

Stellungnahme der DINI AG Forschungsinformationen & Systeme (FIS) zur Online-Konsultation ‚Fahrplan für Open Data‘ des BMDS

BEITRAG ZUR WEITERENTWICKLUNG DES STRATEGISCHEN RAHMENS FÜR OPEN DATA

Vorbemerkung

Open (Government) Data umfassen aus unserer Sicht insbesondere Forschungsinformationen öffentlich finanzierter Wissenschaftsorganisationen. Darunter sind Metadaten zu verstehen, die Forschungsaktivitäten und -ergebnisse und deren Kontext beschreiben. Dazu gehören beispielsweise Informationen zu Forschungsprojekten, Publikationen (u.a. Text, Forschungsdaten, Software), beteiligten Akteuren und Einrichtungen sowie Forschungsinfrastrukturen.

Auch international gewinnt die offene Bereitstellung von Forschungsinformationen zunehmend an Bedeutung. Die 2024 veröffentlichte *Barcelona Declaration on Open Research Information*¹, die inzwischen von rund 150 Organisationen aus 35 Ländern unterzeichnet wurde, fordert, Informationen über die Durchführung und Kommunikation von Forschung offen zugänglich und nachnutzbar zu machen und die dafür notwendigen Infrastrukturen zu stärken. Zugleich setzen Initiativen wie DORA² und CoARA³ darauf, wissenschaftliche Leistungen breiter und differenzierter abzubilden und die Abhängigkeit von einzelnen publikations- und zitationsbasierten Kennzahlen zu verringern. Offene, qualitativ hochwertige Forschungsinformationen bilden hierfür ebenfalls eine wichtige Grundlage.

Die *DINI AG Forschungsinformationen & Systeme* (FIS) begleitet die Entwicklungen rund um Forschungsinformationen und Forschungsinformationssysteme in Deutschland seit mittlerweile 15 Jahren und hat sich als zentraler Knotenpunkt für den institutionsübergreifenden Austausch und die Vernetzung der FIS-Community etabliert. Im Mittelpunkt stehen dabei insbesondere Standards und Interoperabilität sowie praktische Fragen zum Betrieb und zur Weiterentwicklung von Forschungsinformationssystemen. Die DINI (Deutsche Initiative für Netzwerkinformationen e.V.) selbst gehört zu den Unterzeichner:innen der Barcelona Declaration und unterstützt ausdrücklich das Ziel, Forschungsinformationen offen, nachhaltig und nachnutzbar bereitzustellen. Mit der *FIS-Landkarte*⁴, die in Kooperation mit der *Kommission für Forschungsinformationen in Deutschland* (KFiD) entwickelt wurde, bietet die DINI AG FIS zudem nicht nur ein Informations- und Vernetzungstool, sondern auch eine aktuelle Übersicht der an deutschen Hochschulen und Forschungseinrichtungen eingesetzten und geplanten Forschungsinformationssysteme.

¹ <https://barcelona-declaration.org>

² <https://sfdora.org>

³ <https://www.coara.org>

⁴ https://kfid-online.de/fis_landkarte_visual

Die folgenden Anmerkungen zur Weiterentwicklung des Open-Data-Rahmens erfolgen aus dieser Perspektive und richten den Blick insbesondere auf die offene, interoperable und nachhaltige Zugänglichkeit und Nachnutzung von Forschungsinformationen:

1. Welche zentralen Herausforderungen und Chancen sehen Sie aktuell für die Bereitstellung und Nutzung von Open Data in Deutschland? An welchen Indikatoren machen Sie dies fest?

