

# EMPFEHLUNGEN ZU OBJEKTORIENTIERTEN DATENINFRASTRUKTUREN UND FAIR DIGITAL OBJECTS (FDO) FÜR DIE 21. LEGISLATURPERIODE

## FORMULIERUNGSHILFEN

- Wir unterstützen die Weiterentwicklung von Internet und WWW von einem Netz der Datenübertragung, über ein Netz der Dokumente zu einem Netz der Daten, Register und Bedeutungen. Um dieses wie bisher skalierbar zu halten, unterstützen wir die Entwicklung eines "Global Integrated Data Space" (GIDS) aus atomaren und komplexen, vertrauenswürdigen, persistenten und maschinenlesbaren Informationseinheiten.
- Wir setzen uns für Technologien und Dateninfrastrukturen ein, die vertrauenswürdige Datenräume und qualitätsgesichertes Wissen ermöglichen.
- Wir fördern die Entwicklung von Dateninfrastrukturen (wie z.B. Linked (Open) Data und FAIR Digital Objects) und offenen, verallgemeinerten Datenstandards als Fundament für vertrauenswürdige Datenräume. Dieser Standard gewährleistet die FAIR-Prinzipien sowie semantische Interoperabilität und trägt somit zur Qualitätssicherung künstlicher Intelligenz, zur Realisierung digitaler Produktpässe sowie zum Aufbau einer leistungsfähigeren Forschungsdateninfrastruktur für Deutschland und Europa bei.
- Wir leisten als Bundesrepublik Deutschland einen Beitrag beim Aufbau eines Global Integrated Data Space (GIDS).
- Wir fördern semantische Interoperabilität nach FAIR-Prinzipien, um Datenräume, Prozesse und Wissen domänenübergreifend zu verknüpfen.
- Wir fördern Dateninfrastrukturen und Institutionen, die Datenqualität und Orientierungswissen für evidenzbasierte Entscheidungen und vertrauenswürdige KI sicherstellen.
- Wir fördern die Entwicklung eines verallgemeinerten Datenstandard, mit dem sich konkrete Sachverhalte, Beziehungen zwischen Rechtssubjekten und tatsächliche Lebensverhältnisse strukturiert in maschinenlesbare Form übertragen lassen, um sowohl künstlicher Intelligenz in der Verwaltung eine solide Grundlage zu bieten als auch das Once-Only-Prinzip wirksam umzusetzen.

## ERLÄUTERUNG

Deutschland und die EU steuern in eine schwerer als bisher absehbare Zukunft, mit der Gefahr von Souveränitätsverlusten, Wohlstandsverlusten und Verteilungskonflikten. Einem wirtschaftlichen und intellektuellen Abstieg muss Deutschland als informierte Gesellschaft entgegenen und mit Daten-Instrumenten **evidenzbasiert** und mit **Orientierungswissen** in eine gute Zukunft navigieren. Dazu muss Deutschland endlich in einer Daten- und KI-Realität ankommen. Das unterstreicht auch das Kapitel 6 aus dem Jahresgutachten 23/24 der Wirtschaftsweisen.

Hierzu müssen mehr als zuvor Datennutzung und Datendemokratie (siehe D21 e.V.) vorangetrieben werden. Dazu gehören unter anderem die Umsetzung europäischer und nationaler Datenstrategien und Datenregulierungen, die gezielte intelligente Vernetzung des deutschen und europäischen Datenökosystems, der Ausbau vertrauenswürdiger Datenräume und **Datentreuhänder** sowie die Förderung des Aufbaus und Betriebs **objektorientierter Dateninfrastrukturen** (wie z.B. Linked (Open) Data und FAIR Digital Objects), welche Persistenz, Nachvollziehbarkeit, reiche Metadaten und Kontexte ermöglichen. Insbesondere hervorzuheben ist hier die Anwendung bei künstlicher Intelligenz, digitalen Produktpässen, interdisziplinären Forschungsdateninfrastrukturen oder der Modernisierung von Registern.

