

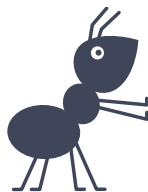
11

Working Paper
2025

KonsortSWD 
NFDI4Society

Empfehlung des FDI Ausschusses für eine einheitliche Bezeichnung von Datenprodukten der Forschungsdatenzentren

Daniel Fuß, Friederike Schlücker



Februar 2025

www.konsortswd.de

Empfehlung des FDI Ausschusses für eine einheitliche Bezeichnung von Datenprodukten der Forschungsdatenzentren

Daniel Fuß¹, Friederike Schlücker¹

Unter Mitwirkung von:

Daniel Buck², Hanna Brenzel³, Jan Goebel⁴, Marie-Christine Laible⁵,
Laura Menze⁶, Dana Müller⁷, Dagmar Pattloch⁶, Katharina Werhan⁸

Februar 2025

¹ Leibniz-Institut für Bildungsverläufe (IfBi)

² Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW)

³ Statistisches Bundesamt (Destatis)

⁴ Sozioökonomisches Panel (SOEP) am Deutschen Institut für Wirtschaftsforschung e.V. (DIW)

⁵ Bundesamt für Migration und Flüchtlinge (BAMF)

⁶ Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA)

⁷ Institut für Arbeitsmarkt- und Berufsforschung (IAB)

⁸ Deutsche Rentenversicherung Bund (DRV)

Abstract

Der Text beinhaltet die vom FDI Ausschuss diskutierten und beschlossenen Empfehlungen zur Vereinheitlichung der Bezeichnungen von Datenprodukten, die von Forschungsdatenzentren (FDZ) zur wissenschaftlichen Nutzung angeboten werden. Ziel ist es, die Verständlichkeit und Nachvollziehbarkeit der Datenprodukte seitens der Datennutzenden durch eine gemeinsame Bezeichnungslogik zu verbessern. Die vorgeschlagenen Bezeichnungen orientieren sich an den zwei zentralen Dimensionen Nutzungszweck (Public Use, Campus Use, Scientific Use) und Zugangsweg (Off-site, Remote, On-site).

Danksagung

Die Autorengruppe dankt den Mitgliedern des FDI Ausschusses für ihr wertvolles Feedback. Für verbleibende Fehler sind allein wir verantwortlich.

Keywords: Forschungsdatenzentren (FDZ), Datenprodukte, Bezeichnungspraxis, Nutzungszweck, Zugangsweg, Public Use File (PUF), Campus Use File (CUF), Scientific Use File (SUF), Forschungsdateninfrastruktur, Empfehlung

Motivation der Empfehlung

Zielgruppe: Die vorliegende Empfehlung des ständigen Ausschusses Forschungsdateninfrastruktur (FDI Ausschuss)¹ richtet sich insbesondere an bestehende und im Aufbau befindliche Forschungsdatenzentren (FDZ), aber auch an weitere Akteure der Bereitstellung von Forschungsdaten. Sie soll zum einen dazu anregen, historisch gewachsene und oftmals für externe Datennutzende nur bedingt verständliche Bezeichnungen für die angebotenen Datenprodukte² kritisch zu prüfen. Zum anderen soll die Empfehlung einen Standard für eine einheitlichere Bezeichnung von Datenprodukten innerhalb der Forschungsdateninfrastruktur in Deutschland anbieten.

Intention: Mit der vorliegenden Empfehlung wird auf eine Bezeichnungspraxis hingewirkt, die sich am Nutzungszweck der Forschungsdaten ausrichtet und die für Datennutzende einfacher nachvollziehbar ist. Diese gemeinsame Praxis soll bestehende Bezeichnungen ablösen, sofern diese missverständlich oder für Außenstehende nur schwer zu erschließen sind.