Aus Sicht der DINI AG Forschungsinformationen & Systeme zeigen sich die zentralen Herausforderungen von Open Data derzeit besonders deutlich im Bereich der öffentlich geförderten Forschung. Informationen zu Forschungsprojekten liegen in Deutschland nach wie vor stark fragmentiert vor: Systeme wie PROFI⁵ und FOEKAT⁶ des Bundes, GEPRIS⁷ und GERi⁸ der DFG sowie zahlreiche weitere institutionelle Repositorien existieren nebeneinander, ohne dass eine systematische Zusammenführung stattfindet. Verschärft wird diese Zersplitterung durch anhaltende rechtliche Unsicherheiten, da Metadaten zu Projekten und Forschungsergebnissen häufig nicht eindeutig lizenziert sind. Eine verbindliche CC0-Kennzeichnung fehlt in der Breite, was die automatisierte Weiterverwendung dieser Daten faktisch verhindert. Hinzu kommt ein Mangel an durchgängiger Interoperabilität: Die Außerachtlassung etablierter Standards wie des KDSF⁹ (Standard für Forschungsinformationen in Deutschland) oder des europäischen Formats CERIF¹⁰ erschwert sowohl die nationale Aggregation als auch den Abgleich mit internationalen Infrastrukturen wie OpenAIRE¹¹ und schränkt damit die Nutzbarkeit der Daten für Forschungseinrichtungen erheblich ein. Die Integration und Verknüpfung von Forschungsinformationen wird zudem dadurch erschwert, dass insb. im Kontext der Forschungsförderung kaum etablierte Systeme zur persistenten Identifikation bspw. von Projekten, Förderungen, Förderprogrammen etc. genutzt werden. Zudem sind die Daten meist nur für die menschliche Interaktion aufbereitet in Webportalen verfügbar und nicht über technische Schnittstellen in maschinenlesbarer Form zugänglich.

Eine zunehmend drängende, aber bislang wenig adressierte Herausforderung stellt der Umgang mit KI-Bots und automatisierten Zugriffen dar: Um zu gewährleisten, dass offene Daten auch tatsächlich offen bleiben, bedarf es geeigneter technischer Werkzeuge. Derzeit reagieren viele Anbieter auf die Gefahr automatisierter Angriffe jedoch eher mit einer präventiven Abschottung ihrer Systeme, was dem eigentlichen Open-Data-Gedanken zuwiderläuft. Eng damit verknüpft ist die Frage der KI-gestützten Metadatenkuratierung: Der Einsatz von KI zur Extraktion, Vervollständigung, Anreicherung oder Klassifikation von Metadaten birgt erhebliches Potenzial,

⁵ <https://foerderportal.bund.de/profionline/>

⁶ <https://foerderportal.bund.de/foekat>

⁷ <https://gepris.dfg.de>

⁸ <https://gerit.org>

⁹ <https://kerndatensatz-forschung.de/>

¹⁰ <https://eurocris.org/services/main-features-cerif>

¹¹ <https://www.openaire.eu>

gleichzeitig aber auch das Risiko von Halluzinationen und einer schleichenden Kontamination der gemeinsamen Metadatenallmende. Diesem Risiko sollte durch eine verpflichtende, transparente Kennzeichnung KI-generierter oder KI-angereicherter Inhalte begegnet werden, wie es etwa im Bereich der Softwareentwicklung mit entsprechenden KI-Policies bereits praktiziert wird.

Der Konflikt zwischen Sicherheit und Offenheit wird im Forschungskontext häufig überzeichnet: Die überwiegende Mehrheit der Forschungsmetadaten ist unkritisch, wird aber pauschal unter Verschluss gehalten, statt risikobasiert differenziert zu werden. Nötig ist daher kein genereller Schutzvorbehalt, sondern eine gezielte Inhaltsklassifikation, die echte Schutzinteressen von unbegründeter Vorsicht trennt. (Vgl. Frage 3.)

Strukturell fällt zudem auf, dass Deutschland in der Forschung zu Forschungsinformationen selbst zurückliegt, mangels gezielter Fördermöglichkeiten. Die dafür nötigen Informationsinfrastrukturen sind vielfach unterfinanziert und von kurzfristigen Projektmitteln abhängig, ihre Nachhaltigkeit ist damit strukturell gefährdet.

