*Lösungsbaustein **fTTP (federated Trusted Third Party)** als ein Enabler für vernetzte medizinische Forschung mit dezentraler Datenhaltung*

Bahls T, Hampf C, Bialke M, Hoffmann W; Universitätsmedizin Greifswald; 10.10.2022

Infrastrukturen für vernetzte medizinische Forschung, wie sie aktuell in der Medizininformatik-Initiative (MII) und dem Netzwerk Universitätsmedizin (NUM) prospektiv deutschlandweit etabliert werden, sollen die Datenbestände der einzelnen Klinika für gemeinsame Verbundforschung verfügbar machen und somit die Bearbeitung komplexer und längsschnittlicher medizinischer Fragestellungen anhand von Daten aus der realen Versorgung ermöglichen. Ein Kernstück dieser Infrastrukturen sind **Prozesse und Schnittstellen zwischen den dann vernetzten Einrichtungen**, die es so zusätzlich zu den lokalen und oftmals bereits etablierten Prozessen erlauben, Daten in einer "föderierten" Weise lokal zu selektieren und anschließend übergreifend über mehrere Einrichtungen zusammenzuführen. Die **federated Trusted Third Party (fTTP)** ist einer der erforderlichen Lösungsbausteine zum erfolgreichen Aufbau solcher Infrastrukturen. Sie ermöglicht es, diese Zusammenführung konform zu den für personenbezogene Patientendaten geltenden Regularien von Ethik und Datenschutz umzusetzen. Wesentliches Element dieser föderierten Arbeitsweise ist der **Verbleib der Hoheit über die Daten in den jeweiligen lokalen Einrichtungen**. So verbleiben beispielsweise die identifizierenden Daten und Einwilligungen der betreffenden Personen stets an den jeweiligen lokalen Einrichtungen. Eine dauerhafte einrichtungsübergreifende Speicherung identifizierender Informationen an zentraler Stelle erfolgt nicht. Wissenschaftliche Methoden, beispielsweise auf Basis von Bloomfiltern, erlauben es dennoch, **ohne Kenntnis der ursprünglichen identifizierenden Daten ein einrichtungs- und datensatzübergreifendes Record Linkage vorzunehmen**, so dass über Einrichtungs- und Datensatz-Grenzen hinweg erkannt werden kann, wenn Daten zu ein und derselben Person gehören. Während Datensätze im Umfeld großer Volkskrankheiten eher als statistisch stabil angesehen werden können, entscheidet ein solches Record Linkage beispielsweise im Umfeld der Seltenen Erkrankungen maßgeblich über die Qualität und Durchführbarkeit einer Analyse.

Im Einzelfall kann ein **Clearing-Prozess** erfolgen, nach dessen Abschluss die hierfür zentral verarbeiteten Daten unmittelbar wieder gelöscht werden. Ein solches Clearing ist dann erforderlich, wenn etwa aufgrund abweichender Schreibweisen oder "Zahlendrehern" beispielsweise im Geburtsdatum in den beteiligten Einrichtungen die auf Basis der Bloomfilter erzielbare Wahrscheinlichkeit nicht ausreicht.

Die fTTP bietet in Ergänzung der lokalen Prozesse in den beteiligten Einrichtungen, die jederzeit die Hoheit über "ihre" Daten behalten, Möglichkeiten für:

- wahrscheinlichkeitsbasiertes Record Linkage und einen zusätzlichen (optionalen) Clearing-Prozess über Einrichtungs- und Datensatz-Grenzen hinweg
- einheitliche Pseudonymisierung und Wiedererkennung von Personen über Einrichtungs- und Datensatz-Grenzen hinweg, auch in Hinblick auf zusammengeführte Analysedatenbestände, die gemäß Use & Access Regularien in pseudonymisierter Form an Forschende übergeben werden
- föderiertes Consent-Management zur automatisierten Prüfung des tagesaktuellen Einwilligungsstatus und zum einrichtungsübergreifenden Abgleich ggf. verschiedener Einwilligungstexte, -versionen oder -module, der jeweiligen Gültigkeitsdauer sowie der Berücksichtigung komplexer Anforderungen bei der Umsetzung von teilweisen oder vollständigen Widerrufen und Datenauskünften der teilnehmenden Patienten

