daten, selbst wenn diese mit anderen geteilt werden, was durch klare Richtlinien, Identitätsmanagement und Zugriffskontrollen sichergestellt wird. Der Datenaustausch erfolgt dabei dezentral über vertrauenswürdige Peer-to-Peer-Verbindungen mittels sogenannter Konnektoren, wobei die Daten an ihrem Ursprungsort bleiben und lediglich Metadaten in Katalogen geteilt werden.

>>FDO-Connect baut auf FDO One auf und arbeitet an FAIR Digital Objects. Was ist aktuell das Hauptziel dieser Initiative?

<<FDO Connect fokussiert sich auf die Entwicklung eines FDO-Profils

STRATEGISCHE BETRACHTUNGEN_ÖKONOMISCHE BETRACHTUNGEN

für Forschungsdaten und die anschließende Transformation sowie Integration bestehender Datensätze. Die Synergie mit Datenräumen bildet das Fundament, um die rechtskonforme Datennutzung von FAIR Digital Objects effizient in den verschiedenen Anwendungsdomänen zu gewährleisten.

>>Welche technische Anforderung hat sich als besonders zentral für die praktische Nutzbarkeit der FDO-Architektur herauskristallisiert?
<<Eine zentrale Anforderung aus allen Use Cases ist ganz klar die Durchsuchbarkeit und Auffindbarkeit der FDO-Infrastruktur, da fehlende Suchfunktionalitäten in Handle-Systemen ein wesentliches Hindernis für die praktische Umsetzbarkeit darstellen. Außerdem muss die Infrastruktur extrem skalierbar sein, um sowohl einfache Metadaten-Einträge als auch hochvolumige Datentransfers effizient unterstützen zu können.

digitaler Infrastrukturen in Wissenschaft und Industrie. FDOs können bei konsequenter Umsetzung und institutioneller Unterstützung zu einem wichtigen Baustein einer offenen Datenökonomie werden. Es gibt allerdings noch einige Aufgaben auf dem Weg dorthin. Strategische Positionierung und technologische Weiterentwicklung sind, Stand heute, die wichtigsten Schritte.

FAZIT UND EMPFEHLUNGEN

NOTWENDIGE WEITERE SCHRITTE

> *Weiterentwicklung und Erweiterung der Anwendungsfälle (Use Cases): Die bestehenden, nur prototypisch oder auf Proof-of-Concept-Level befindlichen Use Cases sollen weiterentwickelt und um neue Anwendungsfälle ergänzt werden, um die entwickelte Software marktreifer zu machen (höheres TRL) und insbesondere den domänenübergreifenden Aspekt der FAIR Digital Objects (FDOs) zu zeigen.*

> *Integration in bestehende Datenraumprojekte: Das FDO-Konzept und die entwickelte Software sollen in existierende Datenraumprojekte, wie beispielsweise den Datenraum Mobilität (gefördert von acatech oder dem BMFTR), integriert werden, um die operative Betreibbarkeit und die Mehrwerte der FDOs zu demonstrieren.*

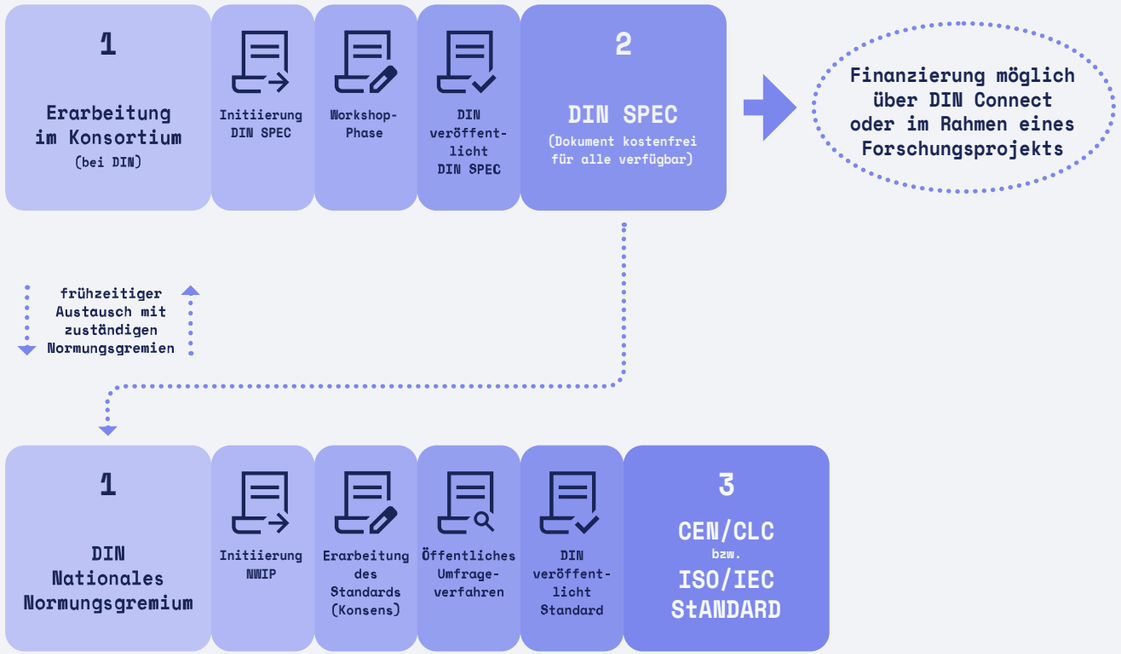
> *Anerkennung als öffentliche Dateninfrastruktur: Aufgrund der infrastrukturellen Aspekte des FDO-Konzepts soll die Lösung als Teil einer öffentlichen Dateninfrastruktur anerkannt werden, sodass bestimmte Services mittels öffentlicher Gelder oder von öffentlichen Stellen betrieben werden, beispielsweise verankert in der NFDI.*

Transferstrategie für Normung und Standardisierung

Im Rahmen des Projekts FAIR Data Publisher wurde auch eine Transferstrategie für Normung und Standardisierung erarbeitet. Ziel dabei ist es, FDOs mit Hilfe des FAIR Data Publisher-Konzepts nachhaltig in der nationalen, europäischen und internationalen Normung zu verankern. Die Strategie basiert auf Erkenntnissen aus der Innovationsforschung, die Normung als Katalysator für technologische Entwicklung und Marktzugang identifiziert. Es geht auch darum, Normung als strategisches Instrument zu verstehen, das frühzeitig in Innovationsprozesse integriert werden sollte.

¬ Das Konzept der FAIR Digital Objects kommt aus Wissenschaft und Forschung, die bereits seit Jahrzehnten vor der Herausforderung stehen, weltweit Daten auffindbar, zugänglich und interoperabel auszutauschen. Das Teilen dieser Erkenntnisse mit der Industrie und die Überzeugung weiterer Akteure ist essenziell, um die Interoperabilität von Datenräumen mit Hilfe von FDOs zu unterstützen.

Im ersten Schritt sollen die Projektergebnisse in den einschlägigen Normungsgremien vorgestellt werden, um gemeinsam mit den Normungsexperten das Konzept zu diskutieren und Anknüpfungspunkte für bestehende Projekte zu identifizieren. Durch die frühzeitige Einbindung in die Normungsgremien können schnell Rückschlüsse und Empfehlungen für die weitere Implementierung gewonnen werden. Abhängig vom Reifegrad der Projektidee kann gemeinsam mit den einschlägigen Expertinnen und Experten das beste Vorgehen für ein internationales Normungsprojekt erarbeitet werden.



•DIN-NORMUNGSPROZESS•

In der Transferstrategie wird mit Hilfe des Projekts FAIR Data Publisher ein strukturierter DIN-Normungsprozess für FDOs empfohlen. Schritt 1 beschreibt den frühzeitigen Austausch mit den einschlägigen Normungsgremien, um gemeinsam mit den Normungsexperten das Konzept zu diskutieren und Anknüpfungspunkte für bestehende Projekte zu identifizieren. Es geht auch darum, Normung als strategisches Instrument zu verstehen, das frühzeitig in Innovationsprozesse integriert werden sollte. Schritt 2 zeigt die Entwicklung einer DIN SPEC durch konsortiale Erarbeitung: Nach der Initiierung folgen Workshop-Phasen, wobei die Finanzierung über DIN Connect oder Forschungsprojekte erfolgen kann. Das Ergebnis ist eine kostenfreie, öffentlich verfügbare DIN SPEC. Schritt 3 beschreibt die Weiterentwicklung zum vollwertigen Standard: Das nationale Normungsgremium initiiert ein New Work Item Proposal (NWIP), gefolgt von der konsensbasierten Standarderarbeitung und einem öffentlichen Umfrageverfahren. Das Ergebnis dieses Prozesses kann ein internationaler Standard (DIN ISO) oder im Rahmen der Wiener Vereinbarung (Vienna Agreement) ein parallel europäischer und international abgestimmter Standard (DIN EN ISO) sein.

→ DIN SPEC

DIN SPEC ist ein schnelles Standardisierungsverfahren des DIN, das es ermöglicht, innovative Ideen und Forschungsergebnisse innerhalb weniger Monate in marktfähige technische Spezifikationen zu überführen. Im Gegensatz zu klassischen DIN-Normen, deren Entwicklung Jahre dauern kann, folgt die DIN SPEC dem PAS-Verfahren (Publicly Available Specification) und richtet sich an Unternehmen, Forschungseinrichtungen und Organisationen, die zeitnah technische Regelwerke benötigen, um Innovationen zu beschleunigen. DIN SPECs werden kostenlos als PDF bereitgestellt und haben eine maximale Lebensdauer von sechs Jahren, mit systematischen Überprüfungen nach drei und sechs Jahren. Danach müssen sie entweder überarbeitet, in eine vollwertige DIN-Norm überführt oder zurückgezogen werden. Das Verfahren bietet eine flexible Alternative zur traditionellen Normung und eignet sich besonders für dynamische Technologiefelder, wo schnelle Markteinführung und Konsensbildung zwischen wenigen, aber engagierten Akteuren im Vordergrund stehen.