All diesen Herausforderungen stehen erhebliche Chancen gegenüber: Maschinenlesbare Forschungsmetadaten schaffen die Grundlage für KI-Training, Marktanalysen und Technologie-Monitoring und stärken damit die AI-Readiness des Forschungssystems. Eine größere Transparenz zu laufenden Projekten verbessert die Koordination zwischen Universitäten, außeruniversitären Einrichtungen und der Wirtschaft, ermöglicht neue Kooperationen und vermeidet Doppelforschung. Sie stärkt zugleich das epistemische Vertrauen der Gesellschaft in die Wissenschaft, da Bürgerinnen und Bürger nachvollziehen können, wofür öffentliche Forschungsmittel verwendet werden. Und nicht zuletzt entlastet die systematische Bereitstellung von Forschungsinformationen durch den Bund Forschende und Forschungsadministration gleichermaßen: automatisierter Import statt mehrfacher manueller Erfassung. Ein entscheidender Beitrag zum Bürokratieabbau.

Offene Forschungsinformationen sind zudem ein unmittelbarer Hebel für Wissens- und Technologietransfer. Wenn Unternehmen und Transferstellen strukturiert auf Projektmetadaten, Kompetenzprofile und Ergebnisdaten zugreifen können, verkürzen sich Anbahnungswege für Kooperationen erheblich, und Innovationspotenziale aus der Forschung erreichen die Anwendung schneller. Open Data wird damit vom Transparenzinstrument zum aktiven Standortfaktor – vorausgesetzt, die Daten sind nicht nur verfügbar, sondern auch in einer für Transferakteure nutzbaren Form aufbereitet.

2. Welche Rahmenbedingungen sollten aus Ihrer Sicht wie verändert werden, um eine wirksame Open-Data-Kultur in Deutschland zu schaffen; welche Themen und Handlungsfelder sollten in den kommenden Jahren priorisiert werden?

Um eine wirksame Open-Data-Kultur im Forschungssektor zu schaffen, sollten folgende Rahmenbedingungen priorisiert werden:

Ein zentrales Forschungsinformationsgesetz auf Bundesebene sollte einen Rechtsanspruch auf die Veröffentlichung von lizenzfreien Metadaten im Falle öffentlicher Forschungsförderung sicherstellen. Dieses Gesetz muss auch klare Regelungen für die Veröffentlichung von personenbezogenen Daten in Forschungsinformationen enthalten, beispielsweise zur Nennung von Projektleitenden oder Organisator:innen von wissenschaftlichen Konferenzen. Die Umsetzung dieses Vorhabens sollte in enger Abstimmung mit der Wissenschaftsministerkonferenz (GWK) erfolgen, um landesrechtliche Vorgaben mit der Bundesebene zu harmonisieren und bundesweit einheitliche Bedingungen für die Digitalisierung von Forschungseinrichtungen und den damit verbundenen Bürokratieabbau zu schaffen.

Darüber hinaus ist – wie in anderen Ländern (z.B. Finnland¹², Norwegen¹³, Niederlande¹⁴) schon produktiv – der Aufbau eines nationalen, eines Deutschen Forschungsinformationssystems (DE-FIS) anzustreben, dass erstmalig Einsichten ermöglichen würde, welche Forschung in Deutschland durchgeführt wird, welche Schwerpunkte bestehen, welche gefördert werden sollten. DE-FIS würde Transparenz in der Forschungsförderung gewährleisten und somit epistemisches Vertrauen in die Wissenschaft befördern. Weiterhin bestünde hiermit erstmals eine verlässliche Datenbasis für evidenzbasierte Wissenschaftspolitik. Eine zentrale Plattform dient hierbei als nationaler Knotenpunkt und erhöht die Sichtbarkeit und Nutzung von Forschungsinformationen auf nationaler und internationaler Ebene. Die Pflege der Forschungsinformationen kann zum Teil dezentral durch Zusammenarbeit der Forschungsförderorganisationen mit den Ländern und den Forschungseinrichtungen erfolgen.

Alle Forschungsinformationen, die im Rahmen öffentlich geförderter Forschung erstellt und veröffentlicht werden, sollten verpflichtend unter einer offenen Lizenz wie der international etablierten CC0 (Public Domain Dedication) in maschinenlesbarer Form und unter Nutzung etablierter persistenter Identifikationssystemen (PID) zur Verfügung gestellt werden. Sie erlaubt maximale Nachnutzbarkeit und fördert die Integration disparater Datenbestände.