Im Rahmen der rasanten Entwicklung von neuronalen Netzen und den damit verknüpften Herausforderung der systemischen Transparenz sind für eine nachhaltige und vertrauensvolle Nutzung von KI besonders folgende Faktoren notwendig: Institutionen der unabhängigen und qualitätsgesicherten Wissensproduktion, qualitätsgesicherte Daten sowie intelligent vernetzte und objektorientierte Dateninfrastrukturen, um kanonische Zusammenhänge und eine "Ground Truth" für maschinelles Lernen systematisch sicherzustellen.

An diesem derzeitigen neuralgischen Kippunkt bei der Entwicklung des Welt- und Domänenwissens können objektorientierte Dateninfrastrukturen einen wichtigen Beitrag leisten.

## Gründe für objektorientierte Dateninfrastrukturen

Weiterentwicklung von Internet und WWW von einem Netz der Datenübertragung, über ein Netz der Dokumente zu einem Netz der Daten, Register und Bedeutungen und um dieses skalierbar zu halten: einem "Global Integrated Data Space" (GIDS) aus atomaren und komplexen, vertrauenswürdigen, persistenten und maschinenlesbaren Informationseinheiten.

Die Entwicklung des Global Integrated Data Space (GIDS) soll mit einer geschichteten Architektur erfolgen, um die Komplexität zu bewältigen, vergleichbar mit dem OSI-ISO Modell des Internets. Die Motivation hierfür liegt in der zunehmenden Abhängigkeit der Gesellschaft von vertrauenswürdigen und hochwertigen Daten und der Notwendigkeit, Barrieren beim Datenaustausch zu überwinden. Im Gegensatz zum Internet, bei dem nur Pakete ausgetauscht werden, geht es beim GIDS um bedeutungsvolle Objekte (Digitale Objekte bzw. 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.

Die Objekte sind durch globale, eindeutige, auflösbare und persistente Identifikatoren (PID) identifizierbar und werden zu einem PID Record aufgelöst. Dieser PID Record beinhaltet den Typ des Objekts sowie Mechanismen, um dessen Bitsequenzen, Metadaten und Verweise auf programmatische Operationen abzurufen. Auf diese Weise wird ein "Signposting" ermöglicht, das die Auflösung der Metadaten bedeutungssicher und bedarfsorientiert erlaubt. Die relevanten und kontextuellen Daten können somit als "Crates" gebündelt werden.

Deswegen müssen die Objekte den FAIR-Prinzipien folgen und als FAIR Digital Objects (FDO) alle für die FAIR-Verarbeitung notwendigen Informationen bündeln. Sie stellen ein Schalen- bzw. Mehrebenen-Modell dar, in dem je nach Schicht weiterführende Informationen beim Auflösen der jeweiligen Ebenen abgerufen werden können.

Der Typ eines FDO definiert Attribute, um solche FDO semantisch zu beschreiben und mit anderen Konzepten zu verknüpfen. Die systematische Entwicklung eines Ökosystems solcher Objekte soll Nutzern die Organisation komplexer digitaler Entitäten mit zugehörigen Identifikatoren, Metadaten und automatisierter Verarbeitung ermöglichen.

Interoperabilität zwischen verschiedenen Datenräumen und Repositorien wird durch FDO-Adapter ermöglicht, ohne dass Änderungen an bestehenden Systemen notwendig sind. Das einheitliche Digital Object Interfacing Protocol (DOIP) der DONA Foundation ist der minimale Vereinheitlichungsmechanismus, der den weltweiten Zugriff auf alle FDO ermöglicht.

Ziel ist es zudem, FDO nicht nur als passive Objekte zu betrachten, sondern künftig auch als aktive Einheiten mit zugeordneten Operationen einzusetzen.

## Unterschiede zwischen FAIR Digital Objects (FDO) und Linked (Open) Data

Ein zentraler Unterschied liegt in der **Struktur und den Anforderungen**. Im Gegensatz zu FDO wird **Linked Data** als ein etabliertes Set von **Prinzipien** basierend auf Semantic Web Technologien betrachtet, das die Vision der FAIR-Prinzipien erreichen kann. Der Fokus von Linked Data liegt jedoch auf der **Wiederverwendung existierender URIs** anstelle der Schaffung neuer Identifikatoren und betont den Aufbau **navigierbarer Datenressourcen** anstelle elaborierter Graphdarstellungen.