INTERVIEW
ANNA PETSCHULIES,
DIN

>>Warum ist DIN an dem Projekt beteiligt?
<<DIN als nationale Normungsorganisation und Bindeglied zur europäischen und internationalen Normung berät die Projektpartner hinsichtlich Normung und Standardisierung und erstellt einen Bericht inkl. Handlungsempfehlungen für den Transfer in die Normung und Standardisierung. Ziel ist es, nationale Initiativen mit internationalen Entwicklungen zu synchronisieren, um Fragmentierung zu vermeiden und Innovation zu fördern. Die Förderung von Interoperabilität und die Vermeidung von Insellösungen sind entscheidend, um die Wettbewerbsfähigkeit der Wirtschaft in Deutschland und der EU zu stärken. Durch einheitliche europäische und internationale Standards werden Innovationen erleichtert und die Grundlage für einen funktionierenden

STRATEGISCHE BETRACHTUNGEN_TRANSFERSTRATEGIE FÜR NORMUNG UND STANDARDISIERUNG

europäischen Datenraum geschaffen.

>>Welchen Nutzen hat eine DIN SPEC?
<<Die Marke DIN genießt weltweit hohes Vertrauen. Eine DIN SPEC erleichtert die Akzeptanz neuer Technologien und Lösungen. DIN SPECs können innerhalb weniger Monate erstellt werden und bieten eine schnelle Möglichkeit, Innovationen zu standardisieren und am Markt zu platzieren. DIN SPECs dienen außerdem häufig als vornormative Textbasis und erleichtern die Überführung in europäische (CEN/CENELEC) und internationale (ISO/IEC) Normungsprozesse. Dadurch können nationale Interessen frühzeitig eingebracht und die internationale Anschlussfähigkeit verbessert werden.

Parallel dazu wird empfohlen, eine nationale DIN SPEC als Kommunikationsinstrument zu erstellen. Die DIN SPEC hat immer das Ziel, zu einem späteren Zeitpunkt in die internationale Normung überführt zu werden. Sie vereinfacht den Start eines internationalen Normungsprojektes, weil damit bereits eine vornormative Textbasis geschaffen wird. Im Gegensatz zu klassischen DIN-Normen, die oft langwierige Abstimmungsprozesse durchlaufen, ermöglicht die DIN SPEC eine deutlich schnellere Veröffentlichung, da sie in einem Konsortium erarbeitet und nicht auf die strengen konsensbasierten Verfahren bestehen muss. Dies ist besonders für Ergebnisse aus Forschungsprojekten von Bedeutung, die in einem dynamischen Umfeld agieren und frühzeitig Standards setzen wollen. Bevor eine DIN SPEC gestartet werden kann, erfolgt immer eine interne Prüfung auf vergleichbare oder bereits vorhandene Projekte (europäisch und international), denn auch hier gilt, dass keine Doppelnormung erfolgen darf.

Diese dreistufige Strategie gewährleistet eine schrittweise Validierung und Akzeptanz der FAIR Data Publisher-Konzepte, beginnend mit einer flexiblen Spezifikation bis hin zur verbindlichen internationalen Normung, wodurch eine breite Adaption und Interoperabilität in der globalen Dateninfrastruktur ermöglicht wird.

FAZIT UND EMPFEHLUNGEN

NOTWENDIGE WEITERE SCHRITTE

> *Bewerbung zur Durchführung eines DIN-SPEC-Verfahrens*

Verweise Normung:
↗ Weitere Informationen zum Thema DIN SPEC
↗ Übersicht relevanter Normungsprojekte
↗ Basiswissen Normung
↗ Rolling Plan 2025
↗ EU Normungsstrategie
↗ Europäische Datenstrategie
↗ Mehr über WIPANO
↗ Mehr über DIN Connect

Strategische Zusammenfassung und Anregungen

Notwendige weitere Schritte, die sich direkt aus dem Projekt FAIR Data Publisher ergeben

¬ Der FAIR Data Publisher stellt eine zentrale technologische Initiative dar, die darauf abzielt, die Auffindbarkeit, Zugänglichkeit, Interoperabilität und Nachnutzbarkeit von Daten radikal zu verbessern. Diese Initiative fungiert als Einstiegspunkt in eine vernetzte und intelligente Datenökonomie und kann als mögliches Fundament für die Entwicklung und Skalierung von KI-Anwendungen in Deutschland dienen. Durch die Verwendung von FDOs können neue Potenziale in einem global integrierten Datenraum erschlossen werden, beispielsweise durch den automatisierten Austausch von Forschungsdaten, plattformübergreifende Produktionsoptimierung oder die nahtlose Bereitstellung von Trainingsdaten für KI-Anwendungen. Das Projekt zeigt im Ergebnis, dass die Anwendung der FDO-Grundsätze in einer verteilten Architektur den Wert von Daten signifikant erhöht. FDOs beseitigen Defizite im Umgang mit wertvollen Daten und verhindern die Entstehung von „Dark Data“, „Stranded Data“ und „Siloed Data“. Sie können als selbstbeschreibende Objekte wesentlich dazu beitragen, die Compliance gegenüber Regulierungen und Verträgen sicherzustellen und sogar größtenteils zu automatisieren.

DIE FDO-SERVICE-ARCHITEKTUR

Die entwickelte FDO-Service-Architektur, welche Handle-basierte FDOs mit Nanopublikationen und Linked Data verbindet, generiert eine Reihe signifikanter Mehrwerte. Sie überbrückt erfolgreich die „Handle-Welt“ mit der „Linked-Data-Welt“ und schafft dadurch eine technologisch unabhängige Basis für den systemübergreifenden Austausch und die Nutzung digitaler Objekte. Da digitale Objekte nun maschinenlesbar und -aktionsfähig sind, ermöglicht dies eine hochautomatisierte Verarbeitung und Analyse von Daten. Ein dezentrales Nanopublikations-Netzwerk gewährleistet die Unveränderlichkeit (Immutability) von Daten und Identifikatoren, wodurch eine langfristige und manipulationssichere Verfügbarkeit sichergestellt wird. Zudem ermöglicht das Design der Nanopublikations-Architektur die effiziente Verwaltung sehr großer Mengen digitaler Objekte durch horizontale Skalierung. Durch autonome FDOs (aFDOs) wird eine dezentrale Datengovernance ermöglicht, welche selbstständige Entscheidungsfindung, Interaktion und die Durchsetzung granularer Zugriffsrechte ohne zentrale Kontrollinstanz erlaubt, was die Datensicherheit und -integrität maßgeblich erhöht.

Jedoch ist die vollständige Dokumentation der umfassenden API-Spezifikationen für die verschiedenen Abfrage-Operationen noch erforderlich. Zentral für die Akzeptanz des dezentralen

Ökosystems ist die Weiterentwicklung, um sicherzustellen, dass Abfragen sowohl Nanopublikationen als auch Handle-basierte FAIR Digital Objects berücksichtigen. Es ist zudem notwendig, ein generisches und vollständig spezifiziertes Typensystem zur Assoziierung von FDOs mit ihren Operationen zu erarbeiten, um die Interoperabilität zu maximieren. Die Testbed-Infrastruktur muss professionalisiert und um redundante Dienste ergänzt werden, um die volle Produktionsreife zu erlangen. Zusätzlich zu den Referenzen auf die Registratur, in der die Profile veröffentlicht sind, sollten weitere Informationen zu den Profilen im FDO Manager deren Sichtbarkeit erhöhen. Eine dauerhaft implementierte Laufzeitumgebung für autonome FDOs sowie eine tiefgehende Analyse der benötigten Ressourcen stehen noch aus.

TOOL ZUR GENERIERUNG VON FDOS AUS UNSTRUKTURIERTEN DATEN

Das Tool zur automatisierten FDO-Generierung (AP2) wurde als Software-Lösung entwickelt, die eine automatisierte FDO-Generierung aus unstrukturierten Daten, wie PDF-Dokumenten, ermöglicht. Die Hauptfunktion des Tools ist die effiziente Extraktion und Transformation menschlich lesbarer Dokumente in maschinell verarbeitbare Daten. Der gesamte Workflow zur Erstellung von FDOs wird durch die Funktionalitäten und die intuitive, browserbasierte grafische Benutzerinnen- und Benutzeroberfläche (GUI) vereinfacht. Das System bietet außerdem eine KI-gestützte Profilauswahl (Autodetection) und Vorschläge für passende FDO-Profile, wodurch der Bedarf an Expertenwissen reduziert wird. Diese Technologie bildet eine entscheidende Grundlage für zukünftige autonome KI-Systeme.

Jedoch muss das System zukünftig zusätzliche Eingabeformate unterstützen, beispielsweise Grafiken, die in die Informationsgewinnung einfließen sollen, um die Vielseitigkeit der Datenquellen zu erweitern. Es ist eine kontinuierliche Verbesserung der KI-Algorithmen notwendig, um die Genauigkeit und Effizienz der intelligenten Wissensextraktion aus unstrukturierten Daten zu optimieren. Zusätzlich sollte der Workflow in der Lage sein, die Parametrisierung der Extraktion zu unterstützen, falls zusätzliche Informationen benötigt werden, etwa wenn nur Daten zu einer von mehreren in einem Dokument spezifizierten Komponenten extrahiert werden sollen. Die Profile-Matching-Verfahren müssen erweitert werden, um die Zuordnung unstrukturierter Daten zu FDO-Profilen präziser und adaptiver zu gestalten.

