GEFÖRDERT VOM

Bundesministerium
für Bildung
und Forschung

Zuwendungsempfänger: Europäisches Laboratorium für Molekularbiologie (EMBL)

Thema der Förderung: Verbundprojekt 16DTM108B: Datenschutzrechtliches Reallabor für eine Datentreuhand in der Netzwerkmedizin – TrustDNA; Teilvorhaben: Datenportabilität, Interoperabilität, technisch- organisatorische Maßnahmen

Verantwortliche: Prof. Dr. Jan O. Korbel,

Förderkennzeichen: 16DTM108B

Die Verantwortung für den Inhalt dieser Veröffentlichung liegt beim Autor.

Trust DNA

Im Teil I ist ein Kurzbericht (max. 2 Seiten) mit einer Darstellung des Projektergebnisses vorzulegen. Dieser soll in allgemein verständlicher Form das Vorhaben darstellen und umfasst:

- ***die ursprüngliche Aufgabenstellung sowie den wissenschaftlichen und technischen Stand, an den angeknüpft wurde***

- ***den Ablauf des Vorhabens***

- ***die wesentlichen Ergebnisse sowie ggf. die Zusammenarbeit mit anderen Forschungseinrichtungen***

Das Projekt konzentrierte sich auf die Entwicklung rechtlicher und technischer Grundlagen für den sicheren und FAIRen (auffindbar, zugänglich, interoperabel und nachnutzbar) Datentausch im Rahmen der DNA-MED-Initiative für vernetzte Medizin. Sein primäres Ziel bestand in der Definition eines Rahmens für den Zugriff auf sensitive Omics-Daten durch innovationsgetriebene Unternehmen und internationale Forschende, unter Sicherstellung der Einhaltung deutscher und europäischer Datenschutzvorschriften. Ein entscheidender Bestandteil des Projekts war die Entwicklung und pilotierende Implementierung eines Daten-Treuhandmodells, das Patienten als aktive Teilnehmer in die Governance und Verwaltung ihrer eigenen Daten einbindet.

Am EMBL wurde das Projekt in enger Anbindung an das German Human Genome-Phenome Archive (GHGA) umgesetzt, das Teil der Nationalen Forschungsdateninfrastruktur (NFDI) ist. GHGA bietet als nationales Archiv für aus Menschen gewonnenene Omics-Datensätze eine sichere Infrastruktur, die strengen Datenschutz mit Forschungszugänglichkeit in einem „Zero-Trust-Sicherheitsmodell“ vereint. Dieses folgt dem Grundsatz, keinem Gerät oder keine Applikation automatisch zu vertrauen. Die Architektur umfasst Ende-zu-Ende Verschlüsselung (mit Crypt4GH), dezentral ausgelegte Microservices sowie die lokale Datenspeicherung in Deutschland, um technischen und rechtlichen Anforderungen gerecht zu werden. Das GHGA-Infrastruktur-Team — insbesondere in Heidelberg und Tübingen — hat das Softwaresystem in Python entwickelt, unter Verwendung eines als „Triple Hexagonal Architecture“ formalisierten Rahmenwerks, das als Python-Paket Hexkit implementiert wurde, um Wartbarkeit und Interoperabilität sicherzustellen.

Das EMBL leistete insbesondere Beiträge zur Standardisierung der Risikokategorisierung von Datenfreigabeverfahren auf technischer Ebene und evaluierte Pseudoanonymisierungsstrategien im Zusammenhang mit technischen und organisatorischen Maßnahmen zum Datenschutz. Darüber hinaus hat das EMBL die technischen Anforderungen für die Portierbarkeit und Interoperabilität sensitiver Forschungsdaten in nationalen und internationalen Kontexten ausgearbeitet. Dabei wurden sowohl technische als auch rechtliche Aspekte betrachtet, um einen sicheren und standardkonformen Zugriff für externe Forschende zu gewährleisten. Zur Unterstützung dieser Tätigkeiten wurde außerdem ein hybrides Treuhandmodell geprüft, das verschiedene Governance-Ansätze in einer für die Akteure jeweils passende Weise miteinander vereint.

