



Reihe: FAIR-Prinzipien und Datenschutz.

Wie kann FAIR Datenmanagement mit der Datenschutzgrundverordnung vereinbart werden?

Factsheet 2

Methoden zum FAIRen und datenschutzkonformen Umgang mit personenbezogenen Daten

Autor:innen:

Katharina Deschler¹, Andreas Hübner², Hans März³, Kati Mozygemba⁴, Jochem Rieger⁵, Ulrich Sax⁶

Affiliations:

- 1) GHGA (Deutsches Human Genom-Phänom Archive) (DFG Projektnummer [441914366](#)), DKFZ, Heidelberg, DE
- 2) NFDI4Earth (DFG Projektnummer [460036893](#)), Freie Universität Berlin, Universitätsbibliothek, Berlin, DE
- 3) EvH Bochum, Bochum, DE
- 4) KonsortSWD (DFG Projektnummer [442494171](#)), Universität Bremen, Qualiservice, Bremen, DE
- 5) NFDI4Energy (DFG Projektnummer [501865131](#)), Carl von Ossietzky Universität Oldenburg, Department für Psychologie, Oldenburg, DE
- 6) NFDI4Health (DFG Projektnummer [442326535](#)), Universitätsmedizin Göttingen, Institut für Medizinische Informatik, Göttingen, DE

Stand: 15.07.2025

Dieses Factsheet ist Teil einer Reihe zum Themenbereich "FAIR-Prinzipien und Datenschutz". Diese Factsheet-Reihe ist in einer Arbeitsgruppe zum Thema Datenschutz im Rahmen der [Nationalen Forschungsdateninfrastruktur](#) entstanden.

Die Factsheets verfolgen ausschließlich informative Zwecke und ersetzen keine individuelle Rechtsberatung.

Lizenz:  CC BY 4.0 – <https://creativecommons.org/licenses/by/4.0/>

INHALT

Methoden zum FAIRen und datenschutzkonformen Umgang mit personenbezogenen Daten	2
Forschungsdaten finden - Findability	2
Zugriff auf Forschungsdaten - Accessibility	3
Interoperabilität - Interoperability	4
Nachnutzung - Reusability	4
Querverweise	6
Quellen	6

Methoden zum FAIRen und datenschutzkonformen Umgang mit personenbezogenen Daten

Entlang der FAIR-Prinzipien (Wilkinson et al., 2016) geben wir Beispiele, wie Persönlichkeitsrechte durch Pseudonymisierung, Aggregation, Nachnutzung und Metadaten geschützt werden und gleichzeitig Forschung mit personenbezogenen Daten möglich ist.

FAIR Prinzipien	FINDABLE	ACCESSIBLE	INTEROPERABLE	REUSABLE
Datenschutz-Maßnahmen	• Pseudonymisierung • Daten-Aggregation • Nicht-personenbezogene Metadaten	• Zugangsbeschränkung • Datenzugriffskomitee • Verschlüsselung • Zugriffsberechtigungen	• Kompatible Metadaten-Schemata • Genormte Begriffe • Datenschutz-Folgenabschätzung	• Einwilligung • DUO und Provenance • Datennutzungsverträge und Lizenzen • EU Standard Contractual Clauses

Abbildung: Beispiele für Maßnahmen zum Datenschutz bei der Archivierung und Bereitstellung von personenbezogenen Forschungsdaten.

Forschungsdaten finden - Findability

Datensätze mit personenbezogenen Daten dürfen nicht ohne weiteres online gestellt werden. Wird von den Daten jeglicher Personenbezug entfernt, können diese als anonym geltenden Daten sicher geteilt werden. Dies ist allerdings in Forschungsbereichen wie z. B. der Genomforschung nicht möglich, da Gensequenzen immer Rückschluss auf eine Person erlauben. Auch pseudonymisierte Daten, bei denen im Prinzip noch eine Verbindung zu den personalisierten Daten besteht, sind besonders geschützt und können deshalb nicht ohne weiteres zugänglich gemacht werden (siehe [Art. 4 Abs. 5 DSGVO](#) sowie [Erwägungsgrund 26](#)

Wie können Daten leichter gefunden werden?