USE CASE – PROOFS OF CONCEPT – DIGITALE MEDIZIN

Der Anwendungsfall zeigt, dass FDOs die systemische Fragmentierung der medizinischen Datenlandschaft durch FDO-Grundsätze überwinden können. Die FDO-Architektur ist darauf ausgelegt, die Anforderungen des EU-Digitalrechtsrahmens (EHDS, Data Act, AI Act, DSGVO) durch maschinenlesbare Compliance-Informationen und persistente Identifikatoren zu unterstützen. Die standardisierte Verknüpfung heterogener medizinischer Daten schafft eine essenzielle Basis für die Entwicklung KI-basierter Lösungen in der personalisierten Diagnostik und Therapie. FDOs ermöglichen als selbstbeschreibende digitale Objekte eine dauerhafte Referenzierbarkeit und nahtlose Interoperabilität medizinischer Daten.

Jedoch bedarf es einer Validierung der entwickelten FDO-Lösungen mit realen Patientendaten, um die Praxistauglichkeit unter Berücksichtigung des Datenschutzes zu belegen. Eine DSGVO-konforme Anonymisierungspipeline sollte praktisch implementiert und validiert werden. Zudem muss die vollständige technische Integration der FDO-Profile mit der LinkAhead-Datenbankinfrastruktur über die konzeptionelle Ebene hinaus realisiert werden, um eine robuste lokale Datenhaltung zu ermöglichen. Es ist notwendig, die FDO-Architektur umfassend in bestehende reale klinische Informationssysteme zu integrieren, um deren Nutzbarkeit und Akzeptanz in der klinischen Praxis zu demonstrieren. Schließlich muss der konzipierte dreistufige Qualitätssicherungsprozess für medizinische Daten, der strukturelle, semantische und Cross-Source-Validierung umfasst, vollständig implementiert und angewendet werden.

USE CASE – PROOFS OF CONCEPT – INDUSTRIELLE FERTIGUNG

Der Anwendungsfall zeigt, dass die Erfüllung regulatorischer Anforderungen, wie EU-Batterieverordnung (digitaler Produktpass ab 2027) und AI Act, durch standardisierte Dokumentation und integrierte Compliance-Mechanismen der FDOs unterstützt werden kann. Der Service ermöglicht die effiziente automatisierte Extraktion und Strukturierung von Spezifikationsdaten aus unstrukturierten Produktblätter-PDFs, wodurch manuelle, fehleranfällige Prozesse ersetzt werden. FDOs fungieren als technologieunabhängiges Regelwerk, das den maschinenlesbaren Austausch und die automatisierte Verarbeitung von Daten und Diensten über heterogene Systeme (AAS, EDC) hinweg ermöglicht. Die Strukturierung unterstützt Decision Support Systeme bei der parametrischen Optimierung von Produktdesign und Fertigungsprozessen, und die FDOs können als AAS-Submodelle exportiert werden, was eine nahtlose Integration in digitale Produktpässe (DPP) und die Industrie 4.0-Infrastruktur sicherstellt, wodurch Vendor-Lock-in verhindert wird.

Jedoch muss die automatische Extraktion noch eine ausreichende Genauigkeit für den vollautomatischen Einsatz ohne manuelle Nachkontrolle erreichen, weshalb die Implementierung einer robusteren Validierungspipeline erforderlich ist. Dies bestätigt die fortgesetzte Notwendigkeit des Human-in-the-Loop-Ansatzes. Das aktuelle Geschäftsmodell leidet unter dem Fehlen „Unfairer Vorteile“ und der transienten Natur des Bedarfs, was die Entwicklung eines nachhaltigen und differenzierten Geschäftsmodells erfordert, das sich beispielsweise durch spezialisierte Modelle stärken lässt. Der Demonstrator dient lediglich als Proof of Concept für Datenextraktion und -strukturierung und ist daher noch nicht für die regulatorische Compliance oder den Produktiveinsatz geeignet, was eine umfassende Validierung für Produktivsysteme und regulatorische Compliance bedingt. Die fundamentale Datenfragmentierung auf Seiten der Komponentenhersteller bleibt ein Hindernis, solange die Hersteller nicht selbst standardisierte FDO-Daten bereitstellen. Hier ist der Aufbau einer breiten Akzeptanz und Standardisierung von FDO in der Lieferkette erforderlich. Darüber hinaus sind Skalierungsprobleme bei der Verarbeitung großer Datenmengen und die vollständige Integration in umfassende Datenökosysteme mit automatisierten Update- und Delete-Funktionen der FDO-Infrastruktur noch nicht vollständig gelöst.

USE CASE – PROOFS OF CONCEPT – ROBOTIK

Der Anwendungsfall zeigt, dass FDOs eine granulare Adressierbarkeit und dadurch eine präzise Verknüpfung wissenschaftlicher Aussagen sowie eine flexible Rekombination von Datensatzkomponenten ermöglichen. Durch die Entwicklung der Experimental Research Ontology (ERO) wird die implizite Semantik formalisiert und eine domänenübergreifende Interoperabilität gewährleistet. Die systematische Dokumentation von Protokollen erhöht die Reproduzierbarkeit experimenteller Forschungsergebnisse. Die FDO-Struktur unterstützt zudem eine kontinuierliche Datenpflege und Evolution des Datenkorpus. Die geschaffenen Strukturen antizipieren zukünftige regulatorische Anforderungen des europäischen Datenraums.

Jedoch muss die FDO-Modellierung hochauflösender Sensordaten angegangen werden, um den Kern robotischer Experimente vollständig abbilden zu können. Um die Experimentwiederholung zu gewährleisten, müssen die Spezifikationsbereiche für Software, Sensoren und Datenformatierung vollständig erarbeitet und in die FDO-Struktur integriert werden. Das Netzwerk hinter der Nanopublikations-Infrastruktur muss weiter wachsen, um die Kapazität für die breite Anwendung von FDOs zu schaffen. Für den Einsatz in Produktivsystemen ist es erforderlich, den Nanopublikations-Client um Update- und Delete-Methoden zu erweitern. Schließlich muss die globale UID-Generierung für Produktivsysteme durch robustere und zukunftssichere Mechanismen ersetzt werden.

USE CASE – PROOFS OF CONCEPT – KOMPLEXE SYSTEME

Der Anwendungsfall zeigt, dass Kenntnis von Kausalität gezieltes Eingreifen in Prozesse ermöglicht und der KI hilft, Zusammenhänge klarer zu verstehen, was zu zuverlässigeren und effizienteren Ergebnissen führt. Die Kombination von FDOs und kausaler Inferenz eröffnet neue Möglichkeiten, Zusammenhänge besser zu verstehen und automatisch verknüpfte Datenindizes zu erstellen. Der Prozess ist durch FDO-Kapseln hoch reproduzierbar und erreichte eine perfekte Validierung auf Surrogates (100 % Precision/Recall), was eine erhöhte FAIRness durch strukturierte, portable Artefakte mit sich bringt.

Jedoch muss die Limitierung auf Triaden überwunden werden, um die Analyse von komplexeren Kausalnetzwerken mit einer größeren Anzahl an Variablen zu ermöglichen. Vorverarbeitungsschritte erfolgen noch manuell. Für die nächste Phase wurden bereits Erweiterungen vorgesehen, u. a. die Integration nichtlinearer Tests, $\tau=0$ Fälle, sowie die Anbindung an den Nanopublikations-Adapter und das Profil der FAIR Data Publisher-Architektur.

USE CASE – PROOFS OF CONCEPT – QUANTEN COMPUTING

Der Anwendungsfall zeigt, dass Q-FDOs einen entscheidenden ersten Schritt hin zu einer ganzheitlichen, transparenten und offenen wissenschaftlichen Vergleichbarkeit im Bereich QML darstellen. Die Modularisierung erlaubt eine gezielte systematische Analyse einzelner Experiment-Aspekte

zur Hypothesenprüfung über Ursache-Wirkungs-Zusammenhänge, und die Übertragbarkeit der Q-FDO-Strukturen auf klassische ML-Ansätze führt zu einer breiteren Interoperabilität zwischen Quanten- und klassischen Lernverfahren. Q-FDOs besitzen einen operativen Mehrwert, da sie unmittelbar in experimentelle Pipelines eingebunden werden können, um Konfigurationen zu steuern.

Jedoch ist es noch notwendig, eine vollautomatisierte Konvertierungspipeline zu entwickeln, die den manuellen Transfer zwischen JSON-Konfigurationen, Nanopublikationen und Q-FDO-Abstraktionen eliminiert. Dadurch soll die produktive Nutzung für die iterative Quantenforschung ermöglicht werden. Die Modellvielfalt muss erweitert werden, um über den Data Reuploading Classifier hinauszugehen und weitere QML-Architekturen wie Variational Quantum Circuits, Quantum Neural Networks und hybride Ansätze standardisiert abzubilden. Zudem sollte das Zusammenspiel von Nanopublikationen und Q-FDOs weiter präzisiert und operationalisiert werden. Wichtig ist ferner die Integration unüberwachter Lernverfahren, um neben Klassifikationsproblemen auch Clustering, Dimensionsreduktion und andere Zweige des Quantum Machine Learning systematisch zu erfassen. Es sollte eine automatische Korrelationsanalyse implementiert werden, die Muster zwischen Experimenten erkennt und systematische Vergleiche verschiedener Hardware-Backends, Algorithmus-Parameter und Datensätze ermöglicht. Für die Praxistauglichkeit ist eine produktive Validierung mit realen Quantencomputern durchzuführen, die über synthetische MNIST-Daten hinausgeht und die Eignung der Q-FDO-Struktur in echten Forschungsumgebungen demonstriert.

