Viewpoint

Revolutionizing Medical Data Sharing Using Advanced Privacy-Enhancing Technologies: Technical, Legal, and Ethical Synthesis

James Scheibner[1,2], BComp, LLB, PhD; Jean Louis Raisaro[3,4], BSc, MSc, PhD; Juan Ramón Troncoso-Pastoriza[5], BSc, MSc, MPhil, PhD; Marcello Ienca[1], BA, MA, MSc, PhD; Jacques Fellay[3,6,7], MD, PhD; Effy Vayena[1], BA, MSc, PhD; Jean-Pierre Hubaux[5], Dr-Eng

[1]Health Ethics and Policy Laboratory, Department of Health Sciences and Technology, Eidgenössische Technische Hochschule Zürich, Zürich, Switzerland

[2]College of Business, Government and Law, Flinders University, Adelaide, Australia

[3]Precision Medicine Unit, Lausanne University Hospital, Lausanne, Switzerland

[4]Data Science Group, Lausanne University Hospital, Lausanne, Switzerland

[5]Laboratory for Data Security, School of Computer and Communication Sciences, École polytechnique fédérale de Lausanne, Lausanne, Switzerland

[6]School of Life Sciences, École polytechnique fédérale de Lausanne, Lausanne, Switzerland

[7]Host-Pathogen Genomics Laboratory, Swiss Institute of Bioinformatics, Lausanne, Switzerland

Corresponding Author:
James Scheibner, BComp, LLB, PhD
College of Business, Government and Law
Flinders University
Ring Road, Bedford Park
Adelaide, 5042
Australia
Phone: 61 (08) 8201 3196
Email: james.scheibner@flinders.edu.au

Abstract

Multisite medical data sharing is critical in modern clinical practice and medical research. The challenge is to conduct data sharing that preserves individual privacy and data utility. The shortcomings of traditional privacy-enhancing technologies mean that institutions rely upon bespoke data sharing contracts. The lengthy process and administration induced by these contracts increases the inefficiency of data sharing and may disincentivize important clinical treatment and medical research. This paper provides a synthesis between 2 novel advanced privacy-enhancing technologies—homomorphic encryption and secure multiparty computation (defined together as multiparty homomorphic encryption). These privacy-enhancing technologies provide a mathematical guarantee of privacy, with multiparty homomorphic encryption providing a performance advantage over separately using homomorphic encryption or secure multiparty computation. We argue multiparty homomorphic encryption fulfills legal requirements for medical data sharing under the European Union's General Data Protection Regulation which has set a global benchmark for data protection. Specifically, the data processed and shared using multiparty homomorphic encryption can be considered anonymized data. We explain how multiparty homomorphic encryption can reduce the reliance upon customized contractual measures between institutions. The proposed approach can accelerate the pace of medical research while offering additional incentives for health care and research institutes to employ common data interoperability standards.

(J Med Internet Res 2021;23(2):e25120) doi: 10.2196/25120

KEYWORDS

encryption; anonymization; pseudonymization; centralized approach; decentralized approach; federated approach; Interoperability; privacy; GDPR; General Data Protection Regulation; data privacy; data protection; ethics; research; data sharing; data governance; patient data privacy

Introduction

The current biomedical research paradigm has been characterized by a shift from intrainstitutional research toward multiple collaborating institutions operating at an interinstitutional, national or international level for multisite research projects; however, despite the apparent breakdown of research barriers, there remain differences between ethical and legal requirements at all jurisdictional levels [1]. There are numerous organizational strategies that have been used to resolve these issues, particularly for international academic consortia.

For example, the International Cancer Genome Consortium endeavors to amass cancer genomes paired with noncancerous sequences in a cloud environment, known as pancancer analysis of whole genomes. The International Cancer Genome Consortium's data access compliance office was unable to establish an international cloud under the Pancancer Analysis of Whole Genomes Project because of conflicts between United States and European Union data privacy laws [2]. These conflicts will be likely exacerbated with the Court of Justice of the European Union (CJEU) invalidating the United States–European Union Privacy Shield agreement. This decision will prevent private research organizations from transferring personal data from the European Union to the United States without organizational safeguards [3]. In addition, the COVID-19 pandemic has made sharing data for clinical trials and research imperative. However, a series of COVID-19 papers retracted due to data unavailability emphasizes the need for data sharing to encourage oversight [4]. Furthermore, within the European Union there is the potential for differences in how countries regulate the processing of health-related personal data [5]. There are also different grounds to justify processing of health-related data under separate branches of EU law. The Clinical Trials Regulation and the European Union General Data Protection Regulation (GDPR) require different standards of consent for processing health-related data, depending on whether those data are collected as part of a clinical trial protocol or not. The effect of this difference is that data collected for one purpose, such as a trial protocol, may not be made available for a secondary research purpose if appropriate consent has not been obtained [6]. Finally, given study restrictions it may be impossible to share data between institutions or jurisdictions [7]. Although reforms to EU data protection law have been proposed to encourage scientific data sharing [8], at present the best available solutions remain contractual and technological measures.

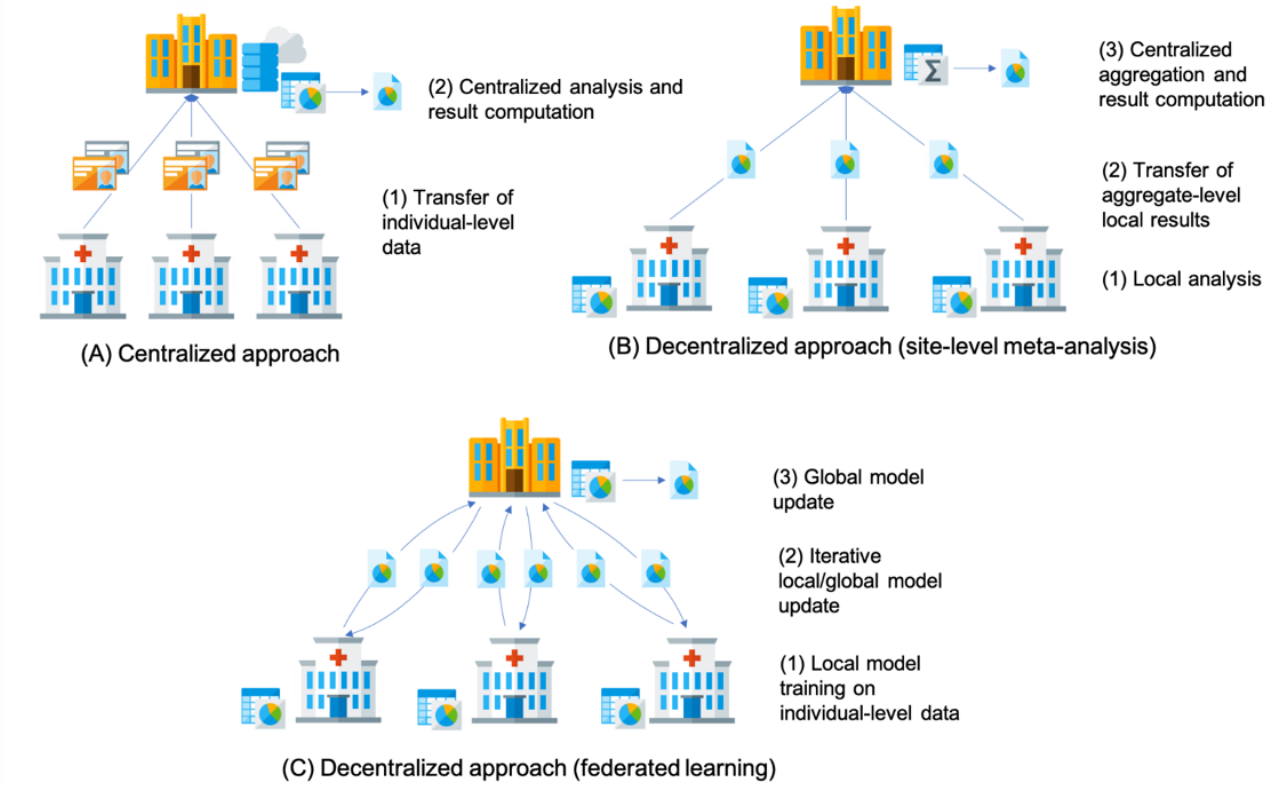
In this paper, we describe how traditional data-sharing approaches relying upon conventional privacy-enhancing technologies are limited by various regulations governing medical use and data sharing. We describe two novel privacy-enhancing technologies, homomorphic encryption and secure multiparty computation, that extend the capacity of researchers to conduct privacy-preserving multisite research. We then turn to analyze the effects of regulation on using these novel privacy-enhancing technologies for medical and research data sharing. In particular, we argue these privacy-enhancing technologies guarantee anonymity as defined under the EU GDPR and are, therefore, key enablers for medical data sharing. We focus on the GDPR, as it currently represents a global benchmark in data protection regulations. We argue that using these technologies can reduce the reliance upon customized data-sharing contracts. The use of standardized agreements for multiparty processing of data in concert with privacy-enhancing technologies can reduce the bottleneck on research. Finally, we turn to address how these novel privacy-enhancing technologies can be integrated within existing regulatory frameworks to encourage increased data sharing while preserving data privacy.