Ein weiterer Unterschied wird in der **Akzeptanz und Implementierung** gesehen. Forscher und Entwickler neuer Plattformen für Computer und Datenmanagement zögern oftmals, einen vollständigen **FAIR Linked Data-Ansatz** zu adaptieren. Dies mag an der Heterogenität und der scheinbaren Komplexität der Semantic-Web-Ansätze liegen. Stattdessen bevorzugen sie oft eigene Modelle und JSON-basierte Formate von RESTful Web Services.

FDO jedoch erfordern **JSON** als primäres Syntax-Format, insbesondere bei der Realisierung mit Digital Object Interfacing Protocol (DOIP), während das Web (und damit auch Linked Data) eine Vielzahl von syntaktischen Formaten zulässt. Semantisch verlangt FDO, dass jedes Objekt einen **wohl-definierten Typ und eine strukturierte Form** besitzt, was im Web nicht zwingend der Fall ist.

Ein wichtiger Punkt ist auch die **Grundlage der Identifizierung**: FDO legt einen **starken Fokus auf die Verwendung von PIDs**, während Linked Data **HTTP-URLs** als Basis für Identifikatoren verwendet, deren Persistenz nicht immer gewährleistet ist.

## ANWENDUNGSBEISPIELE

### FAIR Digital Objects (FDO) für vertrauenswürdige Datenräume

Vertrauenswürdige Datenräume entstehen aus allgemeinen Datenräumen, in denen Prinzipien umgesetzt werden, die die Vertrauenswürdigkeit sowie den Datenaustausch zwischen verschiedenen Rechteinhabern, Rollen und Disziplinen gewährleisten. FDO können solche Prinzipien transportieren und maßgeblich zur Einhaltung von Compliance-Anforderungen beitragen. Durch die Standardisierung von Datenformaten und -beschreibungen

erleichtern FDO zudem die Interoperabilität zwischen unterschiedlichen Systemen und fördern so die Einhaltung von regulatorischen und vertraglichen Anforderungen, Anforderungen an die Datenqualität oder Anforderungen an die Souveränität.

## FAIR Digital Objects (FDO) bei der Nutzung für Recht

FDO erleichtern die **Einbettung von Rechtsmetadaten** wie Lizenzen, Nutzungsbeschränkungen und IP-Rechten in standardisierte Schemata (z.B. Dublin Core oder DataCite), automatisieren die Verbreitung von Rechtsbedingungen und machen sie für Nutzer zugänglich. Maschinenlesbare Lizenzen, etwa Software Package Data Exchange (SPDX) oder Creative Commons Rights Expression Language (ccREL) in FDO, ermöglichen dabei die automatische Überprüfung und Durchsetzung der Datennutzung, wodurch das Risiko eines unbeabsichtigten Missbrauchs reduziert wird. Gleichzeitig gewährleistet die **Zugangskontrolle und Authentifizierung** durch die Verknüpfung von FDO mit Systemen wie OAuth, OpenID Connect (OIDC) oder Shibboleth über persistente Identifikatoren (PID) die Einhaltung von Datenschutzgesetzen wie der DSGVO.

Die **Rückverfolgbarkeit und Provenienz** von Daten wird durch detaillierte Aufzeichnungen des Ursprungs und der Nutzungshistorie mittels Metadatenstandards wie PROV-O sichergestellt. Dies ermöglicht nicht nur eine rechtliche Rechenschaftspflicht, sondern auch umfassende Audits. Darüber hinaus wird durch die Nutzung globaler Metadatenstandards und persistenter Identifikatoren eine **Interoperabilität über Gerichtsbarkeiten hinweg** erreicht. Harmonisierte Datenfreigabepraktiken erleichtern internationale Kollaborationen und vereinfachen die rechtliche Handhabung über verschiedene Rechtssysteme hinweg.

Ein weiteres zentrales Element ist die **automatisierte Durchsetzung von Richtlinien**, beispielsweise bei Embargos oder geografischen Beschränkungen. Durch die Kopplung von FDO mit regelbasierten Engines, die Metadaten auslesen und entsprechende Regeln automatisch durchsetzen, wird sichergestellt, dass rechtliche Vorgaben eingehalten werden. Ergänzend dazu ermöglicht die **dynamische Aktualisierung von Rechtsbedingungen** mittels versionskontrollierter Metadaten eine kontinuierliche Anpassung an neue rechtliche Rahmenbedingungen.