Weiterhin sollten das Bundesministerium für Forschung, Technologie und Raumfahrt (BMFTR) und das Bundesministerium für Digitalisierung und Staatsmodernisierung die Barcelona Declaration on Open Research Information zu unterzeichnen und offiziell zu unterstützen. Dies gilt auch für alle forschungsfördernden Einrichtungen im Bund. Diese Institutionen sollten sich verpflichten, die Prinzipien offener Forschungsinformationen in ihr Förder- und Verwaltungshandeln zu integrieren.

3. Balance zwischen Transparenz und Sicherheit

Transparenz und Sicherheit sind im Umgang mit Forschungsinformationen kein Widerspruch. Maßgeblich sollte der Grundsatz „so offen wie möglich, so geschlossen wie nötig“ sein. Er ermöglicht es, die Vorteile offener Forschungsinformationen zu nutzen und zugleich berechnete

¹² <https://research.fi/en/>

¹³ <https://nva.sikt.no/>

¹⁴ <https://netherlands.openaire.eu/>

Schutzinteressen angemessen zu berücksichtigen.

Grundlegende Metadaten zu öffentlich geförderter Forschung, wie etwa Projekttitel, Laufzeit, Fördersumme, beteiligte Institutionen und – soweit rechtlich möglich bzw. freigegeben – Abstracts, sollten grundsätzlich offen zugänglich und frei nachnutzbar sein. Für Metadaten sollte daher eine offene Lizenzierung, also CC0, angestrebt werden. Das schafft Transparenz über öffentlich finanzierte Forschung und ermöglicht die institutions- und systemübergreifende Nachnutzung und Verknüpfung von Forschungsinformationen. Die strukturierte Bereitstellung von Forschungsinformationen trägt zugleich dazu bei, bestehende Transparenzpflichten der Bundesländer zu erfüllen, wie sie z.B. in den Transparenzgesetzen von Hamburg, Rheinland-Pfalz, Bremen, Thüringen und Sachsen bereits verankert sind.

Namen von Forschenden und verknüpfte Identifikatoren wie ORCID sind DSGVO-konform veröffentlichbar; darüberhinausgehende personenbezogene Informationen bleiben geschützt – anders als bei sicherheitsrelevanten Projekten geht es hier um Individualrecht, nicht um institutionellen Schutzbedarf.

Auch sicherheitsrelevante Forschungsprojekte können in Forschungsinformationssystemen erfasst werden, ohne dass die betreffenden Informationen / Metadaten öffentlich zugänglich sein müssen. Vielmehr sollten sie mit abgestuften Zugriffs- und Sichtbarkeitsregeln geschützt werden. Wo eine spätere Veröffentlichung vertretbar ist, können zeitlich begrenzte Embargos eingesetzt werden.

Neue Sicherheitsanforderungen sollten nicht dazu führen, die Vorteile von Open Data bzw. Open Science (insb. Transparenz, Nachnutzbarkeit und Reproduzierbarkeit) grundsätzlich einzuschränken/aufzugeben. Vielmehr gilt es, den Open-Science-Gedanken durch geeignete Risikomanagement- und Sicherheitsmaßnahmen weiterzuentwickeln. Ein risikobasierter Ansatz ermöglicht es, Offenheit dort zu erhalten, wo keine konkreten Schutzinteressen entgegenstehen und Einschränkungen gezielt dort vorzunehmen, wo sie tatsächlich erforderlich sind. Sicherheit und Offenheit sind damit keine Gegensätze, sondern Bestandteile eines verantwortungsvoll ausgestalteten Open-Science-Ansatzes.

4. Wie können Open-Data-Wertschöpfungsketten und die Weiternutzung von Daten stärker in den Blick genommen werden? Sollten spezifische Zielgruppen als Nutzer von Open Data adressiert werden? Wenn ja, welche und auf welche Weise?

Weiternutzung entsteht nicht von selbst, sie muss aktiv betrieben werden – zielgruppenspezifisch, mit jeweils passenden technischen Zugängen:

Für die Wirtschaft zählt vor allem der Zugriff auf Projektmetadaten, um Kooperationspartner zu identifizieren und Technologietrends frühzeitig zu erkennen. Entscheidend ist dabei nicht die Datenmenge an sich, sondern die Integrationsfähigkeit: Ohne dokumentierte, offene Schnittstellen (APIs) und standardisierte Daten bleibt jede Anbindung an betriebliche Informationssysteme ein

aufwendiges Einzelprojekt statt eines skalierbaren Regelfalls.