Privacy and Security Issues of Current Medical Data-Sharing Models

Overview

Before examining novel privacy-enhancing technologies, it is necessary to examine the main models for exchanging medical data for research purposes and the limitations of conventional privacy protection mechanisms that are currently used to reduce the risk of reidentification. We synthesize the data-sharing models into three categories and analyze their main technological issues (Figure 1).

Figure 1. Overview of the three main data-sharing models: (A) centralized, (B) decentralized (site-level meta-analysis), and (C) decentralized (federated learning).



Centralized Model: Trusted Dealer

The centralized model requires medical sites (ie, data providers) that are willing to share data with each other to pool their individual-level patient data into a single repository. The data repository is usually hosted by one medical site or by an external third party (eg, a cloud provider), playing the trusted dealer role. The main advantage of this model is that the trusted dealer enables authorized investigators to access all the patient-level information needed for data cleaning and for conducting statistical analysis. Moreover, such a data-sharing model minimizes infrastructure costs at medical sites, as data storage and computation are outsourced. However, from a data privacy perspective the centralized model is often difficult to realize, especially when medical and genetic data should be exchanged across different jurisdictions. The central site hosting the data repository represents a single point of failure in the data-sharing process. All participating sites must trust such single entity for protecting their patient-level data [9].

To minimize sensitive information leakage from data breaches, traditional anonymization techniques include suppressing directly identifying attributes, as well as the generalizing, aggregating or randomizing quasi-identifying attributes in individual patient records. In particular, the *k*-anonymity privacy model [10] is a well-established privacy-preserving model that aims to reduce the likelihood of reidentification attacks singling out an individual. Specifically, the *k*-anonymity model ensures that for each combination of quasi (or indirect) identifier, there exists at least *k* individuals who share the same attributes.

However, given the increased sophistication of reidentification attacks [10-16] and the rising dimensionality (number of clinical and genetic attributes) of patient data, the above-mentioned countermeasures are inadequate to ensure a proper level of anonymization and preserve acceptable data utility. As a result, these conventional anonymization techniques for individual-level patient data are rarely used in practice. Researchers prefer to rely upon simple pseudonymization techniques (such as replacing direct identifiers with pseudonymous codes) combined with legal measures defining each party's responsibilities regarding data transfer, access, and use. This process generates administrative overheads that slow down the pace of biomedical research. Furthermore, although designed to comply with data protection regulations, contractual safeguards may not eliminate the risk of individuals being reidentified [17]. As we argue below, combining traditional pseudonymization mechanisms and governance strategies meets the legal standard of pseudonymization but not anonymization under the GDPR.

Decentralized Model: Site-Level Meta-analysis

As opposed to the centralized data-sharing model, the decentralized model does not require patient-level data to be physically transferred out of the medical sites' information technology infrastructure. Medical sites keep control over their individual-level patient data and define their own data governance principles. For each clinical study, the statistical analysis is first computed on local data sets. The resulting local statistics are then sent to the site responsible for the final meta-analysis that aggregates the separate contribution of each data provider [18] to obtain the final result of the analysis. Under

this model, the site performing the meta-analysis is trusted by all other sites for the protection of their local statistics. As local statistics have a significantly lower dimensionality with respect to individual-level data, there is a lower risk of reidentification in the decentralized data-sharing model.

However, the sharing of only aggregate-level data does not guarantee patients' privacy by itself. Some aggregate-level statistics may be too low for certain subpopulations (such as patients with rare diseases) and can be considered personally identifying. Moreover, in some circumstances aggregate-level data from local analyses can be exploited to detect the presence of target individuals in the original data set. For example, an attacker may already hold the individual-level data of 1 or several target individuals [19-23]. This membership information can be subsequently used to infer sensitive and sometimes stigmatizing attributes of the target individuals. For example, detecting the membership of an individual in a HIV-positive cohort reveals their HIV status. The intuition behind these attacks is to measure the similarity between the individual-level target data with statistics computed from the study data set and statistics computed from the general population. The attacker's certainty about the target's membership in the data set increases with the similarity of the target's data to the statistics derived from the study data set.

To address these inference attacks, clinical sites can anonymize their local statistics by applying obfuscation techniques that mainly consist in adding a certain amount of statistical noise on the aggregate-level data before transfer to third parties. This process enables data providers to achieve formal notions of privacy such as differential privacy [24,25]. In the statistical privacy community, differential privacy is currently considered as guaranteeing the likelihood of reidentification from the release of aggregate-level statistics can be minimized to an acceptable value. Similar to anonymization techniques for individual-level data, statistical obfuscation techniques degrade the utility of aggregate-level data. Consequently, the amount of noise introduced by data obfuscation should be carefully calibrated to reach the desired compromise between utility and privacy. Often, when each data provider adds the required amount of noise to reach an acceptable level of privacy, the resulting aggregated results stemming from a meta-analysis are too distorted to be reliable [26].

Beyond privacy considerations, this approach also suffers from a lack of flexibility as the medical sites involved in the analysis must coordinate before the analysis on the choice of parameters and covariates to be considered. This coordination often depends on manual approval, impeding the pace of the analysis itself. Finally, as opposed to the centralized approach, accuracy of results from a meta-analysis that combines the summary statistics or results of local analysis can be affected by cross-study heterogeneity. This can lead to inaccurate and misleading conclusions [27].