STRATEGISCHER TEIL (AP4) – RECHTLICHE EINORDNUNG VON FDOS

Auch im rechtlichen Kontext bieten FDOs erhebliche Möglichkeiten. Der FDO-Ansatz hat das konkrete Potenzial einer umfassenden Compliance-Unterstützung, indem er durch einheitliche technische Mechanismen die Rechenschaftspflichten der DSGVO, die Dokumentationsanforderungen des AI Acts und die Transparenzberichte des Digital Services Acts bedient. Gleichzeitig gewährleistet dieser Standard eine native Interoperabilität, die simultan die Anforderungen an die Datenportabilität gemäß DSGVO, die Zugangs- und Weiternutzungsrechte des Data Acts, den Aufbau von Datenräumen im Rahmen des Data Governance Acts sowie die Weiterverwendung von Informationen des öffentlichen Sektors nach der PSI-Richtlinie erfüllt.

Jedoch sollten zur Realisierung hierzu unmittelbar und zeitnah die Entwicklungen nachgelagerter Rechtsakte, bei deren Ausgestaltung man mitwirken kann und muss, aufmerksam beobachtet und adressiert werden. Beispielsweise bieten die bis März 2027 zu erlassenden Durchführungsrechtsakte der FDO-Community ein konkretes Zeitfenster zur Mitgestaltung: Die technischen Spezifikationen für Datensätze, Kodiersysteme und sichere Verarbeitungsumgebungen können durch FDO-Prinzipien – maschinenlesbare Metadaten, persistente Identifikatoren, FAIR-konforme Qualitätsmetriken – maßgeblich geprägt werden. Ungeachtet dieser Rechtsakte ist für jeden Anwendungsfall ein Legal Requirement Engineering notwendig. Nanopublikationen kollidieren in ihrer derzeitigen Konzeption mit DSA-Entfernungspflichten für rechtswidrige Inhalte und DSGVO-Speicherbegrenzung, da kryptografisch gesicherte Persistenz technisch keine nachträgliche Löschung ermöglicht. Hier müssen organisatorische und technische Überlegungen angestellt werden, wie damit umgegangen wird. Die funktionale Qualifikation einzelner KI-Komponenten als eigenständige Systeme unter der KI-VO erfordert präzise technische Abgrenzung – eine generelle Herausforderung modularer Infrastrukturen, nicht FDO-spezifisch.

STRATEGISCHER TEIL (AP4) – FDO-INTEGRATION IN DIE ÖFFENTLICHE REGISTERLANDSCHAFT

Im Kontext der Registermodernisierung (RegMod) eignen sich FDOs als technische Grundlage für ein strukturiertes und verknüpftes Datenmanagement und unterstützen den Aufbau eines interoperablen Datenökosystems. FDOs können in bestehende Systeme integriert werden, um Registereinträge strukturiert und eindeutig referenzierbar zu machen. Die FDO-Architektur erlaubt es, Zugriffs- und Nutzungsrechte in Metadaten zu hinterlegen, um die Zugriffsberechtigung automatisch zu überprüfen und zu protokollieren. Zudem ermöglichen versionierte Ergänzungen der Metadaten eine verbesserte ex-post-Kontrolle durch die Betroffenen. Metadaten können durch versionierte Ergänzungen dargestellt werden, ohne dass die ursprünglichen Daten überschrieben werden, was eine verbesserte ex-post-Kontrolle durch die Betroffenen gewährleistet. Die Datennutzlast kann auf den jeweiligen behördlichen Servern verbleiben und Metadaten werden ausschließlich in staatlich kontrollierten Infrastrukturen verarbeitet. Das Konzept eines XFDO-Transformationsframework ist darauf ausgelegt, bestehende Verwaltungsdaten semantisch aufzuwerten und die Anbindung des NOOTS an das European-Once-Only-Technical-System (EU OOTS) zu erleichtern.

Jedoch sollte die notwendige Software zur behördlichen FDO-Erstellung auf staatlichen Infrastrukturen laufen und von staatlichen Betreibern angeboten werden. Technische und organisatorische Maßnahmen sind erforderlich, um Spannungen mit dem Zweckbindungsgrundsatz einzuhegen, wobei diese Maßnahmen äquivalent zu denen bei der Einführung einer Personenidentifikationsnummer sein müssen. Eine unabhängige Kontrollstelle sollte regelmäßig Protokolldaten evaluieren und auf unbefugte Zugriffe überprüfen. FDOs müssen durch eindeutige und dauerhafte Referenzen eine lückenlose Dokumentation, die Datenherkunft, Verwendungszweck und beteiligte Stellen umfasst, ermöglichen. Eine Verordnungsermächtigung in den maßgeblichen Gesetzen ist nötig, um technische Fragen der Implementierung, wie die der Datenorganisation durch FDOs, nachträglich klären zu können.

STRATEGISCHER TEIL (AP4) – FDO-USE-CASE NUTZUNGSRECHTE-MANAGEMENT

Im Bereich des Nutzungsrechte-Managements ermöglichen FDOs die automatisierte Rechtsverwaltung, da rechtliche Aspekte als maschineninterpretierbare Metadaten ausgedrückt werden können, was eine automatisierte Richtliniendurchsetzung und Lizenzierung ermöglicht. Das Konzept unterstützt die sichere Handhabung des gesamten Spektrums der Datennutzung – von offen bis restriktiv proprietär. Die gestärkte Datensouveränität durch vertragliche Regelungen wird ermöglicht, da individuelle Nutzungsvereinbarungen (DUAs) mittels ODRL technisch umgesetzt werden können, was besonders für urheberrechtsfreies Faktenwissen essenziell ist. Die Verwendung standardbasierter, maschinenlesbarer Policy-Sprachen wie ODRL trägt zur Gewährleistung der Interoperabilität in Datenräumen bei. Für Datenkonsumenten wird der Rechtsprozess technisch erleichtert, da die Darstellung rechtlicher Informationen als maschineninterpretierbare Metadaten im FDO das sofortige Verständnis der Bedingungen ermöglicht. Das Konzept integriert zudem weitere Zuteilungsrechte (wie Patentrechte, Geschäftsgeheimnisse und Datenschutzanforderungen) über das Urheberrecht hinaus.

Jedoch sollte im nächsten Schritt zur Stärkung der Robustheit eine Infrastruktur zur Geolokalisierungsverifizierung entwickelt werden, um die aktuellen Selbstangaben der Teilnehmerinnen und Teilnehmer durch die Integration mit externen Geolokalisierungsdiensten abzulösen.

STRATEGISCHER TEIL (AP4) – ÖKONOMISCHE BETRACHTUNGEN

Aus ökonomischer Sicht können FDOs wesentlich zum wirtschaftlichen Wachstum beitragen, indem sie eine robuste Infrastruktur schaffen, die den Austausch von Daten und Wissen deutlich effizienter macht. Insbesondere in technologieintensiven Sektoren wie Telekommunikation, Elektronik und IT-Systemen spielen standard-essentielle Patente (SEPs) eine zentrale Rolle für die langfristige Wachstumsdynamik. FDOs ermöglichen durch die Nutzung persistenter Identifikatoren und standardisierter Metadaten einen effizienten Zugang zu Daten und Lizenzen. Einheitliche Standards senken dabei die Transaktionskosten und Interoperabilität erleichtert die Integration neuer Innovationen und unterstützt die Entstehung von Folgeinnovationen. Vor allem die Kombination von FDOs und FRAND-Bedingungen (Fair, Reasonable and Non-Discriminatory) bringt zusätzliche Vorteile, da FDOs effizienten Zugang zu Daten und Lizenzen ermöglichen, was volkswirtschaftliche Kosten reduziert und produktivitätssteigernd wirkt.

Für eine fundierte wirtschaftspolitische Bewertung des FDO-Konzepts wären empirische Komponenten, konkretere Quantifizierungen und eine stärkere Integration der FDO-spezifischen Eigenschaften erforderlich. Dies setzt allerdings eine fundierte Datenlage voraus, um kausale Effekte von FDOs auf Wachstum ableiten zu können.

STRATEGISCHER TEIL (AP4) – BETRIEBS- UND GESCHÄFTSMODELLE VON DATENRÄUMEN

Die bestehenden, nur prototypisch oder auf Proof-of-Concept-Level befindlichen Anwendungsfälle sollen weiterentwickelt und um neue ergänzt werden, um die entwickelte Software marktreifer zu entwickeln, indem ein höheres Technology Readiness Level (TRL) erreicht wird. Insbesondere soll der domänenübergreifende Aspekt der FDOs gezeigt werden. Das FDO-Konzept und die entwickelte Software sollen in existierende Datenraumprojekte, wie den Datenraum Mobilität, integriert werden, um die operative Betreibbarkeit und die Mehrwerte der FDOs zu demonstrieren. Aufgrund der infrastrukturellen Aspekte des FDO-Konzepts soll die Lösung als Teil einer

STRATEGISCHE BETRACHTUNGEN_STRATEGISCHE ZUSAMMENFASSUNG UND ANREGUNGEN

öffentlichen Dateninfrastruktur anerkannt werden, was bedeutet, dass bestimmte Services mittels öffentlicher Gelder oder von öffentlichen Stellen betrieben werden, beispielsweise verankert in der Nationalen Forschungsdateninfrastruktur (NFDI).