Während der gesamten Projektlaufzeit wurde in enger Abstimmung mit den nationalen Infrastrukturpartnern und relevanten deutschen Institutionen gearbeitet. Ein besonderer Schwerpunkt lag

auf der Ermittlung allgemein gültiger technischer Spezifikationen und Bewertungsrahmen, um die Einbindung von DNA-MED-Daten in übergeordnete Forschungsökosysteme — zum Beispiel das German Human Genome-Phenome Archive (GHGA) sowie das Federated European Genome-Phenome Archive (fEGA), das derzeit einen internationalen Rahmen für den Austausch humaner Omics-Daten entwickelt — nachhaltig zu unterstützen.

Diese Tätigkeiten wurden im Einklang mit den übergeordneten Zielen der DNA-MED-Initiative ausgeführt — insbesondere der Förderung eines sicheren Datenzugangs, der internationalen Zusammenarbeit sowie der verantwortungsvollen Innovation in der Vernetzten Medizin.

Trust DNA

Im Teil II sind die im Rahmen des Vorhabens durchgeführten Arbeiten ausführlicher darzustellen, insbesondere im Vergleich zur ursprünglichen Vorhabenbeschreibung. Bei Einzelvorhaben soll möglichst ein Umfang von 20 Seiten nicht überschritten werden. Die Verwendung der Zuwendung sowie die erzielten Ergebnisse im Einzelnen müssen nachvollziehbar sein. Dabei sind ergänzend zu den Inhalten darzustellen:

• die wichtigsten Positionen des zahlenmäßigen Nachweises

Im Laufe des Projekts wurden zahlreiche quantifizierbare Meilensteine erreicht, die den Umfang, die Effektivität sowie den Ressourceneinsatz der DNA-MED-Initiative und ihre Integration in GHGA deutlich machen.

Infrastrukturen und Sicherheitsmaßnahmen:
Im Rahmen der Nationalen Forschungsdateninfrastruktur (NFDI) wurde die GHGA-Plattform weiter ausgebaut. Dabei kamen die **Triple Hexagonal Architecture**, Ende-zu-Ende-Verschlüsselung (unter Verwendung von Crypt4GH) sowie das **Hexkit-Python-Paket** zum Einsatz, um die Modularität und Interoperabilität des Systems weiter zu verbessern.

Darüber hinaus wurde bei GHGA ein **mehrschichtiges Sicherheitsframework** implementiert, das folgende Elemente umfasst:

- Alle Daten werden sowohl im Ruhezustand als auch bei der Übertragung verschlüsselt.
- Nach der Aufnahme werden die Daten unter Verwendung von in HashiCorp Vault gespeicherten Secrets neu verschlüsselt.
- Der Download-Zugriff ist zeitlich begrenzt, Token-basiert und ausschließlich für autorisierte Nutzer zugänglich.
- Regelmäßige Schwachstellenprüfungen sowie Penetrationstests durch Dritte gewährleisten die kontinuierliche Sicherheit von Code und Infrastruktur.
- Aus rechtlicher Sicht werden bei GHGA ausschließlich anonymisierte Metadaten entgegengenommen; sämtliche Metadaten werden in aggregierter Form bereitgestellt, um die Identität von Patienten zu schützen.
- Ein vertragsbasierter Zugriffsvorgang unter Einbindung eines „Data Access Committee“ ermöglicht ein standardisiertes und rechtskonformes Verfahren für den Datenzugang.

Die technischen Spezifikationen für die Bereitstellung von DNA-MED-Daten über GHGA sowie der Rahmen zur Beurteilung der internationalen Vergleichbarkeit beim Datenschutz wurden in Zusammenarbeit mit den Projektpartnern ausgearbeitet.