Die fTTP stellt somit den Lösungsbaustein dar, der den Betrieb von Infrastrukturen ermöglicht, um Machbarkeitsanfragen in dezentral organisierten Datenbeständen zu bearbeiten, Teildatenkörper zu selektieren und diese einrichtungsübergreifend datenschutzkonform zusammenzuführen sowie mit weiteren Datenquellen zu verknüpfen. Die fTTP dient dabei mit ihren technischen Möglichkeiten der Qualitätsbildung und der konsequenten Wahrung der Betroffenenrechte in diesen Infrastrukturen.

Einheitliche Pseudonymisierung und föderiertes Record Linkage (am Beispiel NUM-RDP)

Die deutschlandweite Vernetzung von Forschungs- und Dateninfrastrukturen einer Vielzahl universitärer Einrichtungen wird aktuell in mehreren Förderlinien verfolgt (Medizininformatik-Initiative, Netzwerk Universitätsmedizin, Nationale Forschungsdateninfrastruktur NFDI). Ziel ist stets die Vernetzung, Standardisierung und Automatisierung der Verbindung der verschiedenen Datenbestände der Standorte zur Durchführung standortübergreifender Forschungsvorhaben. Die Zusammenführung dezentral erhobener und gespeicherter medizinischer Daten (einschließlich Bilddaten und Biomaterialien) zu einem spezifisch zu beforschenden Data Set ist erforderlich, um inhaltlich umfassende, statistisch belastbare und qualitativ hochwertige Datensätze zur Beantwortung spezifischer und komplexer gesundheitsbezogener Fragestellungen zu erzielen. Eine rein dezentral konzipierte und auf die jeweiligen lokalen Daten bezogene Betrachtung würde die gegenwärtige Situation fortsetzen. Die Forschung würde weiterhin durch die methodisch heterogene und im Ergebnis nicht zusammenführbare Vielzahl isolierter Dateninseln behindert und beschränkt - ein solches Vorgehen liefe damit den zuvor genannten Zielen im Grundsatz zuwider. Daher ist eine übergreifende, und in der technischen Betrachtung "föderierte" Infrastruktur zur Vernetzung der Datenbestände unabdingbar. Ungeachtet solch überzeugender und relevanter Zielstellungen sind die gesetzlichen Regularien etwa im Bereich des Datenschutzes auch in solchen einrichtungsübergreifenden Settings strikt einzuhalten. Eine zeitlich oder örtlich isolierte und eventuell nur zum Erhebungszeitpunkt der Daten einmalig durchgeführte Beachtung regulativer Erfordernisse greift somit zu kurz. Stattdessen ist die Informierte Einwilligung Basis aller Datenprozesse von der Erhebung über die einrichtungsübergreifende Zusammenführung, dem Datenmanagement, dem Monitoring und der Qualitätssicherung bis hin zur Auswertung und Publikation.