STRATEGISCHER TEIL (AP4) – TRANSFERSTRATEGIE FÜR NORMUNG UND STANDARDISIERUNG

In der Transferstrategie wird mit Hilfe des Projekts FAIR Data Publishers ein strukturierter DIN-Normungsprozess für FAIR Digital Objects empfohlen. Im ersten Schritt wird die Bewerbung zur Durchführung eines DIN-SPEC-Verfahrens angeregt.

Allgemeine Anregungen zu weiteren Schritten

¬ Da der FAIR Data Publisher explizit auf der technischen Implementierung der FAIR-Prinzipien durch FDOs basiert und gleichzeitig die institutionelle und rechtliche Verankerung in einem nationalen und europäischen Datenraum erlangt werden muss, bilden FAIR Data Spaces eine passende Brücke, da es die Gaia-X-Infrastruktur (für vertrauenswürdige und föderierte Datenräume) mit den FAIR-Prinzipien und der Forschungsdaten-Community (NFDI) zusammenführt. Es dient als die konzeptionelle und strategische Blaupause für die Art von Ökosystem, die Ihr FDO-Projekt technisch realisieren möchte.

Im folgenden einige Anregungen, um FAIR Digital Objects und den FAIR Data Publisher weiter zu etablieren.

KARTIERUNG DER EUROPÄISCHEN UND NATIONALEN FÖRDERKULISSE

Zur Kartierung der Förderkulisse zu FAIR Digital Objects wird eine systematische Recherche europäischer (z. B. Horizon Europe, EOSC, Connecting Europe Facility) und nationaler (z. B. BMFTR, DFG, FWF, SNSF) Programme, Projekte und Aufrufe angeregt. Im Zuge dessen werden Themenfelder, Budgets, Laufzeiten und Zielgruppen analysiert, um Förderlücken und bestehende Überschneidungen sowie komplementäre Effekte zu identifizieren. Im Ergebnis sollen strategisch adressierbare Instrumente und damit verknüpfte Bedingungen gelistet werden. Basierend auf dieser Analyse und politischen Agenden (z. B. Hightech Agenda) sollte zudem die Entwicklung auch FDO-spezifischer Förder-Instrumente vorangetrieben werden. Diese sollten klar definierte Ziele (z. B. FDO-Implementierung, Standardisierung, Pilotprojekte) haben, innovative Ansätze unterstützen und die Zusammenarbeit über Disziplinen und Ländergrenzen hinweg fördern und zudem mit messbaren Erfolgskriterien ausgestattet sein. Ein Feedback-Mechanismus mit der FDO-Community ist für die iterative Optimierung des Förderdesigns essenziell.

ANBINDUNG AN DIE VERWALTUNGSDIGITALISIERUNG

Die föderale Struktur der deutschen Verwaltung erfordert besondere Berücksichtigung bei der Implementierung nationaler Digitalstrategien. Der IT-Planungsrat koordiniert diese Zusammenarbeit. FDO-Standards könnten als übergreifende Datenarchitektur diese föderale Koordination technisch unterstützen und einheitliche Schnittstellen zwischen Bund, Ländern und Kommunen erleichtern. FDOs können in bestehende Systeme integriert werden, um Registereinträge strukturiert und eindeutig referenzierbar zu machen. Sie können dazu beitragen, individuelle Zugriffs- und Nutzungsrechte in den Metadaten zu hinterlegen und Zugriffsberechtigung automatisch zu überprüfen und zu protokollieren. FDOs können die semantische Interoperabilität in Verwaltungsprozessen und im Recht beschleunigen. Das Konzept der Extended FAIR Digital Objects (XFDOs) aus dem FITKO-Umfeld ist mittlerweile in erste erfolgreiche Tests überführt worden. Hier wird ein Auftrag durch den IT-Planungsrat an FITKO benötigt, um diese Entwicklung voranzutreiben.

Die im Koalitionsvertrag 2021-2025 verankerte Vision eines „Deutschland-Stacks“ und der „Deutschen Verwaltungscloud“ formuliert ein ambitioniertes Ziel: interoperabel, souverän und ausschließlich mit vertrauenswürdigen Anbietern zu operieren. Diese strategische Ausrichtung schafft ideale Voraussetzungen für die Integration einer FDO-Service-Architektur als technologisches Fundament digitaler Souveränität. Die Deutsche Verwaltungscloud-Strategie (DVS) konkretisiert bereits wesentliche Prinzipien, die mit FDO-Konzepten konvergieren: Multi-Cloud Strategie und offene Schnittstellen sowie strenge Sicherheits- und Transparenzvorgaben.

Ein zentrales Element digitaler Souveränität ist die Vermeidung von Abhängigkeiten zu einzelnen Technologieanbietern. Die Deutsche Verwaltungscloud adressiert dieses strategische Risiko explizit. z. B. sollen durch offene Standards die Wechselfähigkeit in der öffentlichen Verwaltung garantieren und sogenannte Lock-In-Effekte ausschließen. FDO-Architekturen verstärken diese Wechselfähigkeit durch technologieunabhängige Datenstrukturen, die unabhängig von spezifischen Cloud-Anbietern oder Softwareplattformen operieren können.

Die Integration von FDO-Services in den Sovereign Cloud Stack (SCS) würde die Flexibilität beim Anbieterwechsel auf die Datenebene erweitern und einheitliche, interoperable Datenstrukturen unabhängig vom gewählten Cloud-Provider gewährleisten.

Das Zentrum Digitale Souveränität (ZenDiS) unterstützt diese strategische Ausrichtung durch systematische Förderung von Open Source-Technologien. FDO-Implementierungen auf Open Source-Basis würde dessen Engagement optimal unterstützen und gleichzeitig internationale Standardisierungsprozesse (z. B. mit der direction interministérielle du numérique (DINUM) in Frankreich) vorantreiben.

Die Bundesdruckerei kritisiert aktuell den Deutschland-Stack und fordert ein „Ende-zu-Ende-Denken“. Der D-Stack sollte aber mehr sein als ein App-Store mit einzelnen neuen digitalen Werkzeugen. Nötig sei vielmehr ein Zusammenspiel von Funktionalität, Regulierung, Infrastruktur und Betrieb. Zugleich gelte es, die europäische Dimension mit dem Euro-Stack zu berücksichtigen und ein „Ende-zu-Ende-Denken“ mit aufzunehmen. Nur so könne sich der D-Stack nahtlos in bestehende und künftige IT-Architekturen einfügen.

Verweise:
↗ Deutsche Verwaltungscloud-Strategie
↗ Deutschland-Stack: Rückgrat der staatlichen Daseinsvorsorge in der Digitalwelt

FDOS FÜR HIGH VALUE DATASETS

Die Verordnung (EU) 2023/138 über hochwertige Datensätze (High Value Datasets, HVD) wurde als Durchführungsverordnung zur überarbeiteten PSI- bzw. Open-Data-Richtlinie (EU) 2019/1024 erlassen. Ziel ist es, die in der Richtlinie vorgesehenen Kategorien besonders wertvoller öffentlicher Daten – etwa aus den Bereichen Georaum, Umwelt, Statistik, Unternehmen, Mobilität und Meteorologie – konkret zu bestimmen und ihre Bereitstellungsanforderungen zu harmonisieren. Mit der Verordnung wurden erstmals verbindliche Vorgaben zur kostenlosen, maschinenlesbaren und API-basierten Bereitstellung solcher Daten festgelegt, um ihre Weiterverwendung im Sinne der europäischen Datenstrategie zu fördern und ein interoperables, offenes Datenökosystem zu schaffen.

FAIR Digital Objects (FDOs) können wesentlich zur Umsetzung und Verstetigung der High-Value-Datasets-Verordnung (EU) 2023/138 beitragen, indem sie die darin geforderte maschinenlesbare, interoperable und offene Bereitstellung öffentlicher Daten technisch und konzeptionell absichern. FDOs bieten dafür ein robustes Rahmenmodell: Jeder Datensatz wird als eindeutig identifizierbares, versioniertes und mit standardisierten Metadaten versehenes Objekt beschrieben. Dadurch wird die Nachnutzung rechtssicher, die Verknüpfung zwischen Themenbereichen (z. B. Umwelt und Wirtschaft) vereinfacht und die Interoperabilität über Verwaltungs- und Ländergrenzen hinweg gewährleistet.

STRATEGISCHE BETRACHTUNGEN_STRATEGISCHE ZUSAMMENFASSUNG UND ANREGUNGEN

Im Sinne der Fraunhofer-Studie zu hochwertigen Datensätzen könnten FDOs die dort beschriebenen technischen Herausforderungen – etwa mangelnde Dokumentation, unklare Lizenzinformationen und fehlende Schnittstellen – systematisch adressieren. Sie verbinden rechtliche Klarheit (über eingebettete Lizenz- und Herkunftsangaben) mit technischer Nachhaltigkeit (dauerhafte Referenzierbarkeit und API-Zugriff). Damit schaffen sie die infrastrukturelle Grundlage, um High-Value-Datasets nicht nur bereitzustellen, sondern als Teil eines lebenden, FAIR-orientierten Ökosystems für Open Data kontinuierlich nutzbar zu halten.