Hintergrund: Aktuell gängige Bezeichnungen von Datenprodukten weisen oftmals Bezüge zu drei Dimensionen mit jeweils unterschiedlichen Betonungen und Begrifflichkeiten auf:

- Nutzungsberechtigte bzw. Nutzungszweck der Daten (z. B. Scientific Use File „SUF“, Campus Use File „CUF“, Public Use File „PUF“)³
- Zugangsweg zu diesen Daten (z. B. „On-site“)
- Anonymisierungsgrad (z. B. „formal anonymisierte Daten“)

Der nachfolgende Vorschlag einer einheitlichen Bezeichnung von Datenprodukten fokussiert nur auf die beiden erstgenannten Dimensionen, also den Nutzungszweck und den Zugangsweg. Unterschiedliche Grade der Anonymisierung der Daten sind zwar in aller Regel mit dem jeweiligen Zugangsweg und der Nutzungsberechtigung der Daten verknüpft, der besseren Verständlichkeit halber sollte bei der Bezeichnung von Datenprodukten jedoch nicht explizit darauf Bezug genommen werden. Denn zum einen ist das Schutzniveau für ein bestimmtes Datenprodukt immer abhängig von der jeweiligen Kombination aus inhaltlicher Detailtiefe der potenziell re-identifizierenden Informationen in den Daten sowie der Ausgestaltung der technisch-organisatorischen Maßnahmen zur Nutzung der Daten und deren Kontrolle. Zum anderen gelten für verschiedene Datenprodukte unterschiedliche gesetzliche

¹ Der FDI Ausschuss ist die Vertretung der vom Rat für Sozial- und Wirtschaftsdaten (RatSWD) akkreditierten Forschungsdatenzentren (FDZ). Seine permanenten Aufgaben und Ziele verfolgt der FDI Ausschuss als unabhängiges, inhaltlich souveränes Gremium innerhalb des Konsortiums für Sozial-, Verhaltens-, Bildungs- und Wirtschaftswissenschaften (KonsortSWD).

² Als ein Datenprodukt wird in diesem Dokument die typischerweise unter einer DOI vom FDZ angebotene Sammlung von Daten verstanden. Dabei handelt es sich oftmals um mehrere Einzeldatensätze, die inhaltlich aufeinander bezogen sind (z. B. einer Studie entstammen) und die gemeinsam in einem oder mehreren ZIP-Files (z. B. für unterschiedliche Datenformate) zur Verfügung gestellt werden.

³ Neben den drei genannten Bezeichnungen existiert noch der sog. Secure Use File (SecUF): „Secure-use files´ means confidential data for scientific purposes to which no further methods of statistical disclosure control have been applied“ (COMMISSION REGULATION (EU) No 557/2013). Die Bezeichnung des SecUF findet jedoch weder in der deutschsprachigen noch in der internationalen FDI-Landschaft eine verbreitete Anwendung.

Regelungen, die unmittelbar das Schutzniveau betreffen. Ein Scientific Use File kann folglich unterschiedliche Anonymisierungsgrade je nach Zugangsweg aufweisen. Umgekehrt lässt sich nicht per se von einem Remote-Zugang auf ein gleiches Anonymisierungslevel der Datenprodukte schließen.

Empfehlung

Tabelle 1 veranschaulicht die empfohlenen Bezeichnungen für Datenprodukte auf der Grundlage von Nutzungszwecken und Zugangswegen. Die Logik basiert auf einer einfachen 3x3-Feldertafel, die einerseits zwischen Public Use (ohne Zweckbindung), Campus Use (für Lehrzwecke) und Scientific Use (für Forschungszwecke) und andererseits zwischen Off-site-Zugang (Verarbeitung der Daten am eigenen Rechner), Remote-Zugang (Ferndatenverarbeitung der Daten) und On-site-Zugang (Verarbeitung der Daten in Räumlichkeiten eines FDZ) unterscheidet. Dabei sind nicht alle neun Felder befüllt, da z. B. ein restriktiver Remote- oder On-site-Zugang für einen offenen Public Use File nicht sinnvoll erscheint. Auf explizite Hinweise zum jeweiligen Anonymisierungsgrad der Daten wird in den vorgeschlagenen Bezeichnungen von Datenprodukten aus den oben genannten Gründen verzichtet. Grundsätzlich gilt jedoch: Je restriktiver die Zugangsregelungen und je aufwändiger die technisch-organisatorischen Maßnahmen sind, desto geringer ist der Anonymisierungsgrad und desto mehr Detailtiefe ist typischerweise in den Daten gegeben.