Decentralized Model: Federated Analysis and Learning

The federated model is an evolution of the decentralized model based on site-level meta-analysis. Instead of sharing the results of local analyses, the participating data providers collaborate to perform a joint analysis or the training of a machine learning model in an interactive and iterative manner, only sharing updates of the model's parameters. One of the medical sites participating in the multicentric research project (typically the site responsible for the statistical analysis) becomes the reference site (or central site) and defines the model to be trained (or analysis to be performed) and executed on the data distributed across the network. This model is referred to as the global model. Each participating site is given a copy of the model to train on their own individual-level data. Once the model has been trained locally over several iterations, the sites send only their updated version of the model parameters (aggregate-level information) to the central site and keep their individual-level data at their premises. The central site aggregates the contributions from all the sites and updates the global model [28]. Finally, the updated parameters of the global model are shared again with the other sites. The process repeats iteratively till convergence of the global model.

With respect to the distributed data-sharing approach based on site-level meta-analysis, this federated approach is more robust against heterogeneous distributions of the data across different sites, thus yielding results accuracy that is comparable to the results obtained with the same analysis conducted using the centralized model. Moreover, this approach does not suffer from the loss in statistical power of conventional meta-analyses. Prominent projects that have attempted to employ federated approaches to analysis and sharing of biomedical data are the DataSHIELD project [29] and the Medical Informatics Platform of the Human Brain Project [30].

The federated data-sharing approach combines the best features of the other two approaches. However, although the risk or reidentification is reduced compared to the centralized approach, the federated approach remains vulnerable to the same inference attacks of the meta-analysis approach. These inference attacks exploit aggregate-level data released during collaboration [31-34]. The potential for an inference attack is even increased compared to a meta-analysis-based approach. This is due to the iterative and collaborative nature of the data processing, allowing adversaries to observe model changes over time and with specific model updates. Melis et al [35] show that updates of model parameters transferred during the collaborative training phase can be used to infer the membership of a target individual in the training data sets as well as some properties associated with a particular subset of the training data. This inference is possible if the context of the data release enables the attacker to easily access some auxiliary individual-level information about the target individual. In legal terms (as discussed below), these aggregate-level data can potentially be considered personal data. As for the meta-analysis approach, obfuscation techniques can be used to anonymize the model's updates at each iteration. Nevertheless, the required perturbation can severely affect the performance of the final model [26].

Finally, regardless of the type of distributed data-sharing model, obfuscation techniques for anonymizing aggregate-level data are rarely used in practice in medical research because of their impact on data utility. As a result, these technical privacy limitations are usually addressed via additional legal and organizational mechanisms. For the DataSHIELD project, access is limited to organizations that have consented to the terms of

use for DataSHIELD and have sought appropriate ethics approval to participate in a DataSHIELD analysis [36]. Therefore, implementing the platform will require cooperating with governments and institutions so they are comfortable with exposing sensitive data to the platform [29]. However, as we discuss below, advanced technologies can also guarantee data privacy.

Minimizing Risks by Leveraging Advanced Privacy-Enhancing Technologies

Overview

In the last few years, several cryptographic privacy-enhancing technologies have emerged as significant potential advances for addressing the above-mentioned data protection challenges that still affect medical data sharing in the decentralized model. Although hardware-based approaches could be envisioned for this purpose, they are usually tailored to centralized scenarios and introduce a different trust model involving the hardware provider. Furthermore, they also depend on the validity of the assumptions on the security of the hardware platform, for which new vulnerabilities are constantly being discovered. In this paper, we focus on two of the most powerful software-based privacy-enhancing technologies: homomorphic encryption and secure multiparty computation. Both rely upon mathematically proven guarantees for data confidentiality, respectively grounded on cryptographic hard problems and noncollusion assumptions.

Homomorphic Encryption

Homomorphic encryption [37] is a special type of encryption that supports computation on encrypted data (ciphertexts) without decryption. Thanks to this property, homomorphically encrypted data can be securely handed out to third parties, who can perform meaningful operations on them without learning anything about their content. Fully homomorphic encryption schemes, or schemes enabling arbitrary computations on ciphertexts, are still considered nonviable due to the high computational and storage overheads they introduce. Current practical schemes that enable only a limited number of computations on ciphertexts (such as polynomial operations) have reached a level of maturity that permits their use in real scenarios.

Secure Multiparty Computation

Secure multiparty computation [38-42] protocols enable multiple parties to jointly compute functions over their private inputs without disclosing to the other parties more information about each other's inputs than what can be inferred from the output of the computation. This class of protocols is particularly attractive in privacy-preserving distributed analytic platforms due to the great variety of secure computations they enable. However, this flexibility includes several drawbacks that hinder their adoption, including high network overhead and the requirement for parties to remain online during computation.

Multiparty Homomorphic Encryption

The combination of secure multiparty computation and homomorphic encryption was proposed to overcome their respective overheads and technical limitations; we refer to it as multiparty homomorphic encryption [43-46]. Multiparty homomorphic encryption enables flexible secure processing by efficiently transitioning between encrypted local computation, performed with homomorphic encryption, and interactive protocols (secure multiparty computation). It can be used to choose the most efficient approach for each step within a given workflow, leveraging the properties of one technique to avoid the bottlenecks of the other. Moreover, multiparty homomorphic encryption ensures that the secret key of the underlying homomorphic encryption scheme never exists in full. Instead, it distributes the control over the decryption process across all participating sites, each one holding a fragment of the key. All participating sites have to agree to enable the decryption of any piece of data, and no single entity alone can decrypt the data.

Unlike homomorphic encryption or secure multiparty computation alone, multiparty homomorphic encryption provides effective, scalable, and practical solutions for addressing the privacy-preserving issues that affect the distributed or federated approach for data sharing. For example, systems such as Helen [47], MedCo [48], or POSEIDON [49] use multiparty homomorphic encryption to guarantee that all the information interchanged between the sites is always in encrypted form, including aggregate data such as model parameters and model updates, and only the final result (the computed model or the predictions based on this model) is revealed to the authorized user. Finally, multiparty homomorphic encryption reduces the need of obfuscation techniques to protect aggregate-level data from inference attacks. Furthermore, data utility, which is typically lost with privacy-preserving distributed approaches that only rely upon obfuscation techniques, can be significantly improved. As aggregate-level data transfer and processing across participating sites during the analysis or training phase remains always encrypted, obfuscation can be applied only to the decrypted final result of the analysis that is released to the data analyst, instead of being applied to all local model updates at each iteration. Hence, multiparty homomorphic encryption enables a much lower utility degradation for the same level of reidentification risk.

Regulatory Hurdles for the Use of Encryption Technologies

Overview

In this section, we focus on the features of EU data protection law concerning encryption and data sharing. We focus on the GDPR because of the persistence of national divergences in member state law, despite the passage of the GDPR. In particular, the GDPR provides member states can introduce further conditions, including restrictions on processing of genetic data, biometric data, or health-related data. These exceptions exist outside the narrow circumstances in which special categories of personal data, which genetic data, biometric data, or health-related data belong to, can be processed [6]. This flexibility increases the potential for divergences in national law that require customized contracts between institutions in different EU member states [5].

Data Anonymization and Pseudonymization

The GDPR defines *personal data* as concerning an identifiable natural person. Therefore, pseudonymized data, where all identifiers have been removed from those data, remain personal data. However, the provisions of the GDPR do not concern anonymized data or data which have been processed so individuals are no longer identifiable. In particular, anonymized data may be used for research or statistical processing without the need to comply with the GDPR.

