Deutsche Version (siehe Unten)
Version française (ci-dessous)

Lay Summary

Project title	RDeID: Risk-based de-identification platform for health-related data
Main applicant	Dr. Jean Louis Raisaro, CHUV
Consortium	**CHUV** Jean Louis Raisaro (main applicant), Solange Zoergiebel (co-applicant) **USZ** Katie Kalt (co-applicant), Patrick Hirschi (co-applicant) **SIB** Frédéric Erard (associated applicant) **Charité (Berlin, Germany)** Fabian Prasser (associated applicant)
Short Summary	Build a web service for automatic risk assessment and de-identification of health datasets.
Background	According to the International Organization for Standardization (ISO), *de-identification* represents a general term for any process of reducing the association between a set of identifying data and the data subject[1]. Particularly, de-identification can lead to *pseudonymized data,* when it removes the association with a data subject and adds an association between a particular set of characteristics relating to the data subject and or more pseudonyms, or to *anonymized data,* when personal data is irreversibly altered in such a way that a data subject can no longer be identified directly or indirectly. In the scope of secondary use of healthcare-related data, de-identification is recognized as an essential method for privacy protection and a prerequisite for broad health data sharing. Although the concept of data de-identification is straightforward, its application is harder in practice. The current legal framework in Switzerland does not provide any specific guidance or methods – as opposed to the American HIPAA regulation and the Safe Harbor Rule – that must be applied to correctly de-identify personal data. Instead, for a correct and compliant de-identification, it is necessary to assess the residual risk that a person with access to the data could re-identify the data subjects, considering all relevant circumstances. To help the broad SPHN research community in performing a sound and re-identification risk assessment and therefore a correct and consistent de-identification across the various stakeholders, a “Data De-identification Project (DeID)” task force, made of Swiss university hospital representatives and technical and legal opinion leaders in the data privacy, was set up by the SPHN Data Coordination Center during the first phase of SPHN (2018-2021). As a result of its activities, the DeID task force elaborated a set of recommendations

[1] https://www.iso.org/obp/ui/#iso:std:iso:25237:ed-1:v1:en

	for de-identifying health-related data in compliance with Swiss legal requirements based on a phased approach where the degree of the de-identification should be based on the risk associated with a given data transfer. An Excel template for "use case evaluation and risk assessment" was also produced with the hope that it could serve as a reference for the broad research community.
Goal	The goal of this demonstrator project is to build on the work of the SPHN DeID task force and implement its recommendations into a modular and extensible Web platform for automating and harmonizing risk assessment and de-identification for biomedical research projects.
Significance	The development of such a platform and its validation across the University Hospitals of Lausanne and Zurich has the potential to establish a common tool for the SPHN community (and beyond) that can tremendously accelerate access to health-related data by researchers by de-facto (i) representing a standard for ethics committees and hospital data protection officers and (ii) harmonizing data de-identification processes across health data providers in Switzerland for both local and multi-centric projects.

Deutsch

Projekttitel	RDeID: Risikobasierte De-Identifizierungsplattform für gesundheitsbezogene Daten
Zusammenfassung	Aufbau eines Webdienstes zur automatischen Risikobewertung und De-Identifizierung von Gesundheitsdatensätzen.
Hintergrund	Laut der Internationalen Organisation für Normung (ISO) gilt 'De-Identifizierung' als allgemeiner Begriff für jeden Prozess, der den Zusammenhang zwischen einer Reihe von identifizierenden Daten und dem Datensubjekt verringert. De-Identifizierung kann zu *pseudonymisierten Daten* führen, wenn die Verbindung zu dem Datensubjekt aufgehoben wird und eine Verbindung zwischen einem bestimmten Satz von Merkmalen des Datensubjekts und oder mehreren Pseudonymen hinzufügt wird. Andererseits kann De-Identifizierung zu *anonymisierten Daten* führen, wenn personenbezogene Daten irreversibel so verändert werden, dass eine betroffene Person nicht mehr direkt oder indirekt identifiziert werden kann. Im Rahmen der Sekundärnutzung von Gesundheitsdaten wird die De-Identifizierung als wesentliche Methode zum Schutz der Privatsphäre und als Voraussetzung für eine umfassende gemeinsame Nutzung von Gesundheitsdaten anerkannt. Obwohl das Konzept der De-Identifizierung von Daten relative klar ist, ist die Anwendung in der Praxis komplex. Die derzeitigen rechtlichen Rahmenbedingungen in der Schweiz bieten - im Gegensatz zur amerikanischen HIPAA-Verordnung und der Safe Harbor Rule - keine spezifischen Leitlinien oder Methoden für eine korrekte und konforme De-Identifizierung personenbezogener Daten. Stattdessen muss dafür das Risiko beurteilt werden, mit dem Datensubjekte - unter Berücksichtigung aller relevanten Umstände -