Daten sind leichter zu finden, wenn sie in Fachrepositorien oder Repositorien von Forschungseinrichtungen gespeichert werden. Diese Repositorien unterstützen die Forschenden, die FAIR-Prinzipien umzusetzen. Sie sind meist spezifisch für Daten aus einem Fachgebiet und bieten Suchfunktionen innerhalb des Repositoriums. Fachspezifische Repositorien können z.B. auf www.fairsharing.org, www.re3data.org, oder www.datasetsearch.research.google.com gesucht werden. Für das Bereitstellen von personenbezogenen Daten in einem Repositorium muss vorher geklärt sein, ob geeignete Datenschutzmaßnahmen, Einwilligung und Verträge vorhanden sind.

In Repositorien werden die Datensätze in der Regel mit einem Digital Object Identifier (DOI) versehen, damit sie nachhaltig auffindbar sind. Die DOI identifiziert den Datensatz eindeutig und ist unabhängig vom Speicherort. Bei der Beschreibung der Daten erleichtert die Verwendung von genormten Begriffen aus Ontologien wie z.B. DUO (<http://obofoundry.org/>) oder kontrollierten Wörterbüchern wie HED-Tags (<https://www.hedtags.org/>) zusätzlich das Auffinden von Datensätzen.

[DSGVO](#) zur Definition der Begriffe *anonym* vs. *pseudonym* sowie die Querverweise zu Anonymisierungs- und Pseudonymisierungstools). Daher werden Datensätze, die personenbezogene Daten wie z.B. Gensequenzen oder personenbezogene audiovisuelle Daten enthalten, meist mit einem unsensiblen Metadatensatz beschrieben, der frei geteilt werden kann, wodurch die verknüpften Datensätze für Wissenschaftler:innen auffindbar werden. Nicht-personenbezogene Metadaten können Studiendetails enthalten (z.B. Studie zu Diabetes Typ 2), aber auch aggregierte, personenbezogene Daten zum Studienobjekt selbst (z.B. Altersgruppe 50-60 Jahre). Die Datenverantwortlichen ([Art. 4 Abs. 7 DSGVO](#)) müssen jedoch dafür Sorge tragen, dass aus dem veröffentlichten Gesamtdatensatz keine Re-identifizierung einer Person möglich ist (für Informationen zur praktischen Umsetzung, siehe z.B. Teil D, Standarddatenschutzmodel (Konferenz der unabhängigen Datenschutzaufsichtsbehörden des Bundes und der Länder, 2024). So könnte z.B. eine Frau zwischen 20-30 Jahren mit einer seltenen Erkrankung, die in einem Postleitzahlenbereich mit wenigen Einwohnern lebt, möglicherweise schon durch diese Metadaten re-identifiziert werden (siehe [FACTSHEET Record Linkage](#)). Solche Risiken müssen minimiert und sollten im Einzelfall mit dem Datenschutzbeauftragten der jeweiligen Einrichtung geklärt werden.

Zugriff auf Forschungsdaten - Accessibility

FAIRe und datenschutzgerechte Daten-Zugangswege

Zum Teilen von personenbezogenen Daten sind sichere Daten-Zugangswege nötig (siehe [Art. 32 DSGVO](#)). Dies kann z.B. bedeuten, dass personenbezogene Daten nur vor Ort in geschützten Räumen eingesehen werden können. Wenn die rechtlichen Voraussetzungen dafür vorliegen, werden personenbezogene Daten auch elektronisch geteilt, wofür ein sicherer und verschlüsselter Übertragungsweg zwingend notwendig ist. Zur Sicherung der Daten-Zugangswege von Cloud- und Archiv-Diensten gehört auch, dass Daten-Manager:innen sowie Nutzenden individuelle Zugriffsberechtigungen zugeteilt werden, um datenschutzrechtliche und technische Sicherheit zu gewährleisten.

Darüber hinaus muss die Weitergabe von Daten gemäß der Einwilligungserklärung vertraglich geregelt werden (z.B. in der Datentransfer- oder Datennutzungsvereinbarung). In Zukunft wird man einen Teil der Forschungsdaten in virtuellen Umgebungen, wie sie z.B. mit der Unterstützung der [European Open Science Cloud](#) entwickelt werden, sicher speichern und cloud-basiert analysieren können. Damit verbunden ist das Ziel, sichere Datenspeicherung leichter umsetzen zu können und datenschutzgerecht Daten verwalten und nutzen zu können, ohne sie aus der Hand geben zu müssen.

Personenbezogene Datensätze werden durch Zugangsbeschränkungen geschützt. Um diese Datensätze anfragen zu können, müssen Kontaktdaten zu den jeweiligen Ansprechpartner:innen hinterlegt sein. Die Bedingungen für den Datenzugriff werden vertraglich geregelt und basieren unter anderem auf der Einwilligungserklärung der betroffenen Personen (siehe auch [FACTSHEET Einwilligung](#)).