Möglichkeiten der Kennzeichnung von Forschungsdaten mit unterschiedlichen Zugangswegen

Ein Datenprodukt mit gleichem Nutzungszweck, aber unterschiedlichen Zugangswegen, variiert demnach in aller Regel hinsichtlich der bereitgestellten Detailtiefe. Das heißt, ein Scientific Use File zur Download-Nutzung (Off-site) enthält weniger (sensible) Informationen als der entsprechende Scientific Use File zur On-site-Nutzung an einem Gastwissenschaftsarbeitsplatz (GWAP) des FDZ. Eine Option zur Kennzeichnung dieser unterschiedlichen Zugangswege bei gleichem Nutzungszweck besteht in der Vergabe separater DOI (Digital Object Identifier) als persistente Identifikatoren eines Datenprodukts.⁴

Beispiel:

- Sozio-oekonomisches Panel (SOEP), Version 39, Daten der Jahre 1984-2022 (SOEP-Core, v39, Onsite Edition). 2024. DOI: 10.5684/soep.core.v39_o
- Sozio-oekonomisches Panel (SOEP), Version 39, Daten der Jahre 1984-2022 (SOEP-Core v39, Remote Edition). 2024. DOI: 10.5684/soep.core.v39_r

⁴ GESIS & ZBW (2021). da|ra Policy. Richtlinien zur Vergabe von DOI-Namen (DOI®) über die Datenregistrierungsagentur da|ra Version 4.0. <https://www.da-ra.de/media/pages/downloads/general/dara-policy/3344cd569c-1679489505/darapolicy-dev4-31122020.pdf>

Eine alternative Möglichkeit, die sich bei der Verwendung einer identischen DOI für verschiedene Zugangswege anbietet, ist ein expliziter Verweis in der Zitationsempfehlung.

Beispiel:

- Briedis, K., Euler, T., Fabian, G., Landers, G., Mühleck, K., Redeke, S., Rehn, T., Schulz, J. & Trennt, F. (2021). DZHW-Absolventenpanel 2013. Datenerhebung: 2014-2019. Version: 3.0.0. Datenpaketzugangsweg: Download-SUF. Hannover: FDZ-DZHW.
Datenkuratierung: Hoffstätter, U., Vietgen, S. & Niebuhr, J.
<https://doi.org/10.21249/DZHW:gra2013:3.0.0>
- Briedis, K., Euler, T., Fabian, G., Landers, G., Mühleck, K., Redeke, S., Rehn, T., Schulz, J. & Trennt, F. (2021). DZHW-Absolventenpanel 2013. Datenerhebung: 2014-2019. Version: 3.0.0. Datenpaketzugangsweg: Remote-Desktop-SUF. Hannover: FDZ-DZHW.
Datenkuratierung: Hoffstätter, U., Vietgen, S. & Niebuhr, J.
<https://doi.org/10.21249/DZHW:gra2013:3.0.0>

Von Bedeutung ist, dass die Datenzitation jeweils exakt Aufschluss über die verwendeten Forschungsdaten gibt. Dazu gehört bei verschiedenen Zugriffsmöglichkeiten auf ein Datenprodukt auch die Angabe über die Art des Zugangs zu diesen Daten. Nur so ist bei unterschiedlichem Informationsgehalt in Abhängigkeit vom jeweiligen Zugangsweg die Replizierbarkeit von Analysen gewährleistet.