Verweis:
↗ Studie des BMWK zu hochwertigen Daten

USE CASE – SMART METADATA UND KOMPLEXE, MULTIVARIATE COMPOSITE-INDIKATOREN FÜR DIE AMTLICHE STATISTIK

Produkte der amtlichen Statistik ermöglichen Politik, Wirtschaft, Wissenschaft und der Gesellschaft im Allgemeinen auf Grundlage (meist weltweit) standardisierter und kontrollierter Erhebungsmethoden wertvolles Orientierungswissen, um Entscheidungen auf Evidenz zu gründen und durch komplexe Herausforderungen zu navigieren. Doch die Möglichkeiten der amtlichen Statistik stoßen mittlerweile an Grenzen, die Analysen in der nötigen Detailtiefe für neue, experimentelle Produkte herunterzubrechen. Um diese Grenzen zu überwinden, sind domänenspezifische Datenzugänge und die Verknüpfung mit metadatenreichen Datensätzen, beispielsweise aus der Forschung, zwingend notwendig. Deutschland hinkt bei der Modernisierung der amtlichen Statistik im weltweiten Vergleich hinterher.

Geplante Mikrodatenzentren sollen als zentrale, sichere Plattformen dienen, um der amtlichen Statistik einen schnellen und unkomplizierten Zugang zu hochaufgelösten Einzeldatensätzen zu gewähren. Um das volle Potenzial dieser Daten zu heben und sie mit externen Quellen kombinierbar zu machen, müssen die Daten selbst und ihre Beschreibungen (Metadaten) maschinenlesbar und intelligent aufbereitet sein.

Konzepte wie „Smart Metadata“ der Paris21 OECD zielen darauf ab, statische Dokumentationen durch dynamische, maschinen-verwertbare Metadaten zu ersetzen. Dies ist die strategische Grundlage, auf der die technische Umsetzung aufbauen kann. Die konsequente Weiterentwicklung dieses Gedankens führt zu FAIR Digital Objects (FDOs). Bei diesem Ansatz werden die Mikrodaten untrennbar mit einem reichen Set an Metadaten in einem digitalen Objekt gebündelt.

Ein besonderes Augenmerk der amtlichen Statistik gilt derzeit dem Biodiversitätsmonitoring, wozu das Bundesnaturschutzgesetzes verpflichtet. Das I-ADOPT-Framework bietet ein konzeptionelles Modell, um Beobachtungs- und Messvariablen präzise und maschinenlesbar zu beschreiben. In Verbindung mit FAIR Digital Objects ermöglicht es, Daten zu Biodiversität – etwa zu Artenvorkommen, Lebensräumen oder ökologischen Indikatoren – standardisiert zu strukturieren und global interoperabel zu machen. Während Biodiversitätsstatistiken bislang oft durch heterogene Formate und uneinheitliche Begriffe erschwert werden, schafft I-ADOPT die semantische Grundlage für Vergleichbarkeit und Integration. Werden diese Konzepte in FDOs umgesetzt, entstehen klar identifizierbare, dauerhaft auffindbare und verknüpfbare Datenobjekte, die sowohl für amtliche Statistik als auch für Forschung und Politik genutzt werden können. Damit ließe sich Biodiversitätsmonitoring zu einem dynamischen, international anschlussfähigen Informationssystem entwickeln, das Transparenz schafft und evidenzbasierte Entscheidungen in Umwelt- und Naturschutzpolitik unterstützt.

Das Nationale Monitoringzentrum zur Biodiversität ruft derzeit zur Einreichung innovativer Projektideen im Rahmen des Förderprogramms „Zukunftsfähiges Biodiversitätsmonitoring“ auf. Gefördert werden Vorhaben, die das bundesweite Biodiversitätsmonitoring gezielt weiterentwickeln – etwa durch neue Methoden, digitale Lösungen oder den Ausbau von Citizen Science und ehrenamtlichem Engagement.

Eine stete Weiterentwicklung findet in der amtlichen Statistik bei der Berechnung von Preisentwicklungen statt. Diese wird zunehmend komplexer, da nicht nur direkte Preisänderungen, sondern auch qualitative Anpassungen wie Shrinkflation (verkleinerte Mengen bei gleichem Preis) oder Sneakflation (veränderte Zusammensetzung bei gleichbleibendem Preis) oder die Quantifizierung eines hedonistischen „Preisgefühls“ (CPI hedonic demand) berücksichtigt werden müssen. Klassische Verbraucherpreisindizes erfassen solche Effekte oft unzureichend, da sie sich primär auf nominelle Preise stützen. Hier setzen Composite-Indikatoren an, die Preis, Menge, Qualitätsanpassungen und Konsumverhalten in ein integratives Maß überführen. Durch die Kombination amtlicher Preisdaten mit ergänzenden Informationen – etwa aus Scannerdaten oder Produktspezifikationen – können verdeckte Inflationsformen sichtbar gemacht und realistischere Kaufkraftentwicklungen abgebildet werden.

Jedes Produkt – samt Preis, Packungsgröße, Komponenten oder Nährwertprofil – könnte hierzu als FDO modelliert und mit standardisierten Metadaten versehen werden. Dadurch wird es möglich, Veränderungen nicht nur im Preis, sondern auch in Menge und Qualität systematisch zu erfassen und über Zeiträume hinweg zu vergleichen. Für die Berechnung komplexer Composite-Indikatoren bedeutet das: Daten aus unterschiedlichen Quellen (amtliche Statistik, Scannerdaten, Unternehmensangaben) lassen sich interoperabel verknüpfen, wodurch verdeckte Preisstrategien sichtbar und robust in Indizes integriert werden können. Hierzu sollte ein Pilotprojekt mit DESTATIS und dem Statistischen Landesamt Bayer gestartet und gefördert werden. Entsprechende Vorgespräche haben bereits stattgefunden.

Verweise:
↗ Kommission Zukunft Statistik Abschlussbericht
↗ I-ADOPT in a nutshell
↗ Competence Centre on Composite Indicators and Scoreboards

USE CASE – MODELLIERBARE WIRKUNGSORIENTIERTE HAUSHALTSPLANUNG

Der Abschlussbericht für das Bundesminsterium der Finanzen über die Erstellung eines Konzepts zur Einführung einer ziel- und wirkungsorientierten Haushaltsführung des Leibniz-Zentrums für Europäische Wirtschaftsforschung zeigt, dass die Entwicklung eines modellierbaren Haushalts eng mit der Verfügbarkeit und Qualität von Daten verknüpft ist. Für eine wirkungsorientierte Haushaltsführung braucht es nicht nur Kennzahlen und Indikatoren, sondern vor allem belastbare Datenquellen, die aus Verwaltung, amtlicher Statistik und Evaluationen stammen. Dabei spielen Datenlabore und deren Vernetzung eine zentrale Rolle: Sie schaffen interministerielle Datenpools, die eine konsistente Indikatorenbildung ermöglichen und die Grundlage für evidenzbasierte Entscheidungen im Haushalt legen.

Diese Zielsetzung lässt sich eng mit den Analysen aus Kapitel 6 des Jahresgutachtens 23/24 des Sachverständigenrats verknüpfen, in dem der Aufbau einer zeitgemäßen Dateninfrastruktur für fundiertere Entscheidungen eingefordert wird. Das Jahresgutachten unterstreicht, dass unzureichende Metadaten und Dokumentationen die Nutzung bestehender Daten erschweren. Das deckt sich mit der Forderung des Abschlussberichts, Haushalts- und Wirkungsdaten nicht nur zu sammeln, sondern systematisch aufzubereiten und zugänglich zu machen.

Eine Metadaten-reiche Dateninfrastruktur kann hier den entscheidenden Mehrwert bringen, indem sie Daten auffindbar, interoperabel und kontextualisierbar macht.

Die amtliche Statistik hat dabei eine Schlüsselfunktion: Sie liefert nicht nur methodisch geprüfte und standardisierte Daten, sondern kann durch Integration in solche Infrastrukturen ihre Rolle als neutrale Referenzquelle weiter stärken. Sie muss aber zudem von einer input- zu einer outputorientierten Logik weiterentwickelt werden, um flexibler auf Informationsbedarfe reagieren zu können.

FAIR Digital Objects (FDOs) können aufzeigen, wie ein moderner, modellierbarer (und lernender) Haushalt aussehen könnte: Daten zu Haushaltsplänen, Maßnahmen und Wirkungsindikatoren werden als digitale Objekte mit klar definierten Metadaten strukturiert, maschinenlesbar bereitgestellt und dauerhaft auffindbar gemacht. Im Ergebnis könnte der Haushalt nicht nur als juristisch-administratives Dokument, sondern als dynamisches Datenmodell verstanden werden, das Planung, Steuerung und Evaluation integriert und zugleich die Basis für innovative Analyse- und Partizipationsformate bietet. Das entspräche genau den in Kapitel 6 des Jahresgutachtens der Wirtschaftsweisen geforderten Mechanismen – von Datentreuhändern zur sicheren Verknüpfung bis hin zu einfacheren Zugangswegen für Forschung und Politik.

Zur wirkungsorientierten Steuerung sollte auch ein stärkeres Augenmerk auf die Erstellung von Strategien in der öffentlichen Verwaltung als semistrukturierte, maschinenlesbare Daten gelegt werden – zudem auch auf die Entwicklung von Metriken zu Objectives and Key Results (OKR). Inspirationen liefert ein Blick in die USA: Bereits seit dem GPRA Modernization Act aus dem Jahre 2010 müssen Behörden ihre „Strategic“ and Performance Plans and Reports" in maschinenlesbaren Formaten, wie zum Beispiel in StratML, verfassen.