Die Zugangsbeschränkungen sollten im Sinne einer offenen Wissenschaft möglichst wenig restriktiv gestaltet werden, trotzdem Persönlichkeitsrechte wahren und gegen das Risiko eines Missbrauchs (z.B. technisch und juristisch) absichern. Dabei ist es möglich, dass verschiedene Level der Aggregation und Pseudonymisierung von Datensätzen unterschiedlichen Zugangsbeschränkungen unterliegen.

Die Nutzungsbedingungen von Datensätzen mit personenbezogenen Daten können z. B. durch eine Datennutzungsvereinbarung (siehe z.B. Konsort SWD Musterverträge (Schallaböck et al., 2022)) geregelt und

durch ein Datenzugriffskomitee (Data Access Committee - DAC) geprüft werden.

Interoperabilität - Interoperability

Um eine effiziente Nachnutzung von Daten zu gewährleisten, sollten die Daten in einem möglichst standardisierten (siehe Box: Wie können Daten leichter gefunden werden?), maschinenlesbaren Format gespeichert und durch Metadaten reichhaltig beschrieben werden. Für die Interoperabilität von unterschiedlichen Datensätzen ist es zudem notwendig, dass die verwendeten Metadaten-Schemata kompatibel sind. Initiativen wie [GA4GH](#), [Elixir](#), [MII/NUM](#) (Ammon et al., 2024), [DDI](#) oder die [NFDI](#) erarbeiten hierfür gemeinsame Standards.

Im Sinne der FAIR-Prinzipien ist es erstrebenswert, möglichst viele sinnvolle Verknüpfungen zwischen (Meta-)Datenressourcen herzustellen, um das kontextuelle Wissen über die Daten zu erweitern. Diese Verknüpfung von unterschiedlichen Datensätzen kann unter Umständen jedoch das Risiko zur Re-Identifikation von Personen erhöhen (siehe [FACTSHEET Record Linkage](#)), wenn aggregierte Daten durch die Kombination von Datensätzen plötzlich Personenbezug herstellen lassen. Vor der Bereitstellung von Metadaten und dazugehörigen personenbezogenen Daten sollte daher mit dem/der Datenschutzbeauftragten der jeweiligen Einrichtung geklärt werden, ob eine Datenschutz-Folgenabschätzung ([Art. 35 DSGVO](#)) durchzuführen ist, die mögliche Risiken unter Berücksichtigung der vorhandenen Schutzmaßnahmen abwägt.

Nachnutzung - Reusability

Forschende haben häufig den Wunsch, Daten möglichst lange aufzubewahren und vielleicht auch für andere Forschungszwecke als ursprünglich beschrieben nutzen zu können. Demgegenüber steht die Forderung der DSGVO, dass personenbezogene Daten nur solange gespeichert werden dürfen, bis der Zweck der Datenerhebung erfüllt ist. Diesem Wunsch wird in der DSGVO ([Art. 89 DSGVO](#)) sowie im Gesundheitsdatennutzungsgesetz ([Art. 1 § 6 GDNG](#)) Rechnung getragen, die unter gewissen Umständen Garantien und Ausnahmen für Forschende vorsehen. Über den Zweck und die Dauer der Speicherung personenbezogener Daten muss bereits bei der Erhebung der Daten informiert werden. Um bei unabsehbarem Zweck die Nachnutzung von personenbezogenen Daten zu erleichtern, kann eine möglichst breite Zweckbeschreibung der Datenverarbeitung zum Zeitpunkt der Erhebung sinnvoll sein (sog. Broad Consent). Eine Voraussetzung für eine breite Zweckbeschreibung ist, dass die weitere Nutzung der Daten “unter Einhaltung der anerkannten ethischen Standards der wissenschaftlichen Forschung geschieht” ([Erwägungsgrund 33 DSGVO](#)). Im medizinischen Bereich z.B. erarbeitet die Medizin Informatik Initiative einheitliche Mustertexte (Kirsten et al., 2024), die eine möglichst breite Einwilligung ermöglichen (siehe auch [FACTSHEET Einwilligung](#)).

Vor dem Teilen der personenbezogenen Daten müssen Datennutzungsverträge abgeschlossen und ggf. Lizenzen vergeben werden (z.B. bezüglich eines Urheber- oder Leistungsschutzrechts bei Datenbanken), die mit der Einwilligung der betroffenen Person(en) in Einklang stehen.

