

A Wenn Du effiziente Services für FAIRe Gesundheitsdaten suchst, dann sprich uns an

Hi, wir sind NFDI4Health

Wir sind die Nationale Forschungsdateninfrastruktur für personenbezogene Gesundheitsdaten und stellen eine technische Infrastruktur und effiziente Services für FAIRe Gesundheitsdaten bereit. Ziel dieser Services ist es, Daten aus der Gesundheitsforschung auffindbar und zugänglich zu machen, während Datenschutz und ethische Standards gewahrt bleiben. Nutzende werden zudem durch ein breites Support- und Trainingsangebot unterstützt.

Wie kommst Du mit uns ins Gespräch

Für wen ist unser Angebot relevant?

Unser Angebot richtet sich an Wissenschaftler:innen und Datenhaltende aus den Bereichen Epidemiologie, klinische Forschung sowie Public Health Studien. Zur NFDI4Health-Community gehören alle Akteure, die FAIRe Gesundheitsdaten schaffen und nutzen möchten. Im Mittelpunkt aller Aktivitäten von NFDI4Health stehen Interaktion, Vernetzung und Austausch mit der klinischen Forschung sowie mit epidemiologischen und Public-Health-Communities. Nur durch die aktive Mitgestaltung der Nutzer:innen können wir Konzepte entwickeln, die den Anforderungen der gesundheitsbezogenen Forschung gerecht werden.

Ergänzend organisieren wir Webinare, Workshops und FAIR Training Angebote.

Welche Leistungen bieten wir an?

- Health Study Hub: Strukturierte Veröffentlichung und Recherche studienbezogener interoperabler Metadaten sowie begleitender Dokumente und Variablenkataloge
- Local Data Hub: Anbindung lokaler datenhaltender Register der datenhaltenden Organisationen (Forschungsdatenmanagementtool)
- Tools zur Generierung synthetischer Daten (<https://github.com/nfdi4health/docker-vambn>) sowie zur Auswertung von synthetischer Datenqualität in Form eines Dashboards (<https://syndat.scai.fraunhofer.de/>) sowie eines Python Packages (<https://github.com/SCAI-BIO/syndat>)
- Anonymisierung: Tool zur Anonymisierung sensibler personenbezogener Daten, das zahlreiche Datenschutz- und Risikomodelle, Transformationsmethoden sowie Nutzenanalysen bietet
- DataquieR: R-Paket zur Durchführung automatisierter und standardisierter Datenqualitätsprüfungen.
- DataShield: Auf R-Paketen basierende föderierte Analyseinfrastruktur für die anonymisierte Analyse sensibler Forschungsdaten ohne Übertragung von Datensätzen auf Einzelpersonenebene
- FDPG: Zentrale Plattform zur Beantragung von Studiendaten (in MII entwickelt und weiterentwickelt in Netzwerk Universitätsmedizin und NFDI4Health)
- Helpdesk für Fragen aller Art, auch rechtliche Expertise
- Trainings zu Forschungsdatenmanagement und/oder NFDI4Health Services

Welche Datentypen können verarbeitet werden?

- Syndat: bietet Interface in Form eines Web-Dashboards sowie Python API (pypi:syndat) zur Bewertung & Visualisierung synthetischer Daten
- VAMBN: Variational Autoencoder Modular Bayesian Networks kann realistische synthetische Daten selbst von longitudinalen, heterogenen Studiendaten zu erzeugen
- DataquieR: numerische tabellarische Daten, darunter bevölkerungsbasierte Kohortenstudien, medizinische Register und Daten aus elektronischen Patientenakten
- LDH: Dokumente wie Projekte, Studien, Publikationen, (bio-)medizinische Daten, Modelle und Software-Tools sowie Studienprotokolle, Datensätze, Verweise auf Publikationen
- DataShield: Daten werden in der Regel in ein Opal-kompatibles Format importiert (z. B. per CSV-Upload)

Technische Rahmenbedingungen

- Das Syndat Dashboard lässt sich über die öffentliche Demo (syndat.scai.fraunhofer.de) sowie über einen Docker Container lokal deployen. Das syndat Python package ist kompatibel mit Python 3.10+ und läuft auf jeder mit Python kompatiblen Hardware
- VAMBN lässt sich lokal über ready-built Docker Container deployen und trainieren
- DataquieR: erfordert R (\geq Version 3.6.0), für Berechnung & Darstellung andere CRAN-Pakete (z. B. ggplot2)
- LDH: 16 GB RAM, 100 GB Speicherplatz; auf eigenständigem PC/Server oder als Teil einer VM-Infrastruktur. Für Installation wird Linux empfohlen.
- DataShield: erfordert verteilte Infrastruktur, in der jeder teilnehmende Standort Opal ausführt, mit relationalem Datenbanksystem wie PostgreSQL oder MySQL