## FAIR Digital Objects (FDO) bei der Nutzung für Registermodernisierung

Es besteht ein Potenzial, wie FDO das Föderale Informationsmanagement (FIM) weiterentwickeln können. Die objektbasierte Natur von FDO und die Nutzung von PIDs und standardisierten Metadaten könnten die starren Katalogstrukturen von FIM ergänzen und die Flexibilität und Zukunftsfähigkeit des föderalen Informationsmanagements erhöhen. FDO könnten helfen, bestehende FIM-Datenstrukturen effizienter zu organisieren: beispielsweise durch den Einsatz von PID für Verwaltungsdaten, semantische Verknüpfungen und besser maschinenlesbare Prozesse. Als verallgemeinerte Register-Architektur, quasi als Single Digital Gateway (SDG), könnten FDO die effiziente Umsetzung von Once-Only voranbringen.

Beispielsweise definiert der Baustein Datenfelder des FIM alle benötigten Datenstrukturen und statischen Inhalte für Verwaltungsleistungen und zugehörige Dokumente. Er nutzt den Datenfeldkatalog und das FIM-Stammdatenschema.

Der darin enthaltene XDatenfelder-Standard standardisiert dabei die Erfassung, Bereitstellung und interoperable Nachnutzung dieser Datenstrukturen. Die semantische Interoperabilität, die Ende-zu-Ende-Digitalisierung und das Once-Only-Prinzip werden durch einheitliche Beschreibung und Registrierung von Datenstrukturen angestrebt.

Während FIM als begrenzt flexibel gilt und kontinuierliche Anpassungen benötigt, bieten FDO eine dynamische, anpassbare Struktur für verschiedene Domänen.

Während FIM weiterhin für prozessuale Standards genutzt werden soll, könnten FDO eine datengetriebene Interoperabilität sicherstellen. Mittel- bis langfristig könnten FDO in einer neuen Version oder Weiterentwicklung von FIM integriert werden und als ein "Metastandard" dienen, der diese Silos übergreifend zusammenführt und den Datenaustausch ermöglicht. FDO-Technologien könnten somit allgemeine Verwaltungsprozesse datengetrieben und plattformübergreifend automatisieren.

## REFERENZEN

<https://fairdo.org/> (wegen Umzug erst ab 21.3. wieder online)

<https://peerj.com/articles/cs-1781/>

Das Konzept der Digitalen Objekte geht auf die Arbeit von Robert Kahn et al zurück: "A Framework for Distributed Digital Object Services"

[https://www.cnri.reston.va.us/tmp\\_hp/k-w.html](https://www.cnri.reston.va.us/tmp_hp/k-w.html)

## UNTERSTÜTZER

### FDO FORUM

<https://fairdo.org/> (wegen Umzug erst ab 21.3. wieder online)

Ansprechpartner: Sven Bingert, Peter Wittenburg (auch Mitglieder des Executive Committee von FDO Forum)

[Peter.Wittenburg@mpi.nl](mailto:Peter.Wittenburg@mpi.nl), [sven.bingert@gwdg.de](mailto:sven.bingert@gwdg.de)

### FDO4DE

<https://fdo4de.de/>

Nutzer- und Use-Case-Initiative für FAIR Digital Objects in Deutschland

Ansprechpartner: Oliver Rack (Sprecher), [oliverrack@posteo.de](mailto:oliverrack@posteo.de)

### FDO ONE

<https://fdo-one.org/>

FDO One zielt darauf ab, seine Datenökonomie durch die Förderung der internationalen Zusammenarbeit über den Global Integrated Data Space voranzutreiben.

Ansprechpartner: Philipp Wieder, [philipp.wieder@gwdg.de](mailto:philipp.wieder@gwdg.de)

### NANOPUBLICATIONS

<https://nanopub.net/>

Stakeholder-übergreifende Gruppe zur Spezifikation und Implementierung von Nanopublikationen

Ansprechpartner: Tobias Kuhn, [tobias@knowledgepixels.com](mailto:tobias@knowledgepixels.com)