Zivilgesellschaft und Medien verfolgen andere Interessen: Wie die Kontrolle der Mittelverwendung und wissenschaftliche Bildung. Für diese Zielgruppe braucht es keine Rohdaten, sondern nutzerfreundliche, bedienbare Frontends und Dashboards, die komplexe Metadaten verständlich visualisieren. Offene Forschungsinformationen liefern zudem die Grundlage, um wissenschaftlichen Input für gesellschaftliche Diskussion und Dialog nutzbar zu machen und Expertinnen und Experten zu konkreten Projekten und Themen schneller zu identifizieren – etwa für Redaktionen, die im Rahmen von Wissenschaftskommunikation auf fundierte Einordnung angewiesen sind. Demokratische Gesellschaften sind auf offene, transparente Infrastrukturen angewiesen, um Vertrauen in staatliches und wissenschaftliches Handeln überhaupt erst herstellen zu können. Ohne diese Übersetzungsleistung und diesen infrastrukturellen Grundkonsens bleibt Transparenz ein Versprechen ohne Wirkung.

Die Wissenschaft selbst ist die naheliegendste, aber oft übersehene Zielgruppe: für Bibliometrie und Wissenschaftsforschung braucht sie Bulk-Downloads vollständiger, unter CC0 stehender Datensätze, die sich für Text- und Data-Mining eignen. Der eigentliche Mehrwert für Forschende liegt jedoch im Bürokratieabbau: Sind Projektinformationen einmal erfasst und offen nachnutzbar, entfällt die mehrfache manuelle Eingabe in verschiedene Systeme – Open Data wird hier zum Entlastungsinstrument für die eigene Community.

Zielgruppenadressierung allein genügt jedoch nicht, wenn die personellen Kapazitäten fehlen, Daten tatsächlich nutzbar zu machen und nachzunutzen. Wertschöpfung aus offenen Forschungsinformationen setzt eine gezielte Stärkung der Data Literacy voraus: Auf Seiten der Datenbereitstellenden ebenso wie auf Seiten der Nutzenden. Hierzu gehört der systematische Aufbau und die Etablierung von Berufsbildern wie Data Stewards und Data Scientists an Hochschulen und Forschungseinrichtungen, mit klaren Karrierepfaden statt befristeter Projektstellen. Nur wenn Kuratierung, Qualitätssicherung und Aufbereitung von Forschungsdaten als eigenständige, professionalisierte Tätigkeit anerkannt und entsprechend ausgebildet wird, lässt sich die Lücke zwischen technischer Verfügbarkeit und tatsächlicher Nachnutzung schließen. Diese Professionalisierung ist für eine Datengesellschaft die notwendige Voraussetzung.

5. Welche messbaren Zielvorgaben und -indikatoren sollte ein neuer „Open-Data-Fahrplan“ für Deutschland enthalten?

Folgende messbaren Zielvorgaben sollten im Open-Data-Fahrplan verankert werden:

Zielvorgabe: Alle Bundesministerien sowie durch den Pakt für Innovationen öffentlich geförderte Forschungsorganisationen wie die Deutsche Forschungsgemeinschaft (DFG) veröffentlichen ihre Förderinformationen in offener und dokumentierter API.

Indikatorik: Alle Bundesministerien und forschungsfördernden Einrichtungen im Pakt für Innovationen verfügen bis Ende 2027 über eine offene und dokumentierte API, über die ihre

erzeugten Forschungsinformationen in standardisierter und maschinenlesbarer Form veröffentlicht werden. Die Umsetzung wird durch das Bundesministerium für Digitales und Staatsmodernisierung (BMDS) koordiniert und gemonitort.

Zielvorgabe: Abfragen und Reportinganforderungen durch die oben genannten Bundesministerien und forschungsfördernden Einrichtungen werden einheitlich auf Basis des KDSF – Standard für Forschungsinformationen in Deutschland erhoben. Dies fördert die Integration der deutschen Open-Data-Landschaft im Forschungssektor und erleichtert die Berichterstattung durch die Forschungseinrichtungen.