Spindler and Schmechel [50] note there are two conflicting approaches to classifying personal and anonymized data. The first is an absolute approach, where anonymized data constitute personal data if there is even a theoretical chance of reidentification. This approach represents the state of national law in a minority of EU member states, such as France [51]. The second is the relative approach, where anonymized data are no longer personal data if it is reasonably likely that methods do not exist to reidentify individuals [50]. This approach represents the state of national law in countries such as Ireland, where the Irish Data Protection Commission has held that data are anonymized if it is unlikely current technology can be used to reidentify those data [52]. Likewise, the German Federal Ministry for Economic Affairs and Energy held that data (including health-related personal data) are anonymized under the *Bundesdatenschutzgesetz* (*German Federal Data Protection Act*) where individuals cannot be reidentified with reasonable effort [53]. In both these jurisdictions, if an unreasonable effort were required to reidentify anonymized data, then it would no longer be personal data [50].

At the supranational level, the former Article 29 Working Party (now the European Data Protection Board) has favored a relative over an absolute approach to anonymization. First, the Article 29 Working Party held that the words "means reasonably likely" suggests a theoretical possibility of reidentification will not be enough to render those data personal data [54]. A subsequent opinion of the Working Party reinforced this support for the relative approach and compared different techniques for anonymization or pseudonymization. For example, encrypting data with a secret key means that data could be decrypted by the key holder. For this party, the data would therefore be pseudonymized data. But, if a party does not have the key, the data would be anonymized. Likewise, if data are aggregated to a sufficiently high level, these data would no longer be personal data [55]. Nevertheless, following the Article 29 Working Party's ruling, no single anonymization technique can fully guard against orthogonal risks of reidentification [56].

Data Processing

The GDPR's provisions apply to data controllers, or entities determining the purpose and means of processing personal data. This definition encompasses both health care institutions and research institutions. Data controllers must guarantee personal data processing is lawful, proportionate, and protects the rights of data subjects. In particular, the GDPR provides that encryption should be used as a safeguard when personal data are processed for a purpose other than which they were collected. Although the GDPR does not define encryption, the Article 29 Working Party treats encryption as equivalent to stripping identifiers from personal data. The GDPR also lists encryption as a strategy that can guarantee personal data security. Furthermore, the GDPR emphasizes that data controllers should consider the state of the art, along with the risks associated with processing, when adopting security measures. The GDPR also provides that data processing for scientific purposes should follow the principle of data minimization. This principle requires data processors and controllers to use nonpersonal data unless the research can only be completed with personal data. If personal data are required to complete the research, pseudonymized or aggregate data should be used instead of directly identifying data.

The GDPR imposes obligations on data controllers with respect to the transfer of data, particularly outside of the European Union. Specifically, the GDPR requires the recipient jurisdiction to offer adequate privacy protection before a data controller transfers data there. Otherwise, the data controller must ensure there are organizational safeguards in place to ensure the data receives GDPR-equivalent protection. Furthermore, data controllers must consider the consequences of exchanging data between institutions, and whether these are joint controllership or controller–processor arrangements. Under the GDPR, data subject rights can be exercised against any and each controller in a joint controllership agreement. Furthermore, controllers must have in place an agreement setting out the terms of processing. By contrast, a data controller-processor relationship exists where a controller directs a data processor to perform processing on behalf of the controller, such as a cloud services provider. The GDPR provides that any processing contract must define the subject matter, duration, and purpose of processing. Contracts should also define the types of personal data processed and require processors to guarantee both the confidentiality and security of processing.

Advanced Privacy-Enhancing Technologies and EU Data Governance Requirements

In this section, we argue that multiparty homomorphic encryption, or homomorphic encryption and secure multiparty computation used in concert, meets the requirements for anonymization of data under the GDPR. Furthermore, we argue the use of multiparty homomorphic encryption can significantly reduce the need for custom contracts to govern data sharing between institutions. We focus on genetic and clinical data sharing due to the potential for national derogations pertaining to the processing of health-related data. Nevertheless, our conclusions regarding the technical and legal requirements for data sharing using multiparty homomorphic encryption, or homomorphic encryption and secure multiparty computation, may apply to other sectors, depending on regulatory requirements [57].

Under the GDPR, separating pseudonymized data and identifiers is analogous to separating decryption keys and encrypted data. For pseudonymized data, any entity with physical or legal access to the identifiers will possess personal data [58]. To this end, Spindler and Schmechel [50] suggest that encrypted data remain personal data to the entity holding the decryption keys. The encrypted data also remain personal data for any third party with lawful means to access the decryption keys. Applying this

approach to homomorphic encryption, if a party has access to the decryption key corresponding to the encryption key that was used to homomorphically encrypt data, that party will have access to personal data. Likewise, if a party has lawful access to data jointly processed as part of secure multiparty computation, those data will remain personal data for that party [59].

Whether a party to data processing using advanced privacy-enhancing technologies has lawful access to data or decryption keys depends on the legal relationship between the parties. With respect to joint controllership, recent CJEU case law has established that parties can be joint controllers even without access to personal data [60-62]. The CJEU held that the administrator of a fan page hosted on Facebook was a joint controller despite only having access to aggregate data in paragraph 38 [60];however, Article 26, paragraph 1 of the GDPR [63] requires that joint controllers establish a contract allocating responsibility for processing of personal data. Hospitals or research institutions processing patient data using secure multiparty computation jointly determine how these data are processed. These entities would be classified as joint controllers, at least when engaging in secret sharing (as a joint purpose of data processing). These entities would need an agreement to establish that only the entity with physical access to patient data can access those data. If a request is made to a hospital or research institution that does not possess these data, the request must be referred to the entity that does.

Applying these principles to processing with privacy-enhancing technologies, for homomorphic encryption, there is no mathematical possibility of decrypting the data without the decryption key. This holds true when both the data are at rest or when the data are processed in the encrypted space via secure operations such as homomorphic addition or multiplication. Whether data processed as part of secure multiparty computation or multiparty homomorphic encryption remain personal data depends on whether entities have lawful access to personal data or decryption keys respectively. If entities can only access personal data they physically hold as part of a joint controller agreement, the data fragments exchanged during secret sharing via secure multiparty computation are not personal data. Likewise, under multiparty homomorphic encryption each individual entity only has access to a fragment of the decryption key, which can only be recombined with the approval of all other entities holding the remaining fragments. This argument is reinforced by Recital 57 of the GDPR [63], which provides controllers forbidden from identifying individuals are not required to collect identifying information to comply with the GDPR.

Therefore, we submit that both homomorphic encryption and secure multiparty computation, when used alone or together through multiparty homomorphic encryption can jointly compute health-related data while complying with the GDPR. These data remain anonymous even though entities processing data using multiparty homomorphic encryption are joint controllers. Furthermore, the use of advanced privacy-enhancing technologies should become a best standard for the processing of health-related data for three reasons. First, the Article 29 Working Party has recommended using encryption and anonymization techniques in concert to protect against orthogonal privacy risks and overcome the limits of individual techniques [55]. Second, the GDPR emphasizes the use of state-of-the-art techniques for guaranteeing the processing of sensitive data. Homomorphic encryption, secure multiparty computation, and multiparty homomorphic encryption are considered state-of-the-art technologies in that they carry a mathematical guarantee of privacy. Third, the Article 29 Working Party has held the data controller is responsible for demonstrating that the data have been and remain anonymized [55]. Further support from this argument comes from a case heard before the Swiss Federal Supreme Court [64]; in paragraph 5.12, the Federal Supreme Court endorsed a relative approach to anonymization, but also placed the onus on the data controller to establish anonymization. Switzerland is not a member of the European Union and does not have to comply with the GDPR. However, Switzerland's close proximity to the European Union means the Swiss Federal Act on Data Protection has been revised. These revisions ensure the continued free exchange of data between Switzerland and EU countries [65].