Um FAIR Digital Objects für evidenzbasierte Entscheidungen und wirkungsorientierte, modellierbare Haushalte in Wert zu setzen, sollten sie als Teil der Public Digital Infrastructure institutionalisiert werden. Kurzfristig sollte ein Pilotprojekt starten, bei dem zunächst in ausgewählten Ressorts Haushaltsdaten nach Open-Data- u. FDO-Grundsätzen publiziert werden. Die bereits

STRATEGISCHE BETRACHTUNGEN_STRATEGISCHE ZUSAMMENFASSUNG UND ANREGUNGEN

installierten Datenlabore könnten mit Wissenschaft und Zivilgesellschaft gemeinsam und mit dem Statistischen Bundesamt oder einem Statistischen Landesamt testen, wie FDOs erzeugt und für Modellierungen und Evaluationen genutzt werden können.

Verweise:
↗ Abschlussbericht für das Bundesminsterium der Finanzen. Erstellung eines Konzepts zur Einführung einer ziel- und wirkungsorientierten Haushaltsführung
↗ Jahresgutachten Wirtschaftweise 23/24, Zeitgemäße Dateninfrastrukturen für fundierte Entscheidungen

USE CASE – LAW AS CODE

Die jüngste Modernisierungsagenda für Staat und Verwaltung (Bund) verspricht „Law as Code" (auch bekannt als „Rule as Code") für eine wirksame, effiziente und digital integrierte Rechtsetzung: Das Recht selbst soll vom Entwurf bis zum Vollzug digital werden. Das BMDS setzt dabei auf die Rulemapping-Methode (SPRIND-Projekt). Sie bildet Rechtsregeln visuell ab und übersetzt sie zugleich in ausführbaren Code. In Kombination mit KI-gestützten Werkzeugen, die die Struktur bestehender Regulierung analysieren und die Textarbeit unterstützen, lässt sich der digitale Politikzyklus der modernen Rechtsetzung wirkungsvoll schließen. Den Normadressaten sollen Gesetze und Entscheidungen als Open Source-Bibliothek zur Verfügung gestellt werden, damit diese unmittelbar Anwendungsprogramme übernommen werden können.

Hierzu soll zeitnah ein Open Source-Editor für die Nutzung von Rulemapping zur Verfügung gestellt werden. z. B. zur Anwendung von maschinenlesbarem Format (XML, RDF) und unter Nutzung von bspw. OpenFisca für Simulation/Plausibilität und XML-Formate wie Blawx bzw. AkomaNtoso (AKN4EU mit IMFC–Vokabular in der EU).

Bereits in 2024 ordnete die US-Regierung die vermehrte Anwendung eines Rule-as-Code-Vorgehen an und auch in Europa wird an einem Digital Legal System Framework (HyperModeLex) für „Digital Law-Making" gearbeitet. Das EU-finanzierte Projekt HyperModeLex wird einen soliden rechtstheoretischen Rahmen schaffen, um die Serialisierung von Gesetzen in einem maschinenlesbaren Format zu ermöglichen und gleichzeitig die rechtliche Robustheit zu wahren.

FAIR Digital Objects können hier bei dieser Zielsetzung des BMDS entscheidend beitragen: Jede einzelne Regel oder jeder Paragraf kann als FDO mit einem persistenten Identifikator (PID) versehen werden. Dadurch wird die Regel eindeutig zitierbar und versionierbar. Ändert sich das Gesetz, entsteht ein neues FDO, aber die alte Version bleibt referenzierbar. Die Metadaten eines FDOs können den ursprünglichen Gesetzestext, den Geltungsbereich, das Gültigkeitsdatum und andere juristische Kontexte enthalten. So bleibt die Verbindung zwischen dem Code und seiner rechtlichen Grundlage transparent und nachvollziehbar.

Damit würde auch ein Legal-Knowledge-Graph realer, der darauf abzielt, das komplexe Geflecht aus juristischem Wissen – bestehend aus Gesetzesnormen, Gerichtsentscheidungen, juristischen Konzepten und Akteuren – in einer strukturierten, maschinenlesbaren Form abzubilden.

Verweise:
↗ Modernisierungsagenda – für Staat und Verwaltung (Bund)

USE CASE – GOVDATA

GovData, das zentrale Open-Data-Portal des Bundes und der Länder, nutzt Linked Data-Prinzipien, um Verwaltungsdaten besser auffindbar, zugänglich, interoperabel und nachnutzbar zu machen. Die potenzielle Rolle von FAIR Digital Objects (FDOs) liegt in der Weiterentwicklung von GovData. Während DCAT-AP.de Metadaten beschreibt, könnten FDOs als „Container" für die Daten selbst dienen. Sie würden es ermöglichen, Daten und Metadaten fest miteinander zu verknüpfen und mit einem persistenten Identifikator zu versehen, um Daten maschinell verarbeitbar und Kontexte besser verstehbar zu machen.

Schleswig-Holstein nimmt bei Linked Open Data eine Vorreiterrolle ein und sollte als Entwicklungspartner adressiert werden.

USE CASE – FÖRDERFUNKE

Förderfunke ist ein Open-Source-Projekt, das vom Prototype Fund gefördert wird und darauf abzielt, Informationen über staatliche Leistungen und Fördermittel für Bürger leichter zugänglich zu machen. Anstatt Informationen in isolierten Datenbanken zu speichern, werden sie in einem sogenannten „Citizen-Knowledge-Graph" (Bürger-Wissensgraph) als Linked Data miteinander verknüpft. Das Projekt nutzt Technologien wie SHACL zur Datenvalidierung und SPARQL für Abfragen, um die Anspruchsvoraussetzungen für verschiedene Leistungen zu modellieren. Dazu überführt Förderfunke die komplexen Regeln und Voraussetzungen für staatliche Förderungen in ein maschinenlesbares und vernetztes Datenmodell. Dies ermöglicht es, individuelle Nutzungsprofile automatisiert mit den Anforderungsprofilen von Förderleistungen abzugleichen. So kann das System ermitteln, für welche Leistungen eine Person anspruchsberechtigt ist und welche Informationen dafür noch fehlen. Als FDOs könnte zudem jedes Förderinstrument und jedes Förderprojekt mit eindeutiger Kennung und standardisierten Metadaten beschrieben – von Laufzeit und Fördersumme über Ziele bis hin zu Indikatoren und Evaluationsberichten. Diese Objekte wären direkt mit Haushaltsdaten und relevanten Kennzahlen der amtlichen Statistik verknüpft. So ließen sich Förderlinien automatisiert vergleichen, Überschneidungen erkennen und Wirkungen evidenzbasiert messen. Für Antragstellende bedeutet das eine bessere Auffindbarkeit passender Programme, für Verwaltung und Politik ein Instrument zur Steuerung und Evaluation. Förderfunke könnte so zu einer modellhaften Umsetzung von FDO-Grundsätzen in der Förderpraxis werden.

USE CASE – OPEN-DATA-KATALOG PIVEAU

FAIR Digital Objects könnten die Katalog-Software „piveau" von Fraunhofer FOKUS bereichern, indem sie die bereits auf DCAT-AP und Semantic Web-Technologien basierende Architektur um standardisierte Mechanismen erweitern. Während piveau heute schon Metadaten strukturiert verwaltet, würde das FDO-Framework durch vorhersagbare Identifier-Auflösung und explizite Typ-Definitionen die Maschinenlesbarkeit auf ein neues Level heben. Besonders interessant ist die Komplementarität: Beide Ansätze nutzen RDF und semantische Standards, wodurch sich eine nahtlose Integration anbietet.

FDOs könnten die Qualitätssicherung von piveau durch validierbare Metadaten-Schemas verstärken und gleichzeitig die langfristige Verfügbarkeit und Interoperabilität in föderierten Datenökosystemen verbessern. Dies würde piveau als Open-Data-Lösung für öffentliche Verwaltungen stärken und die europäische Vision eines einheitlichen Datenraums im Rahmen des EOSC unterstützen, indem standardisierte Zugriffsmuster auf Datenobjekte, ihre Metadaten und Typ-Informationen etabliert werden.

GEMEINSAMES SCOPING MIT WIKIMEDIA DEUTSCHLAND

Die angestrebte Etablierung einer FDO-basierten Datenökonomie durch das FDO Connect-Projekt kann von einer technologischen Validierung auf Basis praxiserprobter, skalierbarer Linked-Data-Strukturen profitieren. Genau hier setzt ein strategisches Scoping mit Wikimedia Deutschland an, deren Wikidata Knowledge Graph das FDO-Paradigma bereits in Teilen vorlebt. Wikidata, als quelloffener und mehrsprachiger Wissensgraph mit persistenten Identifikatoren (Q-IDs), demonstriert eindrucksvoll die Machbarkeit von strukturierten, interoperablen Daten, die für maschinelles Lesen optimiert sind.

Die Mechanismen, mit denen Wikidata Quellenangaben und Zitate direkt in seine Datenobjekte integriert, bieten gute Anknüpfungspunkte mit dem FAIR Data Publisher Projekt, um die im Projekt zentrale lückenlose Rückverfolgbarkeit (Provenance) und die Einhaltung der FAIR-Kriterien weiter zu erproben. Darüber hinaus bieten die von Wikimedia entwickelten Architekturen zur Nutzung dieser Graphen über Vector Embeddings und Semantic Search (RAG-Systeme) konkrete technische Vergleichbarkeiten für die FDO-Service-Architektur. Eine solche Zusammenarbeit würde somit nicht nur die technologische Akzeptanz und die Standardisierung der FDO-Technologie signifikant beschleunigen, sondern auch die Überwindung der fundamentalen Datenfragmentierung im deutschen Datenraum erfolgreich vorantreiben.

