

# Föderierte Datennutzung in der Medizininformatik-Initiative und dem Netzwerk Universitätsmedizin: Grundlagen, Vorteile und Herausforderungen

## Hintergrund

Die Medizininformatik-Initiative (MII) und das Netzwerk Universitätsmedizin (NUM), gefördert vom Bundesministerium für Forschung, Technologie und Raumfahrt, verfolgen das Ziel, Forschungs- und Versorgungsdaten aus der gesamten Universitätsmedizin nutzbar zu machen. Hierzu wurden an allen Universitätskliniken sogenannte Datenintegrationszentren (DIZ) aufgebaut. Diese Zentren ermöglichen eine strukturierte und datenschutzgerechte Nutzung von Patientendaten für die medizinische Forschung – mit Einwilligung oder auf Basis anderer Rechtsgrundlagen, wie den Landeskrankenhausgesetzen oder dem Gesundheitsdatennutzungsgesetz. Ein zentrales Prinzip ist dabei die Etablierung von technischen Plattformen, mit denen Patientendaten sicherer für wissenschaftliche Analysen genutzt werden können.

## Föderierte Datennutzung

Ein zunehmend genutzter Ansatz ist die sogenannte „föderierte Datennutzung“. Dabei verbleiben Patientendaten vollständig an den Standorten (in den jeweiligen DIZ). Die Forschungsfrage wird standortübergreifend beantwortet, indem die Computeralgorithmen zu den Daten an den Standorten versandt werden. Alle teilnehmenden Kliniken führen die Berechnung lokal auf ihren geschützten IT-Systemen durch. Nur aggregierte oder modellierte Ergebnisse werden anschließend verschlüsselt zusammengeführt. Dieser Vorgang kann wiederholt ausgeführt werden, bis ein finales Gesamtergebnis berechnet ist. In der MII bzw. dem NUM werden aktuell zwei Plattformen eingesetzt:

- **DataSHIELD:** Erlaubt föderierte statistische Auswertungen (z. B. Mittelwerte, Regressionen), ohne Einzeldaten zu übertragen.
- **FLAME:** Wurde für komplexere Analysen (z. B. KI-Modellierung mit Bild- oder Genomdaten) entwickelt. Es ermöglicht verteiltes Training von KI-Modellen sowie statistische Auswertungen, auch bei sehr großen Datenmengen.

## Herausforderungen

Föderierte Verfahren bieten erhebliche Vorteile für den Datenschutz und die Verarbeitung großer Datenvolumina, da sie auf eine physische Weitergabe von Einzelangaben verzichten. Es gibt jedoch zwei wesentliche Herausforderungen.

Aus der Datenschutzsicht kann nicht automatisch von Anonymität im datenschutzrechtlichen Sinne ausgegangen werden, da die Ergebnisse (z. B. Tabellen, Modellparameter) unter Umständen Rückschlüsse auf Einzelpersonen zulassen – insbesondere bei kleinen Fallzahlen oder hochdimensionalen Daten. Deshalb kommen in den in der MII bzw. dem NUM genutzten Plattformen ergänzende Schutzmaßnahmen zum Einsatz. Diese umfasse, u. a.:

- Statistical Disclosure Control (SDC): Methoden wie Schwellenwertprüfung (zum Beispiel Mindestfallzahlen) und Ausreißer-Unterdrückung.
- Differential Privacy: Hinzufügen statistisch kontrollierter „Unschärfe“, um Rückschlüsse auf Individuen zu verhindern.
- Verschlüsselung und Zugangskontrolle: Ergebnisse werden nur verschlüsselt übertragen und nur autorisierten Personen bereitgestellt.
- Organisatorische Schutzmechanismen: Zugriff nur durch geschulte Forschende unter kontrollierten Governance-Strukturen (z. B. Nutzungskomitees, Datenschutzprüfung).

Diese Maßnahmen werden entlang des etablierten Five Safes Framework kombiniert:

- Safe People – Qualifikation und Verpflichtung der Nutzenden
- Safe Projects – Forschungsvorhaben mit ethischer und rechtlicher Prüfung
- Safe Settings – technische Sicherheitsvorkehrungen an den Standorten sowie bei der Berechnung selbst
- Safe Data – Schutzmaßnahmen auf Daten- und Ergebnisebene
- Safe Outputs – Bereinigung der Auswertungsergebnisse vor Herausgabe

Die organisatorischen und rechtlichen Verantwortlichkeiten der beteiligten Einrichtungen bei förderter Datennutzung sind dabei verbindlich geregelt – insbesondere im Teilnehmerahmenvertrag der MII (inkl. Anlagen) sowie in den Nutzungsverträgen für konkrete Forschungsvorhaben. Diese Vereinbarungen legen fest, wer für welche Prozesse zuständig ist, wie Genehmigungen eingeholt werden, und welche datenschutzrechtlichen Anforderungen einzuhalten sind.

Es ist außerdem zu beachten, dass es wissenschaftliche Fragestellungen gibt, die sich nur dann beantworten lassen, wenn sämtliche Daten physisch zusammengeführt werden – etwa, weil bestimmte statistische Verfahren gefördert technisch nicht umsetzbar sind oder nicht ausreichend validiert wurden. Darüber hinaus erfordert förderierte Forschung eine besonders enge Abstimmung zwischen den beteiligten Kliniken. Da die Forschenden keinen direkten Zugriff auf die Patientendaten haben, müssen alle Verarbeitungsschritte – von der Datenvorbereitung bis zur Interpretation der Ergebnisse – miteinander abgestimmt sein. Das setzt voraus, dass die zugrunde liegenden Daten an allen Standorten vergleichbar strukturiert und kodiert sind, z. B. hinsichtlich genutzter Terminologien Zeitformate oder Einheiten. In der Praxis ist diese Harmonisierung oft nur sehr schwer zu erreichen.

## Abwägung statt Patentlösung

Förderierte Auswertungen bieten ein mehrschichtiges Schutzmodell, das datensparsame Forschung und die Verarbeitung sehr großer Datenmengen ermöglicht. Dennoch sind sie keine Patentlösung. Die Entscheidung, ob ein Forschungsprojekt gefördert umgesetzt werden kann und ob dies unter Abwägung der relevanten Aspekte sinnvoll ist, muss daher immer individuell geprüft werden.