Therefore, we argue that multiparty homomorphic encryption involves processing anonymized data under EU data protection law. Although homomorphic encryption, secure multiparty computation, and multiparty homomorphic encryption do not obliviate the need for a joint controllership agreement, they lessen the administrative burden required for data sharing. Furthermore, they promote the use of standard processing agreements that can help ameliorate the impacts of national differences within and outside the European Union. Accordingly, we submit that multiparty homomorphic encryption, along with other forms of advanced privacy-enhancing technologies, should represent the standard for health data processing in low trust environments [66]. This processing can include performing computations on sensitive forms of data, such as providing genomic diagnoses without revealing the entire sequence for a patient [67]. Furthermore, the encrypted outputs of homomorphic encryption and secure multiparty computation are mathematically private, as they do not reveal any personal data [68]. Finally, the fact that multiparty homomorphic encryption involves processing anonymized data broadens the purposes for which health-related data can be used. For example, as part of a clinical trial protocol data might be collected from patients via their personal devices. These devices can either store these data locally or transmit them to a hospital. The data may then be queried in an anonymized form as part of a research project without needing to seek additional consent that would otherwise be required under data protection law for those data to be processed [69]. The ability to reuse data that are stored on a patient's personal device can also help support innovative forms of clinical trials, such as remote patient monitoring. The various states of data processed using novel privacy-enhancing technologies such as multiparty homomorphic encryption is displayed in Figure 2. Table 1 demonstrates the status of personal data at different stages of processing.

Figure 2. Comparison of the status of personal data under a distributed approach relying upon traditional privacy-enhancing technologies (eg, aggregation and pseudonymization) and a distributed approach relying on multiparty homomorphic encryption (eg, homomorphic encryption and secure multiparty computation).

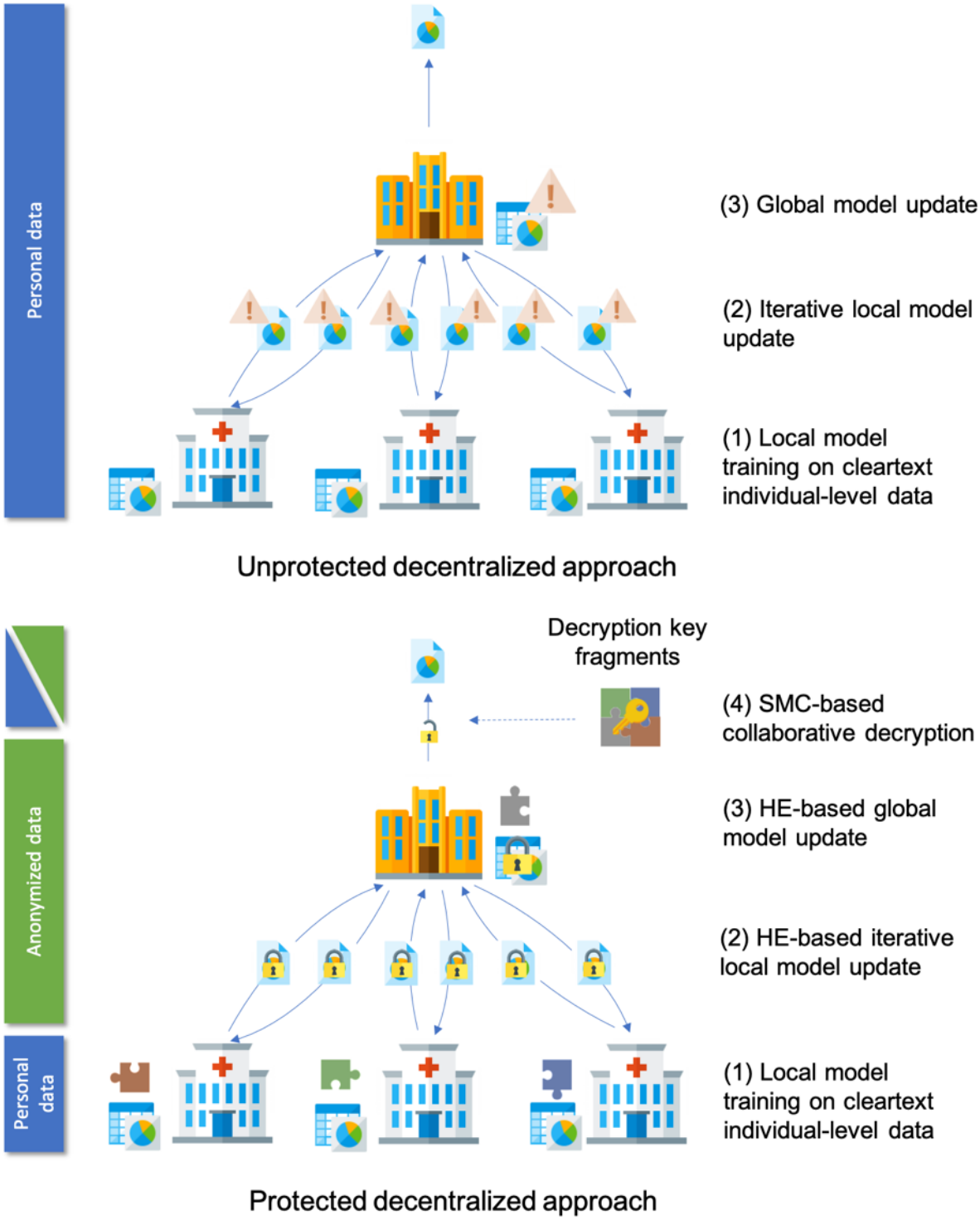


Table 1. Data status at different stages of processing.

Scenario	Description	Status of data based on the scenario
A	Hospital/research institution physically holds personal data	Personal data
B	Hospital/research institution has legal access to decryption key/personal data	Pseudonymized data
C	Hospital/research institution combine decryption keys/personal data to process data	Anonymized data
D	Third party (cloud service provider) carries out processing, hospitals share encryption keys jointly	Anonymized data

The lack of reliance upon custom contracts may encourage institutions to align their data formats to common international interoperability standards. In the next section, we turn to address the standardization of these advanced privacy-enhancing technologies.

Regulatory Instruments to Encourage the Use of Novel Privacy-Enhancing Technologies

At present, regulatory instruments provide limited guidance on the different types of privacy-enhancing technologies required to process medical data in a privacy-conscious fashion. However, the techniques described in this paper may represent a future best standard for processing medical data for clinical or research purposes. Because of the novelty of both technologies, the standardization of homomorphic encryption and secure multiparty computation is ongoing, with the first community standard released in 2018 [70].