Gleichzeitig sind bei allen diesen Prozessen die Anforderungen des Datenschutzes bestmöglich und konsequent umzusetzen. Pseudonymisierung ist ein erforderliches Mittel, um die Identität einer Person auf das erforderliche Maß und den zulässigen Kreis zu beschränken. Das Ziel der Zusammenführbarkeit kann nur dann erreicht werden, wenn einerseits eine Separation mit spezifischen Pseudonymen pro Einrichtung und Data Set erfolgt und andererseits dennoch eine Verbindbarkeit durch eine zu diesen Einrichtungen oder Data Sets "orthogonale" dritte Stelle technisch und methodisch möglich ist. Entscheidende Voraussetzungen sind die methodische Einheitlichkeit und die numerische Eineindeutigkeit bei der Pseudonymisierung über alle Einrichtungen hinweg. Ein Auflösen von Pseudonym-Hierarchien über Einrichtungs-, Datensatz- und Projekt-Grenzen kann erforderlich werden, wenn beispielsweise Re-Kontaktierungsbedarf besteht, Widerrufe umgesetzt und Datenauskünfte zusammengestellt werden müssen. All dies ist nur möglich, wenn zuvor ein netzwerkweit einheitliches Vorgehen bei der Pseudonymisierung erfolgt ist. Die MII und die NUM-Routinedatenplattform (NUM-RDP) setzen dezentrale Strukturen voraus, wodurch Verantwortlichkeiten am Standort verbleiben und zentrale starre Strukturen entfallen. So werden identifizierende Daten (IDAT) nur in den lokalen Treuhandstellen verwaltet und verlassen nicht den Standort, an dem sie gewonnen wurden. Um dennoch die standortübergreifende Zusammenführung medizinischer Daten (MDAT) zu ermöglichen, wird ein Privacy-Preserving Record Linkage (PPRL) benötigt, welches informatorische Verfahren ohne regelhafte Kenntnis von IDATs umsetzt.

Die fTTP stellt auf Basis der erarbeiteten MII-Konzepte den Lösungsbaustein dar, der die genannten Punkte adäquat adressiert. Hierbei werden einheitliche Pseudonymisierungs-Mechanismen über offene und gebräuchliche Schnittstellen für alle Standorte bereitgestellt und Pseudonym-Hierarchien systematisch und konsistent verwaltet. Standortspezifische Pseudonyme stellen sicher, dass die Standorte untereinander nicht abgleichen können, wo weitere Daten bestimmter Patienten vorliegen. Gleichzeitig werden standortübergreifende Pseudonyme vergeben, so dass auch die zentrale Plattform keine Rückschlüsse auf die jeweiligen lokalen Identifier und Datenquellen ziehen kann. Die fTTP kennt die Beziehungen dieser Pseudonyme (jedoch nicht die zugehörigen MDAT und IDAT) und kann diese Beziehungen bei Bedarf auflösen, beispielsweise für Rekontaktierungen oder langzeitliche Nachverfolgungen. Darüber hinaus ist es möglich, für die jeweiligen Verwendungszwecke jeweils weitere entsprechend einheitliche, unabhängige und "frische" Pseudonyme bereitzustellen, beispielsweise Transfer-Pseudonyme für Datenherausgaben im Rahmen des Use & Access Verfahrens.

Um standortübergreifende Datenzusammenführungen zu ermöglichen, implementiert die fTTP ein Privacy-Preserving Record Linkage basierend auf Bloomfiltern. Damit können Dubletten standortübergreifend erkannt werden. Gleichzeitig ist es der fTTP nicht möglich, Rückschlüsse auf die IDAT einer Person zu ziehen. Diese ist ausschließlich in der Treuhandstelle der Institution bekannt, in der ein Patient spezifisch in die Nutzung seiner oder ihrer Behandlungsdaten für Forschungszwecke eingewilligt hat. Dies resultiert in einem hohen Datenschutzniveau und erfüllt die Anforderungen der MII und von NUM-RDP. Die Erzeugung der Bloomfilter findet lokal an den jeweiligen Standorten auf Basis der dort vorliegenden identifizierenden Daten statt. Dabei wird im Zuge des erwähnten netzwerkweiten Datenmanagements einrichtungsübergreifend dasselbe Verfahren angewendet, um der fTTP vergleichbare Bloomfilter zu übermitteln und das PPRL zu ermöglichen.