Indikatorik: Alle Förderprogramme und Berichterstattungspflichten nutzen die Definitionen und das Austauschformat des KDSF. Eine Erhebung unter geförderten Einrichtungen wird durchgeführt, um die Entlastung zu messen. Ziel ist, dass mindestens 50 % der Einrichtungen verringerten Aufwand berichten.

Zielvorgabe: Ein Deutsches Forschungsinformationssystem (DE-FIS) ist bis Ende 2027 etabliert und enthält die Förderinformationen und damit verknüpften Output zu von mindestens 80 % der dem Bund zugehörigen fördernden Institutionen, mindestens jedoch BMFTR und DFG. DE-FIS verfügt über eine offene API, über die unbeschränkter Datenzugriff möglich ist.

Indikatorik: 100 % der in FOEKAT und GEPRIS enthaltenen Projekte sind in DE-FIS enthalten. Informationen zum Output der Projekte liegen für mindestens 50% der seit DE-FIS-Start abgeschlossenen Projekte vor. Alle öffentlichen Forschungseinrichtungen Deutschlands und mindestens 25.000 Forschende sind in DE-FIS erfasst.

Zielvorgabe: Das BMFTR setzt bis 2027 ein neues Förderprogramm zur Forschung und Entwicklung von Open Data im Forschungsinformationskontext auf. Dieses Programm ermöglicht technologische Innovationen, Forschung zur Datensicherheit und zur Interoperabilität. Eine Verschränkung mit der Programmatik im Bereich Wissenschafts- und Hochschulforschung ist sinnvoll.

Indikatorik: Mindestens 25 Projekte werden bis 2028 durch dieses Programm gefördert.

Zielvorgabe: Die vom PID-Network Deutschland erarbeiteten Empfehlungen für den Einsatz von persistenten Identifikatoren (PIDs)¹⁵ für eine nachhaltige und vernetzte Informationsinfrastruktur erhalten bis 2027 den Status einer offiziellen PID-Roadmap für Wissenschafts- und Kultureinrichtungen in Deutschland erhalten und werden bis 2030 operationalisiert.

Indikatorik: 2030 versehen alle öffentlichen Förderorganisationen ihre Daten zumindest zu Projekten, Förderungen, Förderformaten und beteiligten Einrichtungen und Personen mit entsprechenden persistenten Identifikatoren und stelle diese maschinenlesbar bereit.

¹⁵ <https://doi.org/10.5281/zenodo.21703674>

6. Gibt es konkrete Maßnahmen oder Best Practices, die in einen solchen „Open-Data-Fahrplan“ aufgenommen werden sollten?

Folgende konkrete Maßnahmen sollten im „Fahrplan für Open Data“ verankert werden:

- Schnittstellen-Pflicht für Bundesbehörden: Alle Organisationen des Bundes, die Forschungsförderung betreiben, müssen Metadaten aus Systemen wie PROFi (Antrags- und Durchführungsdaten) und PROMO (Programm-Monitoring) über eine API in maschinenlesbarem Format und unter freier Lizenz bereitstellen.
- Pilotprojekt „Forschungsinformations-Aggregator“: Bereitstellung von Mitteln für ein Pilotprojekt, in dem Projektmustadaten von bundesweit tätigen Förderern erstmals aggregiert und öffentlich zugänglich gemacht werden.
- Förderung von digitaler Infrastruktur: Förderprogramme des BMFTR und des BMDS für Forschungsinformationssysteme müssen aufgestellt werden, um Forschungseinrichtungen und Forschungsinfrastruktureinrichtungen bei der technischen Umsetzung der Datenherstellung, -kuratierung und -bereitstellung zu unterstützen.
- Der KDSF wird für alle Berichtsorderungen des Bundes verwendet. Bundeseigene Veröffentlichungen von Forschungsinformationen (z.B. FOEKAT) werden auch dem KDSF angeglichen.
- Das BMFTR fördert die in den Zielvorgaben beschriebene Förderlinie zu Forschungsinformationen.