Tabelle 1: Empfehlung zur einheitlichen Bezeichnung für Datenprodukte im Angebot der Forschungsdatenzentren

	PUBLIC USE FILE (PUF)	CAMPUS USE FILE (CUF)	SCIENTIFIC USE FILE (SUF)
NUTZUNGSZWECK	Offen, nicht restringiert	Lehre	Wissenschaftliche Analyse
ÖFFENTLICHKEIT (Journalistinnen und Journalisten, Privatpersonen, Verbände, etc.)	x		
HOCHSCHULLEHRE (Studierende, Dozierende)	x ⁵	x	
(NACHWUCHS-) FORSCHENDE (Wissenschaftlerinnen und Wissenschaftler, Studierende für Qualifizierungsarbeiten)	x	x	x ⁶
ZUGANGSWEG			
OFF-SITE			
▪ Download ohne Registrierung/Vertrag	PUF Off-site		
▪ Download mit Registrierung/Vertrag	PUF Off-site	CUF Off-site	SUF Off-site
▪ Digitaler oder postalischer Versand	PUF Off-site	CUF Off-site	SUF Off-site
DATENFERNVERARBEITUNG			
▪ Remote Desktop (Datenverarbeitung über eigenen Bildschirm)			SUF Remote
▪ Remote Execution (Datenverarbeitung wird von FDZ ausgeführt)			SUF Remote
ON-SITE			
▪ [vernetzter] Gastwissenschaftsarbeitsplatz			SUF On-site
	inhaltliche Aussagen bei eher niedrigem Detailgrad möglich	teilweise keine inhaltlichen Aussagen möglich, primärer Fokus auf Datenstruktur	inhaltliche Aussagen mit hohem Detailgrad möglich

⁵ Auch Lehrzwecke können „Public Use“ sein. Entsprechend werden Daten für Lehrzwecke in manchen FDZ durch PUF bedient und keine separaten CUF bereitgestellt.

⁶ Die FDZ regeln den Zugang zu SUF-Daten für Qualifizierungsarbeiten unterschiedlich, sowohl in Bezug auf die Antragstellung als auch die Art des angestrebten Abschlusses.

Impressum

Kontakt:

FDI Ausschuss - Koordinationsstelle

Leibniz-Institut für Bildungsverläufe (LifBi)

Wilhelmsplatz 3

96047 Bamberg

Webseite: <https://www.konsortswd.de/ueber-uns/fdi-ausschuss/>

E-Mail: fdi-office@lifbi.de

Februar 2025

KonsortSWD Working Paper:

KonsortSWD baut als Teil der Nationalen Forschungsdateninfrastruktur Angebote zur Unterstützung von Forschung mit Daten in den Sozial-, Verhaltens-, Bildungs- und Wirtschaftswissenschaften aus. Unsere Mission ist es, die Forschungsdateninfrastruktur zur Beforschung der Gesellschaft zu stärken, zu erweitern und zu vertiefen. Sie soll nutzungsorientiert ausgestaltet sein und die Bedürfnisse der Forschungscommunities berücksichtigen. Wichtiger Grundstein ist dabei das seit über zwei Jahrzehnten durch den Rat für Sozial- und Wirtschaftsdaten (RatSWD) aufgebaute Netzwerk von Forschungsdatenzentren.

In dieser Reihe erscheinen Beiträge rund um das Forschungsdatenmanagement, die im Kontext von KonsortSWD entstehen. Beiträge, die extern und doppelblind begutachtet wurden sind entsprechend gekennzeichnet.

KonsortSWD wird im Rahmen der NFDI durch die Deutsche Forschungsgemeinschaft (DFG) gefördert – Projektnummer: 442494171.



Diese Veröffentlichung ist unter der Creative-Commons-Lizenz (CC BY 4.0) lizenziert:

<https://creativecommons.org/licenses/by/4.0/>

DOI: [10.5281/zenodo.14899704](https://doi.org/10.5281/zenodo.14899704)

Zitationsvorschlag:

Fuß, D. & Schlücker, F. (2025). *Empfehlung des FDI Ausschusses für eine einheitliche Bezeichnung von Datenprodukten der Forschungsdatenzentren*. KonsortSWD Working Paper Nr. 11/2025. Konsortium für die Sozial-, Verhaltens-, Bildungs- und Wirtschaftswissenschaften (KonsortSWD). <https://doi.org/10.5281/zenodo.14899704>