GHGA ist nun als deutscher Knotenpunkt des föderierten European Genome-Phenome Archive (fEGA) voll einsatzbereit.

• die Notwendigkeit und Angemessenheit der geleisteten Projektarbeiten

Die im Rahmen der DNA-MED-Initiative durchgeführten Tätigkeiten waren ebenso notwendig wie angemessen, um das übergeordnete Ziel des Projekts zu erreichen: das Ermöglichen eines sicheren, rechtskonformen und FAIR-en (auffindbaren, zugänglichen, interoperablen und nachnutzbaren) Austauschs sensibler Omics-Daten. Die umgesetzten Tätigkeiten — darunter die Entwicklung eines patientenzentrierten Datentreuhandmodells, die Definition technischer und rechtlicher Zugangsbedingungen für internationale Akteure sowie der Aufbau von GHGA als deutschem Knotenpunkt im European Genome-Phenome Archive (EGA) — adressierten unmittelbar den dringenden Bedarf an robusten Infrastrukturen für die Vernetzte Medizin.

Der Wechsel von der Selbsteingabe von Patientendaten zum Hochladen validierter Dokumente — ausgelöst durch eine 78-prozentige Inkonsistenzrate — verdeutlicht ebenso die Relevanz dieser Tätigkeiten wie ihre methodische Angemessenheit. Auch die Implementierung von Ende-zu-Ende-Verschlüsselung, Token-basiertem Zugriff und penetrationserprobter Architektur durch GHGA belegt die technische Solidität dieser Arbeit. Darüber hinaus hat das EMBL bei der Definition von Datenportierbarkeit und Interoperabilität mit seinem hybriden Treuhandmodell die rechtlichen und wissenschaftlichen Erfordernisse nachhaltig in Einklang gebracht.

Trotz personeller Verzögerungen wurde durch die Neuverteilung von Verantwortlichkeiten an erfahrene EMBL-Teammitglieder Kontinuität und Effizienz sichergestellt. Zusammengefasst waren die durchgeführten Tätigkeiten daher für das Erreichen der Projektziele ebenso entscheidend wie für ihre praktische, umsetzbare und rechtskonforme Realisierung.

• der voraussichtliche Nutzen, insbesondere die Verwertbarkeit des Ergebnisses - auch konkrete Planungen für die nähere Zukunft - im Sinne des fortgeschriebenen Verwertungsplans

Die Ergebnisse des Projekts sind von hohem praktischem Nutzen und stehen im Einklang mit der langfristigen Strategie zum Aufbau nachhaltig ausgelegter, sicherer und interoperabler Dateninfrastrukturen für die biomedizinische Forschung. Die Fortsetzung dieser Arbeit durch GHGA — zum Beispiel bei der Einrichtung von „Trusted Research Environments“ , sowie die Einbindung in europäische Standards stellen sicher, dass die Ergebnisse nachhaltig wirken und über die Laufzeit des Projekts hinaus skaliert werden können.

Darüber hinaus ermöglicht die Ermittlung standardisierter Kriterien zur Beurteilung internationaler Datensicherheitsmaßnahmen GHGA und der DNA-MED-Treuhandstelle, länderübergreifenden Datentausch fundiert zu prüfen und zu erleichtern. Dies schließt eine entscheidende Lücke in der europäischen biomedizinischen Forschung, da bisher mangelnde Harmonisierung beim Datenschutz oft internationale Zusammenarbeit verhindert hat.

Ferner wurde mit der Erprobung eines Treuhandmodells, das Patienten als aktive Teilnehmer in die Governance-Strukturen einbindet, ein übertragbares und ethische Leitsätze respektierendes Governance-Konzept entwickelt. Dieses hat das Potenzial, das Vertrauen in und die Teilnahme an Initiativen der Präzisionsmedizin nachhaltig zu erhöhen.