Besonderes Augenmerk sollte auch auf dem Transfer der Daten in Länder liegen, für die die europäischen Datenschutz-Regularien ([DSGVO](#)) nicht gelten. Hier kann die Integration von EU Standard Contractual Clauses (Directorate-General for Justice and Consumers, 2021) in den Verträgen hilfreich sein. Dies sollte aber im Einzelfall immer mit dem/der Datenschutzbeauftragten der jeweiligen Einrichtung oder ggf. dem gewählten Repositorium geklärt werden.

Auch die Angaben zur Einwilligungserklärung und mögliche Nutzungseinschränkungen sollten mit dem Datensatz verknüpft sein. Die Verwendung von sog. Data Use Ontologies (DUO) kann hier für Forschende sehr wertvoll sein. Werden Datenderivate (z.B Analyseergebnisse) geteilt ist zudem die Dokumentation der Herkunft bzw. der Entstehung (Provenance) von zentraler Bedeutung. Auch Angaben, die helfen, die Datenqualität einzuschätzen, sollten in der Beschreibung der Daten erfasst werden.

Eine sichere Nachnutzung ist dabei mit einem hohen technisch-organisatorischen Aufwand verbunden: Einwilligungserklärung, rechtliche Verträge, Gewährleistung der technischen Sicherheit, etc. Dies alles dient dem Schutz der personenbezogenen Daten. Um die Bereitschaft zum Teilen von Daten allgemein zu erhöhen, müssen die genannten Aspekte möglichst transparent und nachvollziehbar sowie möglichst übersichtlich und zeitsparend in das Forschungsprojekt integriert werden können. Dabei können z.B. Datenmanagementpläne helfen.

Querverweise

EDPB Guidelines: Pseudonymisierung

https://www.edpb.europa.eu/system/files/2025-01/edpb_guidelines_202501_pseudonymisation_en.pdf

Pseudonymisierungs- und Anonymisierungstools

1. Amnesia: <https://amnesia.openaire.eu/>
2. Mainzelliste: <https://mainzelliste.de>
3. QualiAnon: <https://www.qualiservice.org/de/helpdesk/webinar/tools.html>
4. Anonymizer: <https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface>
5. ARX: <https://arx.deidentifier.org/>
6. Anonimatron: <https://realrolfje.github.io/anonimatron/>
7. Understand.ai Anonymizer: <https://github.com/understand-ai/anonymizer>

Quellen

- Ammon, D., Kurscheidt, M., Buckow, K., Kirsten, T., Löbe, M., Meineke, F., Prasser, F., Saß, J., Sax, U., Stäubert, S., Thun, S., Wettstein, R., Wiedekopf, J. P., Wodke, J. A. H., Boeker, M., & Ganslandt, T. (2024). Arbeitsgruppe Interoperabilität: Kerndatensatz und Informationssysteme für Integration und Austausch von Daten in der Medizininformatik-Initiative. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz*, 67(6), 656–667. <https://doi.org/10.1007/s00103-024-03888-4>
- Directorate-General for Justice and Consumers. (2021). *Standard contractual clauses for international transfers—European Commission*. https://commission.europa.eu/publications/standard-contractual-clauses-international-transfers_en
- Kirsten, T., Kleinert, P., Gebhardt, M., Drepper, J., Andreeff, A.-K., Prasser, F., & Kohlbacher, O. (2024). Grundlagen für die wissenschaftliche Nutzung umfangreicher Versorgungsdaten in Deutschland – Ergebnisse der AG Data Sharing der Medizininformatik-Initiative. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz*, 67(6), 648–655. <https://doi.org/10.1007/s00103-024-03880-y>
- Konferenz der unabhängigen Datenschutzaufsichtsbehörden des Bundes und der Länder. (2024). *Das Standard-Datenschutzmodell: Eine Methode zur Datenschutzberatung und -prüfung auf der Basis einheitlicher Gewährleistungsziele*. <https://www.datenschutzkonferenz-online.de/media/ah/SDM-Methode-V31.pdf>
- Schallaböck, J., Hoffstätter, U., Buck, D., & Linne, M. (2022). Mustervertrag Datennutzung KonsortSWD. *Zenodo*. <https://doi.org/10.5281/zenodo.5828114>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1), 160018. <https://doi.org/10.1038/sdata.2016.18>