GEMEINSAMES SCOPING MIT DER STADT MÜNCHEN ZUR SMART-DISTRICT-DATA-INFRASTRUCTURE (SDDI) IM CUT-PROJEKT

Die Herausforderung im Aufbau Urbaner Digitaler Zwillinge liegt nicht länger in der reinen Bereitstellung von Schnittstellen, sondern in der Sicherstellung einer tiefgreifenden semantischen Interoperabilität der Daten selbst. Die Smart-District-Data-Infrastructure (SDDI) bietet zwar den notwendigen architektonischen und Governance-Rahmen für verteilte Datenservices, stößt jedoch an ihre Grenzen, da die alleinige Nutzung von standardisierten APIs die tieferliegende Fragmentierung der Datenstrukturen und -bedeutungen nicht überwindet. Dies führt zu hohem Integrationsaufwand, geringer Übertragbarkeit von Anwendungen und weiterhin zu Dateninseln.

FAIR Digital Objects können hier ansetzen, indem sie die SDDI-Architektur auf der Ebene des Datensatzes ergänzen. Und zwar indem sie jedem Datenobjekt eine selbstbeschreibende, maschinenlesbare Struktur mit persistenten Identifikatoren (PIDs) verleihen. Die Integration von FDOs in die SDDI transformiert somit ein architektonisches Konzept in ein technisch abgesichertes, souveränes Datenökosystem.

STRATEGISCHE BETRACHTUNGEN_STRATEGISCHE ZUSAMMENFASSUNG UND ANREGUNGEN

EINBINDUNG IN EIN ZU SCHAFFENDES NATIONALES DATENINSTITUT

Die Etablierung von FAIR Digital Objects (FDOs) stellt einen strategischen Grundpfeiler für die deutsche und europäische Dateninfrastruktur dar, der unabdingbar für den Fortschritt der MISSION KI ist. Angesichts der fundamentalen Bedeutung dieser Technologie für die datensouveräne digitale Transformation bedarf es einer zentralen, nationalen Instanz – wie einem Nationalen Dateninstitut (NDI) – deren Beitrag für die Skalierung des FDO-Ökosystems und die weitere Entwicklung des Projekts FAIR Data Publisher unverzichtbar wäre.

Die Analyse internationaler Vorbilder zeigt eindeutig: Der Luxembourg National Data Service (LNDS) wurde 2022 als wirtschaftliche Interessenvereinigung von der luxemburgischen Regierung gegründet und ermöglicht Wertschöpfung aus der Sekundärnutzung öffentlicher Daten, während das Open Data Institute (ODI) 2012 von Sir Tim Berners-Lee und Sir Nigel Shadbolt gegründet wurde und transparent, rechenschaftspflichtig und innovativ mit Daten arbeitet und neue Use Cases vorantreibt. Beide Institutionen kombinieren technische Infrastruktur mit Kompetenzaufbau und strategischer Vernetzung – genau jene Funktionen, die ein deutsches Dateninstitut übernehmen sollte.

Das Luxemburger Modell zeigt exemplarisch, wie ein kleines Land durch konzentrierte Datenkompetenz internationale Bedeutung erlangt. Das LNDS-Serviceportfolio umfasst Know-how, Fähigkeiten, Werkzeuge, Infrastruktur und Datendienste zur Unterstützung von Datenentdeckung, Zugang, Anreicherung, Pseudonymisierung und Qualität. Deutschland mit seiner fragmentierten föderalen Struktur, diversen Forschungslandschaft und starken Wirtschaft benötigt ein solches Institut umso dringender als koordinierende und standardisierende Instanz.

Wie bereits im Umsetzungskonzept für ein Dateninstitut für die Bundesregierung in Bezug auf FAIR Digital Objects angerissen, muss ein nationales Dateninstitut vier Kernfunktionen erfüllen:

Erstens für die Bereitstellung technischer Infrastruktur für FAIR Digital Objects und Linked Data sorgen, einschließlich Ontologie-Registern und standardisierten Schnittstellen.

Zweitens die aktive Vernetzung der Akteure aus Wirtschaft, Wissenschaft, Verwaltung und Zivilgesellschaft durch Arbeitsgruppen, Communitys of Practice und Fachforen, wie in der Begleituntersuchung detailliert beschrieben.

Drittens eine Data-Stewardship und die Ausarbeitung von Anwendungsfällen sowie die Entwicklung und Pflege domänenspezifischer Datenstandards und die Integration in europäische Initiativen wie EOSC und Gaia-X.

Viertens sollte ein Nationales Dateninstitut die Mitgestaltung nationaler und internationaler Regulierung und Normierung unterstützen.

Die Untersuchung internationaler Dateninstitute und Use Cases – von GreenHome über Health Data Hubs bis zu Mobility Data Specifications – durch die Dateninstitut-Begleitforschung von BMWK und BMI demonstriert das enorme Potenzial datengetriebener Innovation. Aber auch, dass ohne koordinierende Institution diese Ansätze fragmentiert bleiben. Ein deutsches Dateninstitut könnte als Enabler fungieren, bestehende Projekte zu FAIR Digital Objects transformieren, Linked Data-Standards etablieren und damit echte semantische Interoperabilität ermöglichen.

Die Errichtung eines Nationalen Dateninstituts für Deutschland muss durch das BMDS weiter vorangetrieben werden. Die Risiken ohne ein Dateninstitut sind erheblich: Es droht eine weitere Perpetuierung isolierter Datensilos und ungehobener Datenschätze wie die Use-Case-Recherche für BMWK und BMI zeigt – viele innovative Projekte bleiben in proprietären Systemen gefangen und können nicht skalieren. Ohne eine zentrale Plattform für die Belange einer wertschöpfenden Datennachnutzung droht Deutschland weiter, international an Anschlussfähigkeit zu verlieren.

Verweise:
↗ Begleituntersuchung im Rahmen der Gründung eines Dateninstituts Deutschland Modul 2: Stakeholder- und Vernetzungsanalyse
↗ Begleituntersuchung im Rahmen der Gründung eines Dateninstituts Deutschland Modul 4: Rechercheergebnisse zu internationalen Use Cases und Pilotprojekten
↗ Luxembourg National Data Service (LNDS)
↗ Open Data Institute UK

IMPRESSUM

acatech – Deutsche Akademie der Technikwissenschaften e. V.

Postanschrift:
Karolinenplatz 4
80333 München

Kontakt:
Telefon: +49 (0)89/52 03 09-0
Telefax: +49 (0)89/52 03 09-900
E-Mail: info@acatech.de

Vertreten durch den Vorstand i.S.v. § 26 BGB:
Prof. Dr. Claudia Eckert, Prof. Dr.-Ing. Thomas Weber, Wolfgang Siegel

Copyright:
acatech – Deutsche Akademie der Technikwissenschaften e. V.
Lizenz:
↗ CC BY 4.0
Bildlizenzen S. 1, 42, 49, 56, 63, 67: iStock

Projektleitung bei acatech – Deutsche Akademie der Technikwissenschaften e. V. (MISSION KI):
Marius Bensley, Philippe Calvez

Projektumsetzung:
Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG)
Projektleitung bei GWDG:
Sven Bingert, Sonja Sternkopf

Strategische Konzeption, Beratung, Redaktion, Gestaltung & Produktion:
Oliver Rack, Intellectual Services & Strategy

Leitung der Arbeitspakete:
AP1 – Entwicklung einer verteilten FDO-Service-Architektur auf Basis von Nanopublikationen:
Tobias Kuhn, Knowledge Pixels AG

AP2 – Entwicklung eines Software-Tools zur Generierung von FDOs aus unstrukturierten Daten:
Henrik tom Wörden, IndiScale GmbH

AP3 – Entwicklung von Anwendungsfällen für die Nutzung von FDOs in ausgewählten Domänen:
Hans-Günther Döbereiner, Universität Bremen

AP4 – Einbindung der Projektergebnisse in die Landschaft bestehender technischer Lösungen und Standards:
Oliver Rack, Intellectual Services & Strategy

Mitwirkende
(in alphabetischer Reihenfolge):

Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI):
Christian Backe, Patrick Draheim, Veit Briken

Deutsches Institut für Normung e. V.:
Anna Petschulies

FIZ Karlsruhe – Leibniz-Institut für Informationsinfrastruktur:
Fabian Rack, Franziska Boehm

Frankfurt School of Finance and Management:
Jan Nagler

Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e. V.:
Christina Gillmann, Christoph Lange-Bever, Daham Mustafa, Joshua Gelhaar, Maik Mannsfeld, Michael Steinert, Zeyd Boukhers

Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen:
Jana Böhm, Philipp Wieder, Sonja Sternkopf, Sven Bingert

Hochschule Hannover:
Suat Can

IndiScale GmbH:
Henrik tom Wörden, Insa Nüske, Sina Rohde, Timm Fitschen

Oliver Rack, Intellectual Services & Strategy

Karlsruher Institut für Technologie:
Frank Tristram

Knowledge Pixels AG:
Ashley Caselli, Philipp von Essen, Tobias Kuhn, Virginia Balseiro, Ziroli Plutschow

RWTH Aachen:
Igor Garmaev, Marius Politze, Torben Miny

SPIRIT LEGAL Rechtsanwaltsgesellschaft mbH:
David Wagner

Universität Bremen:
Hans-Günther Döbereiner, Lennard Scheurer, Lukas Schmidt, Malte Ohmstede, Robert Porzel, Torben Klarl

Universität der Bundeswehr München:
Prof. Dr. Mario Martini (Inhaber des Lehrstuhls für Recht und Digitalisierung an der Universität der Bundeswehr München, Leiter des Themenbereichs „Digitale Transformation im Rechtsstaat“ am Deutschen Forschungsinstitut für öffentliche Verwaltung), Tom Schuchort

YottaSen Forschungs-GmbH:
Nils Niendorf

↗ FAIR DATA PUBLISHER