Furthermore, there are numerous documents published by data protection agencies that can aid the development of such guidelines. For example, the *Commission Nationale de l'Informatique et des Libertés* (*French Data Protection Agency*) published a set of guidelines following the passage of the GDPR on how to secure personal data. This document provides recommendations on when encryption should be used, including for data transfer and storage [71]. Likewise, the *Agencia Española de Protección de Datos* (*Spanish Data Protection Agency*) has already recommended using homomorphic encryption as a mechanism for achieving data privacy by design pursuant to Article 25 of the GDPR [72].

Nevertheless, any standards will need to be continually updated to respond to new technological changes. For example, one of the most significant drawbacks of fully homomorphic encryption is the complexity of computation. This computational complexity makes it hard to predict running times, particularly for low-power devices such as wearables and smartphones. For the foreseeable future, this may limit the devices upon which fully homomorphic encryption can be used [73]. Therefore, specialized standards may need to be developed for using homomorphic encryption on low-power devices in a medical context. Specifically, these standards must be compliant with the legal requirements for access to and sharing of data by patients themselves, including the right to data portability as contained within Article 20 of the GDPR [**54**]. Although homomorphic encryption and secure multiparty computation offer privacy guarantees, there is still an orthogonal risk of reidentifying individuals from aggregate-level results that are eventually decrypted and can be exploited by inference attacks [19,21,27,74]. However, as mentioned earlier, the use of multiparty homomorphic encryption or secure multiparty computation enables the application of statistical obfuscation techniques for anonymizing aggregate-level results with a better privacy-utility trade-off than the traditional distributed approach, thus facilitating the implementation of end-to-end anonymized data workflows.

A final consideration relates to ethical issues that exist beyond whether homomorphic encryption, multiparty computation, and multiparty homomorphic encryption involve processing anonymized or personal data. First, the act of encrypting personal data constitutes further processing of those data under data protection law. Therefore, health care and research institutions must seek informed consent from patients or research participants [50]. Institutions must consider how to explain these technologies in a manner that is understandable and enables the patient to exercise their rights under data protection law. Second, the institution that holds the data must have procedures in place that govern who can access data encrypted using advanced privacy-enhancing technologies. Institutions should also determine which internal entity is responsible for governing access requests. These entities can include ethics review committees or data access committees [2].

Conclusion

Medical data sharing is essential for modern clinical practice and medical research. However, traditional privacy-preserving technologies based on data perturbation, along with centralized and decentralized data-sharing models, carry inherent privacy risks and may have high impact on data utility. These shortcomings mean that research and health care institutions combine these traditional privacy-preserving technologies with contractual mechanisms to govern data sharing and comply with data protection laws. These contractual mechanisms are context-dependent and require trusted environments between research and health care institutions. Although federated learning models can help alleviate these risks as only aggregate-level data are shared across institutions, there are still orthogonal risks to privacy from indirect reidentification of patients from partial results [66]. Furthermore, changes in case law (such as the already mentioned recent invalidation of the US-EU Privacy Shield [3]) can undermine data sharing with research partners outside the European Union. In this paper, we demonstrated how these privacy risks can be addressed through using multiparty homomorphic encryption, an efficient combination of homomorphic encryption and secure multiparty computation. In particular, we demonstrated how homomorphic encryption and secure multiparty computation can be used to compute accurate federated analytics without needing to transfer personal

data. Combining these technologies (multiparty homomorphic encryption) for medical data sharing can improve the performance overheads of privacy enhancing technology while reducing the risk of GDPR noncompliance. Furthermore, personal data do not leave the host institution where they are stored when processed using multiparty homomorphic encryption. Therefore, the lack of personal data transfer with multiparty homomorphic encryption will encourage increased data sharing and standardization between institutions. Data protection agencies, as well as health care and research institutions, should promote multiparty homomorphic encryption and other advanced privacy-enhancing technologies for their use to become widespread for clinical and research data sharing.

Acknowledgments

We are indebted to Dan Bogdanov, Brad Malin, Sylvain Métille, and Pierre Hutter for their invaluable feedback on earlier versions of this manuscript. This work was partially funded by the Personalized Health and Related Technologies Program (grant 2017-201; project: Data Protection and Personalized Health) supported by the Council of the Swiss Federal Institutes of Technology.

Conflicts of Interest

None declared.

References

1. Dove ES, Knoppers BM, Zawati MH. An ethics safe harbor for international genomics research? Genome Med 2013;5(11):99 [FREE Full text] [doi: 10.1186/gm503] [Medline: 24267880]
2. Phillips M, Molnár-Gábor F, Korbel J, Thorogood A, Joly Y, Chalmers D, et al. Genomics: data sharing needs an international code of conduct. Nature 2020 Feb 05;578(7798):31-33 [FREE Full text] [doi: 10.1038/d41586-020-00082-9]
3. Case C-311/2018 judgment of the court (grand chamber) of 16 July 2020. Data Protection Commissioner v Facebook Ireland Limited and Maximillian Schrems. request for a preliminary ruling from the high court (Ireland). reference for a preliminary ruling — protection of individuals with regard to the processing of personal data — charter of fundamental rights of the European Union — articles 7, 8 and 47 — regulation (EU) 2016/679 — article 2(2) — scope — transfers of personal data to third countries for commercial purposes — article 45 — commission adequacy decision — article 46 — transfers subject to appropriate safeguards — article 58 — powers of the supervisory authorities — processing of the data transferred by the public authorities of a third country for national security purposes — assessment of the adequacy of the level of protection in the third country — decision 2010/87/EU — protective standard clauses on the transfer of personal data to third countries — suitable safeguards provided by the data controller — validity — implementing decision (EU) 2016/1250 — adequacy of the protection provided by the EU-US privacy shield — validity — complaint by a natural person whose data was transferred from the European Union to the United States). Eur-Lex. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A62018CJ0311 [accessed 2021-02-10]
4. Ledford H, Van Noorden R. High-profile coronavirus retractions raise concerns about data oversight. Nature 2020 Jun 5;582(7811):160-160 [FREE Full text] [doi: 10.1038/d41586-020-01695-w] [Medline: 32504025]
5. Chen J. How the best-laid plans go awry: the (unsolved) issues of applicable law in the General Data Protection Regulation. Int Data Priv Law 2016 Dec 10;6(4):310-323 [FREE Full text] [doi: 10.1093/idpl/ipw020]
6. Ienca M, Scheibner J, Ferretti A, Gille F, Amann J, Sleigh J, et al. How the General Data Protection Regulation changes the rules for scientific research. ETH Zurich. 2019 Jul. URL: https://www.research-collection.ethz.ch/handle/20.500.11850/391622 [accessed 2021-02-12]
7. Murtagh MJ, Demir I, Jenkings KN, Wallace SE, Murtagh B, Boniol M, et al. Securing the data economy: translating privacy and enacting security in the development of DataSHIELD. Public Health Genomics 2012;15(5):243-253 [FREE Full text] [doi: 10.1159/000336673] [Medline: 22722688]
8. Bovenberg J, Peloquin D, Bierer B, Barnes M, Knoppers BM. How to fix the GDPR's frustration of global biomedical research. Science 2020 Oct 02;370(6512):40-42 [FREE Full text] [doi: 10.1126/science.abd2499] [Medline: 33004505]
9. Kierkegaard P. Electronic health record: wiring Europe's healthcare. Comput Law Secur Rev 2011 Sep 16;27(5):503-515 [FREE Full text] [doi: 10.1016/j.clsr.2011.07.013]
10. El Emam K, Jonker E, Arbuckle L, Malin B. A systematic review of re-identification attacks on health data. PLoS One 2011;6(12):e28071 [FREE Full text] [doi: 10.1371/journal.pone.0028071] [Medline: 22164229]
11. Lin Z, Owen AB, Altman RB. Genomic research and human subject privacy. Science 2004 Jul 09;305(5681):183 [FREE Full text] [doi: 10.1126/science.1095019] [Medline: 15247459]
12. Gymrek M, McGuire AL, Golan D, Halperin E, Erlich Y. Identifying personal genomes by surname inference. Science 2013 Jan 18;339(6117):321-324 [FREE Full text] [doi: 10.1126/science.1229566] [Medline: 23329047]
13. Lippert C, Sabatini R, Maher MC, Kang EY, Lee S, Arikan O, et al. Identification of individuals by trait prediction using whole-genome sequencing data. Proc Natl Acad Sci U S A 2017 Sep 19;114(38):10166-10171 [FREE Full text] [doi: 10.1073/pnas.1711125114] [Medline: 28874526]