Die **fTTP Wahrscheinlichkeit** ermöglicht die **pseudonymisierte** und verlässliche, konsistente und korrekte **Zusammenführung** medizinischer Daten eines und desselben **Patienten** in vernetzten Forschungsvorhaben. Damit steht eine Lösung bereit, welche den Zielen der MII und von NUM-RDP entspricht und zugleich allen datenschutzrechtlichen Regularien genügt.

Zusätzlicher Clearing-Prozess

Insbesondere in Forschungsvorhaben mit geringen Teilnehmerzahlen, etwa bei Seltenen Erkrankungen (dort zudem mit ausgeprägtem "Patiententourismus" zwischen Einrichtungen), ist ein hohes Maß an Qualität bei der Datenzusammenführung erforderlich, um statistisch dennoch belastbare Datenkörper zur Analyse zu erhalten. Dies kann bei ausschließlicher Verwendung von codierten Daten bzw. Bloomfiltern nicht gewährleistet werden, da diese nur eine wahrscheinlichkeitsbasierte Betrachtung erlauben. Im Fall nicht mehr ausreichender Wahrscheinlichkeit beim PPRL könnten mögliche Dubletten ohne Zusatzinformationen nicht aufgelöst werden – es entstehen Synonymfehler, die die Validität einer Analyse erheblich beeinträchtigen können. In diesen Einzelfällen kann das Record Linkage unter temporärem (!) und technischem Rückgriff auf nur die betreffenden IDAT umgesetzt werden, damit auch diese Fälle korrekt aufgelöst werden können.

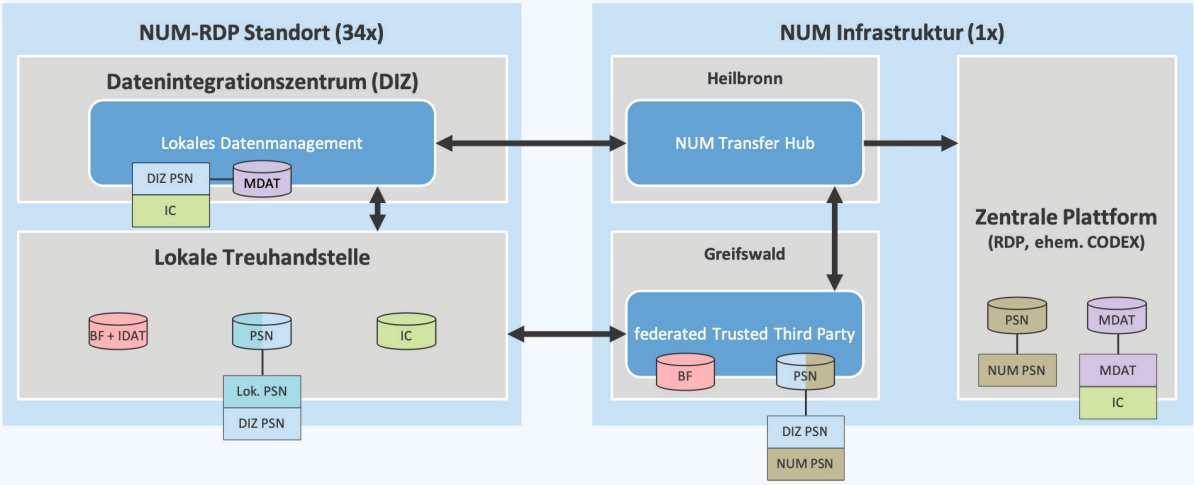

Abbildung 1. Standortübergreifende Datenzusammenführung mittels Privacy-Preserving Record Linkage sowie konsistente Pseudonymisierung durch die fTTP.

Die **fTTP Clearing** sieht für **spezifische** betroffene Datensätze eine Nachlieferung identifizierender Daten vor. Diese werden nur für den Zweck des Record Linkage **temporär** verarbeitet und nach **Auflösung der Dublette** in der fTTP gelöscht. Dieses Vorgehen entspricht den in der MII abgestimmten Prinzipien.