	durch eine Person mit Datenzugang identifiziert werden könnten. Um die breite SPHN-Forschungsgemeinschaft bei der Durchführung einer fundierten Risikobewertung der Re-Identifizierung und damit einer korrekten und innerhalb der diversen Partner, konsistenten De-Identifizierung zu unterstützen, wurde in der ersten Phase von SPHN (2018-2021) die 'Data De-Identification Projekt (DeID)' Task Force geschaffen. Diese setzte sich aus Vertretern der Schweizer Universitätsspitälern sowie technischen und juristischen Meinungsbildnern im Bereich des Datenschutzes zusammen und wurde durch das SPHN-Datenkoordinationszentrum geleitet. Das Resultat der DeID-Taskforce Arbeit sind eine Reihe von gesetzeskonformen Empfehlungen für die De-Identifizierung gesundheitsbezogener Daten. Die Empfehlungen sind dynamisch, so dass der Grad der De-Identifizierung auf dem mit einer bestimmten Datenübertragung verbundenen Risiko basieren sollte. Ausserdem wurde eine Excel-Vorlage erarbeitet zur "praktischen Fall- und Risikobewertung", in der Hoffnung, dass es der Gemeinschaft als Referenz für die risikobasierte De-Identifizierung dienen kann.
Das Ziel	Das Ziel dieses Demonstrator Projekt ist es, auf der Grundlage der Arbeit der DeiD-Task Force eine modulare und erweiterbare Webplattform zu erstellen, mit der Risikobewertungen und Desidentifizierungspraktiken für biomedizinische Forschungsprojekte automatisiert und harmonisiert werden können.
Bedeutung	Die Entwicklung einer solchen Plattform und ihre Validierung in den Universitätsspitälern Lausanne und Zürich hat das Potenzial, ein gemeinsames Instrument für biomedizinische Forschungsgemeinschaft (und darüber hinaus) zu schaffen. Dies wird den Zugang von Forschern zu Gesundheitsdaten beschleunigen, indem es de facto (i) einen Standard für Ethikkommissionen und Datenschutzbeauftragte der Krankenhäuser darstellt und (ii) die Prozesse zur De-Identifizierung von Daten bei Gesundheitsdatenanbietern in der Schweiz sowohl für lokale als auch für multizentrische Projekte harmonisiert..

Français

Titre du projet	RDeID: Service de désidentification de données de santé basé sur une évaluation de risque
Résumé	Développer un service web offrant une évaluation formelle de risque de réidentification, et la désidentification automatique des jeux de données de santé.
Context	D'après la norme ISO, la *désidentification* est un terme général qui désigne tout processus réduisant l'association entre un ensemble de données d'identification et la personne concernée[2]. En particulier, la désidentification peut produire des

[2] https://www.iso.org/obp/ui/#iso:std:iso:25237:ed-1:v1:fr

	données *pseudonymisées* quand elle retire l'association des données avec la personne et ajoute une association entre un ensemble de caractéristiques de la personne et un ou plusieurs pseudonymes. Elle peut aussi produire des données *anonymisées*, quand les données personnelles sont altérées de manière irréversible de manière à empêcher tout lien direct ou indirect avec la personne concernée. Dans le contexte de la réutilisation des données de santé, la désidentification est reconnue comme mesure phare pour la protection de la vie privée et une condition au partage à large échelle des données. Le concept de désidentification est assez équivoque, mais sa mise en œuvre est bien plus compliquée en pratique. Le cadre légal actuel en Suisse ne propose pas de réelle méthode, contrairement à la règle étatsunienne HIPAA dite « Safe Harbor rule » précisant les mesures concrètes à prendre pour désidentifier correctement les données personnelles. Néanmoins, pour une désidentification correcte et légale, il est nécessaire d'évaluer le risque qu'une personne ayant accès aux données puisse réidentifier des personnes concernées, compte tenu de son environnement. Pour aider la communauté de recherche SPHN à établir une évaluation de risque fiable et cohérente – et par là une désidentification correcte pour tous les protagonistes, a été réunie une équipe de représentant.e.s des hôpitaux universitaires suisses ainsi que d'expert.e.s techniques et légistes en protection des données. Cette équipe a mené a bien le projet de De-identification de Données (DeID) durant la première phase d'SPHN (2018-2021). Cette équipe a produit un ensemble de recommandations pour une désidentification des données de santé conforme à la loi. Ces recommandations sont dynamiques, de telle sorte que le degré de désidentification recommandé est basé sur le risque réel associé à un transfert spécifique de données. Un squelette Excel « d'évaluation de cas pratique et de risque » a été écrit, dans l'espoir qu'il serve de référence à la communauté pour la désidentification basée sur évaluation de risque.
But	Le but de ce projet de démonstration est de construire à partir du travail de l'équipe du projet DeiD une plateforme Web modulaire et extensible capable d'automatiser et d'harmoniser les évaluations de risque et les pratiques de désidentifications des projets de recherche biomédicale.
Importance	Le développement d'une telle plateforme et sa validation dans les hôpitaux de Lausanne et Zurich permettront d'établir un outil commun à la communauté de recherche biomédicale. Cela permettra d'accélérer l'accès des chercheurs.es aux données de santé et homogénéiser et de rendre légalement conformes les pratiques de dé-identification des jeux de données à travers les hôpitaux Suisses, et pour des études locales comme collaboratives.