14. El Emam K, Buckeridge D, Tamblyn R, Neisa A, Jonker E, Verma A. The re-identification risk of Canadians from longitudinal demographics. BMC Med Inform Decis Mak 2011 Jun 22;11(1):1-12 [FREE Full text] [doi: 10.1186/1472-6947-11-46] [Medline: 21696636]
15. Loukides G, Denny JC, Malin B. The disclosure of diagnosis codes can breach research participants' privacy. J Am Med Inform Assoc 2010 May 01;17(3):322-327 [FREE Full text] [doi: 10.1136/jamia.2009.002725] [Medline: 20442151]
16. Atreya R, Smith J, McCoy A, Malin B, Miller R. Reducing patient re-identification risk for laboratory results within research datasets. J Am Med Inform Assoc Oxford Academic Jan 1 2013;20(1):95-101 [FREE Full text] [doi: 10.1136/amiajnl-2012-001026] [Medline: 22822040]
17. Austin L, Lie D. Safe Sharing Sites. N Y Univ Law Rev 2019;94(4):581-623 [FREE Full text]
18. Bot BM, Wilbanks JT, Mangravite LM. Assessing the consequences of decentralizing biomedical research. Big Data Soc 2019 Jun 11;6(1):1-6 [FREE Full text] [doi: 10.1177/2053951719853858]
19. Homer N, Szelinger S, Redman M, Duggan D, Tembe W, Muehling J, et al. Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. PLoS Genet 2008 Aug 29;4(8):e1000167 [FREE Full text] [doi: 10.1371/journal.pgen.1000167] [Medline: 18769715]
20. Sankararaman S, Obozinski G, Jordan MI, Halperin E. Genomic privacy and limits of individual detection in a pool. Nat Genet 2009 Aug 23;41(9):965-967 [FREE Full text] [doi: 10.1038/ng.436] [Medline: 19701190]
21. Shringarpure SS, Bustamante CD. Privacy risks from genomic data-sharing beacons. Am J Hum Genet 2015 Nov 05;97(5):631-646 [FREE Full text] [doi: 10.1016/j.ajhg.2015.09.010] [Medline: 26522470]
22. Raisaro JL, Tramèr F, Ji Z, Bu D, Zhao Y, Carey K, et al. Addressing beacon re-identification attacks: quantification and mitigation of privacy risks. J Am Med Inform Assoc 2017 Jul 01;24(4):799-805 [FREE Full text] [doi: 10.1093/jamia/ocw167] [Medline: 28339683]
23. Liu Y, Wan Z, Xia W, Kantarcioglu M, Vorobeychik Y, Clayton EW, et al. Detecting the presence of an individual in phenotypic summary data. AMIA Annu Symp Proc 2018;2018:760-769 [FREE Full text] [Medline: 30815118]
24. Dwork C, McSherry F, Nissim K, Smith A. Calibrating noise to sensitivity in private data analysis. In: Halevi S, Rabin T, editors. Theory of Cryptography. Berlin, Heidelberg: Springer; 2006:265-284.
25. Wood A, Altman M, Bembenek A, Bun M, Gaboardi M, Honaker J, et al. Differential privacy: a primer for a non-technical audience. Vanderbilt J Entertain Technol Law 2019;21(1):209-276 [FREE Full text]
26. Jayaraman B, Evans D. Evaluating differentially private machine learning in practice. 2019 Presented at: 28th USENIX Security Symposium; 14-16 August 2019; Santa Clara p. 1895-1912 URL: https://www.usenix.org/conference/usenixsecurity19/presentation/jayaraman
27. Nasirigerdeh R, Torkzadehmahani R, Matschinske J, Frisch T, List M, Späth J, et al. sPLINK: a federated, privacy-preserving tool as a robust alternative to meta-analysis in genome-wide association studies. BioRxiv. Preprint posted online on June 6, 2020. [FREE Full text] [doi: 10.1101/2020.06.05.136382]
28. Kairouz P, McMahan H, Avent B, Bellet A, Bennis M, Bhagoji A, et al. Advances and open problems in federated learning. ArXiv. Preprint posted online on December 10, 2019. [FREE Full text] [doi: 10.1561/2200000083]
29. Gaye A, Marcon Y, Isaeva J, LaFlamme P, Turner A, Jones EM, et al. DataSHIELD: taking the analysis to the data, not the data to the analysis. Int J Epidemiol 2014 Dec;43(6):1929-1944 [FREE Full text] [doi: 10.1093/ije/dyu188] [Medline: 25261970]
30. Melie-Garcia L, Draganski B, Ashburner J, Kherif F. Multiple linear regression: bayesian inference for distributed and big data in the medical informatics platform of the human brain project. BioRxiv. Preprint posted online on January 5, 2018. [FREE Full text] [doi: 10.1101/242883]
31. Rigaki M, Garcia S. A survey of privacy attacks in machine learning. ArXiv. Preprint posted online on July 15, 2020. [FREE Full text]
32. Carlini N, Liu C, Erlingsson Ú, Kos J, Song D. The secret sharer: evaluating and testing unintended memorization in neural networks. 2019 Presented at: 28th USENIX Security Symposium; 14-16 August; Santa Clara URL: https://www.usenix.org/conference/usenixsecurity19/presentation/carlini
33. Hitaj B, Ateniese G, Perez-Cruz F. Deep models under the GAN: information leakage from collaborative deep learning. In: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security. 2017 Presented at: ACM Conference on Computer and Communications Security (CCS) 2017; October 30-November 3; Dallas p. 603-618. [doi: 10.1145/3133956.3134012]
34. Nasr M, Shokri R, Houmansadr A. Comprehensive privacy analysis of deep learning: passive and active white-box inference attacks against centralized and federated learning. 2019 Presented at: 2019 IEEE Symposium on Security and Privacy; May 19-23; San Francisco p. 739-753 URL: https://ieeexplore.ieee.org/abstract/document/8835245 [doi: 10.1109/sp.2019.00065]
35. Melis L, Song C, De CE, Shmatikov V. Exploiting unintended feature leakage in collaborative learning. 2019 Presented at: 2019 IEEE Symposium on Security and Privacy; May 19-23; San Francisco p. 691-706 URL: https://ieeexplore.ieee.org/abstract/document/8835269 [doi: 10.1109/sp.2019.00029]
36. Wallace SE, Gaye A, Shoush O, Burton PR. Protecting personal data in epidemiological research: DataSHIELD and UK law. Public Health Genomics 2014;17(3):149-157 [FREE Full text] [doi: 10.1159/000360255] [Medline: 24685519]