Consent-Orchestrierung: standortübergreifende Wahrung von Betroffenenrechten (am Beispiel NUM-RDP)

Die gültige Einwilligung des Patienten auf Basis des MII Broad Consent stellt die derzeitige, abgestimmte Rechtsgrundlage der dezentralen Erhebung, Zusammenführung und zentralen Herausgabe von Gesundheitsdaten für die Forschung in NUM dar. Einwilligungen werden lokal an den Standorten dokumentiert und möglichst digital in den lokalen THSen der Standorte abgebildet. Es ist beabsichtigt, ausgewählte Einwilligungsinformationen (pseudonymer Patientenbezug, Umfang der Einwilligung, Gültigkeitsdauer der codierten Einwilligungspolicies) zusammen mit den medizinischen Daten zentral an den NUM Transfer Hub (NTH) wie auch zur zentralen Routinedatenplattform zu übertragen.

Um die Betroffenenrechte des Patienten DSGVO-konform umsetzen zu können, dürfen beispielsweise Datenherausgaben von Forschungsdaten durch die zentrale Plattform stets nur erfolgen, wenn zum Zeitpunkt der Herausgabe die Gültigkeit der Einwilligung und somit die rechtliche Zulässigkeit der Datenweitergabe durch RDP gegeben ist und sichergestellt wurde und eine ausreichende Übereinstimmung bei der Zweckbestimmung zwischen Einwilligungsinhalt und Nutzungsvorhaben besteht. Zudem umfasst die Einwilligung verschiedene Module (z.B. Wiederkontaktierung, Nutzung von Biomaterialien, genetische Analysen), so dass niemals nur ein einziges zum Erhebungszeitpunkt gespeichertes summarisches "Flag" an den Daten oder Biomaterialien ausreichend ist. Vielmehr besteht zum Nutzungszeitpunkt eine komplexe Entscheidungssituation bezogen auf Aktualität der Einwilligung, spezifische Einwilligungsmodule und eingewilligter Zweckbestimmung – und ein Patient kann diese seit der letzten Übertragung ein- oder mehrfach und an einem oder mehreren der Standorte geändert haben, an denen Daten von ihm vorliegen.

Die zunehmend umfassendere Vernetzung der NUM-Standorte und zentralen Infrastruktur-Komponenten erhöht die Komplexität und erschwert die korrekte Umsetzung der Betroffenenrechte:

1. Patienten können stets zu verschiedenen Zeitpunkten an mehreren Standorten und auf unterschiedliche Weise eingewilligt haben (verteilte MDAT, unterschiedliche Nutzungsmöglichkeiten, voneinander abweichende Einwilligungsaussagen).
2. Bei der Ermittlung des "aktuellen Einwilligungsstatus" eines Patienten sind alle Versionen der verwendeten Einwilligungen eines Patienten (ggf. mit inhaltlichen Unterschieden), sowie ggf. Einwilligungen weiterer Studien (unabhängig von NUM und MII) zu berücksichtigen.
3. Durch die dezentrale bzw. verteilte und ggf. mehrfache Verarbeitung von Einwilligungsinformationen (z.B. verschiedene universitäre Einrichtungen/Häuser, NUM Transfer Hub, Routinedatenplattform) steigt das Risiko von Inkonsistenzen und der Entscheidung auf Basis ggf. veralteter Daten. Dies kann eine falsch-positive Entscheidung zur rechtlichen Zulässigkeit von Datenherausgaben und somit eine Verletzung von Betroffenenrechten des Patienten zur Folge haben.
4. Ein einzelner Patienten-Widerruf kann sich in der Umsetzung auf alle datenverarbeitenden Stellen von NUM (Standorte, NUM Transfer Hub, Routinedatenplattform) auswirken. Die Umsetzung des Betroffenenrechts (Art. 7 (3) DSGVO) erfordert de facto komplexe und automatisierte Kommunikationsprozesse. Insbesondere die netzwerkweite Rekontaktierungssperre ausgewählter Patienten ("Verweigerer für alle zukünftigen Studien", vergleichbar zur "Robinson-Liste" muss einheitlich und netzwerkweit geeignet dokumentiert und von allen Standorten berücksichtigt werden.

