

# NFDI4Health

## Forschungsdateninfrastrukturen für personenbezogene Gesundheitsdaten

Prof. Dr. Juliane Fluck, ZB MED, Katzenburgweg 1a, 53115 Bonn, fluck@zbmed.de

Prof. Dr. Iris Pigeot, BIPS, Achterstraße 30, 28359 Bremen, pigeot@leibniz-bips.de

### **Thema/Schlagwort/Fach:**

Medizin, Public Health, Epidemiologie, personenbezogene Daten

### **Beschreibung des Vorhabens:**

Die Erfassung und Analyse personenbezogener Daten zu Gesundheits- bzw. Krankheitsstatus und wichtiger Einflussfaktoren sind eine wesentliche Komponente zur Entwicklung neuer Therapien, übergreifender Versorgungsansätze und präventiver Maßnahmen eines modernen Gesundheitswesens. Die fortschreitende Digitalisierung führt zu einem erheblichen Wachstum des hierfür nutzbaren Datenbestands, aber auch zu einem erhöhten Bedarf an beschreibenden Daten. Diese Daten bieten eine hervorragende Ressource, um die obigen Ziele zu erreichen; jedoch sind personenbezogene Gesundheits- und Krankendaten besonders schützenswert.

Daraus erwachsen besondere Herausforderungen an zukünftige Forschungsdateninfrastrukturen sowohl im Bereich der klinisch-medizinischen Forschung als auch im Bereich der bevölkerungsbezogenen Gesundheitsforschung. Mit der Medizininformatikinitiative des BMBF wird derzeit eine flächendeckende, generalisierte Forschungsdateninfrastruktur an nahezu allen Universitätsmedizinstandorten aufgebaut. Zudem werden auch mit der NAKO oder den DZGs spezifische, verteilte Dateninfrastrukturen implementiert. Es gilt, weitere Datenbestände infolge der FAIR-Prinzipien für die Nutzung durch Forschende zu öffnen und mit den bereits bestehenden Infrastrukturen gemäß der dort etablierten Standards zu verknüpfen. Darüber hinaus liegt ein besonderes Potential in der personenbezogenen Verknüpfung von Primär-, Sekundär- und Registerdaten. Dienste, die eine solche dynamische und sichere Datenverknüpfung ermöglichen, existieren bisher nicht und müssen entwickelt werden.

Eine besondere Herausforderung stellt dabei die Tatsache dar, dass es sich hierbei um „lebende“ Datenkörper handelt, die einer ständigen Pflege, Aktualisierung und Fortschreibung bedürfen. Es muss sichergestellt werden, dass der ursprüngliche Datensatz jeweils um neu generierte Variablen erweitert werden kann, die z.B. bei der Analyse biologischer Proben entstehen.

Wichtige Ziele des Konsortiums sind außerdem die Harmonisierung und die Standardisierung der Metadaten. Für zentrale Metadaten zur Beschreibung von klinischen Versorgungsdaten werden die Standards in der Medizininformatik-Initiative entwickelt. Klinische Studien und epidemiologische Populationsstudien beinhalten jedoch weitere, heterogene und spezifische Datenbestände. Ausgehend von den bereits etablierten Infrastrukturen muss ein Prozess der Standardisierung etabliert werden, um auch diese dem jeweiligen Kontext angemessen unter Einhaltung hoher Qualitätsstandards analysieren zu können. Neben strukturierter Phänotyp-Dokumentation stehen in immer größerem Umfang Biomaterialien sowie Bild-, Ton- und andere Geräteaufzeichnungen zur Verfügung. Um z.B. maschinelle Lernverfahren zur Analyse solcher Datenkörper zu etablieren, bedarf es großer Mengen von Trainingsdaten, die in geschützter Arbeitsumgebung bereitgestellt werden müssen.

Um diesen Herausforderungen zu begegnen, setzt sich das NFDI4Health Konsortium aus Vertretern der Fachcommunities und Entwicklern von Forschungsdateninfrastrukturen zusammen und plant eine enge Zusammenarbeit z.B. mit der Medizininformatik-Initiative, klinischen Studienzentren und weiteren großen Datenhaltern im Gesundheitswesen. Zudem sind eine enge Vernetzung und ein Austausch von Standards, Daten und Erfahrungen mit anderen NFDI Initiativen geplant, insbesondere mit NFDI4Life, NFDI4BioDiversity, sowie NFDI4Neuroscience.