In Zukunft wird GHGA weiter an „Trusted Research Environments“ arbeiten, um Analysen ohne lokalen Download von Daten zu ermöglichen, und so den Datenschutz weiter zu verbessern. Auch die im Rahmen des Projekts ausgearbeiteten rechtlichen Grundlagen und Governance-Strukturen werden in künftigen Leitlinien für den Betrieb genomischer Dateninfrastrukturen in Europa Berücksichtigung finden.

Hinsichtlich Übertragbarkeit und Nachhaltigkeit sind die Ergebnisse so ausgelegt, dass sie weiterbestehen und ausgeweitet werden können; zum Beispiel:

- GHGA ist in das NFDI-Programm eingebettet und daher durch eine langfristige institutionelle und finanzielle Unterstützung abgesichert.

- GHGA hat zum föderierten European Genome-Phenome Archive (fEGA) beigetragen einer Initiative, die sich als internationale Referenzinfrastruktur für den Austausch human-genomischer Daten etabliiert.

- EMBL, DKFZ und die Universität Heidelberg sind entschlossen, die technischen Rahmenwerke für weitere Projekte zu nutzen und auszubauen.

- Die im Rahmen des Projekts entwickelten Werkzeuge und Methoden (Frameworks und Bibliotheken) stehen bereits für den Einsatz in anderen biomedizinischen und datengetriebenen Forschungsprojekten zur Verfügung:

    - Hexkit
    - Metldata
    - Schemapack
    - Microservice Repo Template

• der während der Durchführung des Vorhabens dem Zuwendungsempfänger bekannt gewordenen Fortschritt auf dem Gebiet des Vorhabens bei anderen Stellen

Während der Durchführung des Vorhabens verfolgten die Universität Heidelberg und das EMBL aufmerksam relevante Entwicklungen bei anderen Einrichtungen, insbesondere solche, die mit den Zielen eines sicheren und FAIRen (auffindbaren, zugänglichen, interoperablen und nachnutzbaren) Datenaustauschs im Einklang stehen.

Ein bedeutender Fortschritt wurde im Rahmen des German Human Genome-Phenome Archive (GHGA) erzielt, das Teil der Nationalen Forschungsdateninfrastruktur (NFDI) ist. GHGA entwickelte eine sichere, dezentral angelegte Infrastruktur für menschliche Omics-Daten, einschließlich eines „Zero-Trust“-Sicherheitsmodells, der Verwendung von Crypt4GH zur Ende-zu-Ende-Verschlüsselung sowie der lokalen Speicherung sensitiver Daten in Deutschland zur Erfüllung technischer und rechtlicher Anforderungen.

Darüber hinaus wurden im Projekt laufende Entwicklungen im Rahmen des Federated European Genome-Phenome Archive (fEGA) berücksichtigt, das derzeit einen internationalen Rahmen für den Austausch humaner Genomdaten aufbaut. Diese externen Fortschritte flossen in die Arbeiten des

Projekts zur Standardisierung der Risikokategorisierung von Datenfreigabeverfahren, zur Evaluierung von Pseudonymisierungsstrategien sowie zur Konzeption eines hybriden Datentreuhandmodells ein.

Während der gesamten Projektlaufzeit fand eine enge Abstimmung mit nationalen Infrastrukturpartnern und relevanten deutschen Institutionen statt. Dadurch konnte sichergestellt werden, dass externe Fortschritte erkannt, berücksichtigt und in das Vorhaben integriert wurden, um eine nachhaltige Einbindung von DNA-MED in übergeordnete Forschungssysteme und internationale Standards zu ermöglichen.

• die erfolgten oder geplanten Veröffentlichungen des Ergebnisses nach Nr. 5 der NABF

Zum aktuellen Zeitpunkt wurden die Forschungsergebnisse noch nicht in einer Publikation verwendet. Sollte dies in der Zukunft erfolgen, wird eine Nennung des Projekts und des Fördermittelgebers erfolgen.