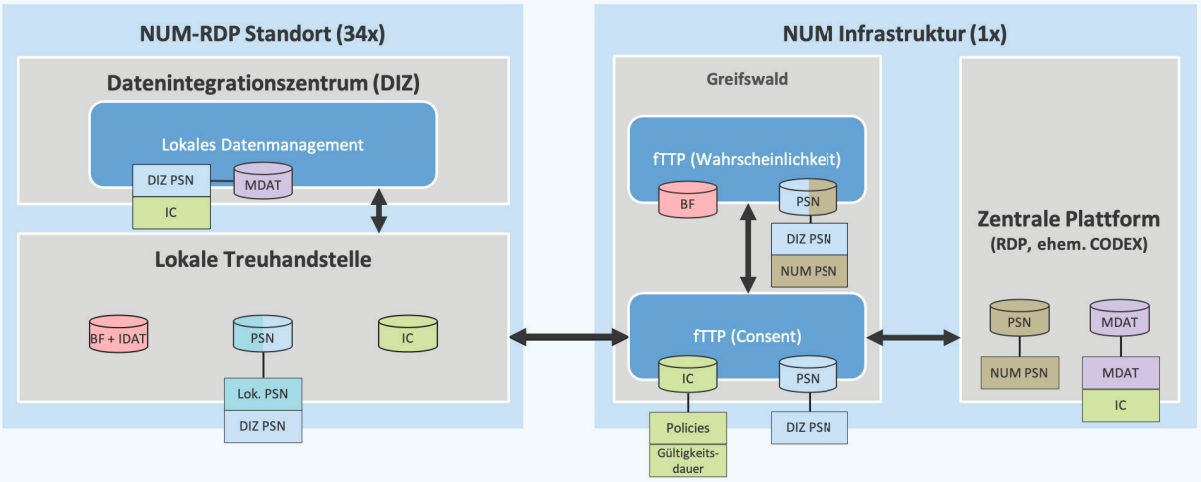

Abbildung 2. Standortübergreifende Vermittlung beim Management ausgewählter Consent-Situationen mit fTTP (Consent).

Ein föderiertes Einwilligungsmanagement adressiert diese Herausforderungen und orchestriert standortübergreifende Einwilligungsprozesse in hoher Qualität. Die fTTP Consent stellt einen einheitlichen, übergreifenden Mechanismus dar, um datenschutzkonform die korrekte und zeitnahe Umsetzung der Betroffenenrechte an den Standorten und in NUM-RDP zu unterstützen. Dabei verarbeitet die fTTP Consent keine direkt personen-identifizierenden Daten, sondern "kennt" stets nur ausgewählte Einwilligungsinhalte mit pseudonymem Personenbezug (beispielsweise Universitäre Einrichtung, Umfang der Einwilligung, Gültigkeitsdauer der codierten Einwilligungs-Policies).

Die **fTTP Consent** ermöglicht
a) die jederzeitige konsistente Auskunft zum **aktuellen Einwilligungsstand** eines Betroffenen abzufragen und so dem Patientenwillen konsequent zu entsprechen sowie
b) die komplexen **Kommunikationsprozesse** zur Widerrufsumsetzung bei verteilter Datenhaltung zu **orchestrieren**, Aufwände durch **Automatisierung** zu reduzieren und somit die korrekte und aktuelle Umsetzung der Betroffenenrechte für alle beteiligten Einrichtungen und in der zentralen Plattform.

Kontakt:
fttp-num@uni-greifswald.de

Weitere Informationen unter:
https://www.ths-greifswald.de/projekte/num-routine-data-platform/