37. Gentry C. Computing arbitrary functions of encrypted data. Commun ACM 2010 Mar;53(3):97-105 [FREE Full text] [doi: 10.1145/1666420.1666444]
38. Yao A. Protocols for secure computations. 1982 Presented at: 23rd Annual Symposium on Foundations of Computer Science; Nov 3-5; Chicago p. 160-164 URL: https://ieeexplore.ieee.org/abstract/document/4568388 [doi: 10.1109/sfcs.1982.38]
39. Bogdanov D, Laur S, Willemson J. Sharemind: a framework for fast privacy-preserving computations. In: Jajodia S, Lopez J, editors. Computer Security European Symposium on Research in Computer Security. Berlin: Springer; 2008:192-206.
40. Damgård I, Pastro V, Smart N, Zakarias S. Multiparty computation from somewhat homomorphic encryption. In: Safavi-Naini R, Canetti R, editors. Advances in Cryptology CRYPTO Lecture Notes in Computer Science vol 7412. Berlin: Springer; 2012:643-662.
41. Damgård I, Keller M, Larraia E, Pastro V, Scholl P, Smart N. Practical covertly secure MPC for dishonest majority - or: breaking the SPDZ limits. In: Crampton J, Jajodia S, Mayes K, editors. Computer Security European Symposium on Research in Computer Security. Berlin: Springer; 2013:1-18.
42. Keller M, Pastro V, Rotaru D. Overdrive: making SPDZ great again. In: Nielsen JB, Rijmen V, editors. Advances in Cryptology EUROCRYPT Lecture Notes in Computer Science vol 10822. Cham: Springer International Publishing; 2018:158-189.
43. Mouchet C, Troncoso-Pastoriza J, Bossuat JP, Hubaux JP. Multiparty homomorphic encryption: from theory to practice. Cryptology ePrint Archive. 2020 Dec 22. URL: https://eprint.iacr.org/2020/304 [accessed 2021-02-12]
44. Cramer R, Damgård I, Nielsen J. Multiparty computation from threshold homomorphic encryption. In: Pfitzmann B, editor. Advances in Cryptology EUROCRYPT 2001 Lecture Notes in Computer Science vol 2045. Berlin: Springer; 2001:280-300.
45. Asharov G, Jain A, López-Alt A, Tromer E, Vaikuntanathan V, Wichs D. Multiparty computation with low communication, computation and interaction via threshold FHE. In: Pointcheval D, Johansson T, editors. Advances in Cryptology EUROCRYPT 2012 Lecture Notes in Computer Science vol 7237. Berlin: Springer; 2012:483-501.
46. Boneh D, Gennaro R, Goldfeder S, Jain A, Kim S, Rasmussen P, et al. Threshold cryptosystems from threshold fully homomorphic Eencryption. In: Shacham H, Boldyreva A, editors. Advances in Cryptology CRYPTO Lecture Notes in Computer Science vol 10991. Cham: Springer International Publishing; 2018:565-596.
47. Zheng W, Popa R, Gonzalez J, Stoica I. Helen: maliciously Ssecure coopetitive learning for linear models. 2019 Presented at: 2019 IEEE Symposium on Security and Privacy; May 19-23; San Francisco p. 724-738 URL: https://ieeexplore.ieee.org/abstract/document/8835215 [doi: 10.1109/sp.2019.00045]
48. Raisaro JL, Troncoso-Pastoriza JR, Misbach M, Sousa JS, Pradervand S, Missiaglia E, et al. MedCo: enabling secure and privacy-preserving exploration of distributed clinical and genomic data. IEEE/ACM Trans Comput Biol and Bioinf 2019 Jul 1;16(4):1328-1341. [doi: 10.1109/tcbb.2018.2854776]
49. Sav S, Pyrgelis A, Troncoso-Pastoriza J, Froelicher D, Bossuat JP, Sousa J, et al. POSEIDON: privacy-preserving federated neural network learning. arXiv. Preprint posted online on January 8, 2021. [FREE Full text]
50. Spindler G, Schmechel P. Personal data and encryption in the European General Data Protection Regulation. J Intellect Prop Inf Technol Electron Commer Law 2016;7(2):163 [FREE Full text]
51. Finck M, Pallas F. They who must not be identified - distinguishing personal from non-personal data under the GDPR. Int Data Priv Law 2020 Feb 1;10(1):11-36 [FREE Full text] [doi: 10.1093/idpl/ipz026]
52. Guidance on anonymisation and pseudonymisation. Coimisiún Chosaint Sonraí Data Protection Commission. 2019. URL: https://www.dataprotection.ie/sites/default/files/uploads/2019-06/190614%20Anonymisation%20and%20Pseudonymisation.pdf [accessed 2021-02-12]
53. Orientierungshilfe zum Gesundheitsdatenschutz. Bundesministerium für Wirtschaft und Energie. 2018 Nov. URL: https://www.bmwi.de/Redaktion/DE/Downloads/M-O/orientierungshilfe-gesundheitsdatenschutz.pdf?__blob=publicationFile&v=16 [accessed 2021-02-12]
54. Article 29 Data Protection Working Party. Opinion 04/2007 on the concept of personal data. European Commission. 2007. URL: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf [accessed 2021-02-18]
55. Article 29 Data Protection Working Party. Opinion 05/2014 on anonymisation techniques. European Commission. 2014 Apr. URL: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf [accessed 2021-02-18]
56. Esayas S. The role of anonymisation and pseudonymisation under the EU data privacy rules: beyond the 'all or nothing' approach. Eur J Law Technol Vol 2015;6(2):1-28 [FREE Full text]
57. Nissim K, Bembenek A, Wood A, Bun M, Gaboardi M, Gasser U, et al. Bridging the gap between computer science and legal approaches to privacy. Harv J Law Technol 2018;31(2):2017-2780 [FREE Full text]
58. Case C-582/14 judgement of the court (second chamber) of 19 October 2016 (request for a preliminary ruling from the Bundesgerichtshof – Germany) – Patrick Breyer v Bundesrepublik Deutschland. Eur-Lex. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A62014CA0582&qid=1614053861638 [accessed 2021-02-10]
59. Mondschein C, Monda C. The EU's General Data Protection Regulation (GDPR) in a research context. In: Kubben P, Dumontier M, Dekker A, editors. Fundamentals of Clinical Data Science. Cham: Springer; 2019:55-71.

60. Case C-210/16 judgment of the court (grand chamber) of 5 June 2018 (request for a preliminary ruling from the Bundesverwaltungsgericht — Germany) — Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein v Wirtschaftsakademie Schleswig-Holstein GmbH. Eur-Lex. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A62016CA0210&qid=1613633585608 [accessed 2021-02-18]
61. Case C-25/17 judgment of the court (grand chamber) of 10 July 2018 (request for a preliminary ruling from the korkein hallinto-oikeus — Finland) — proceedings brought by Tietosuojavaltuutettu. Eur-Lex. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A62017CA0025&qid=1613633621924 [accessed 2021-02-18]
62. Case C-40/17 judgement of the court (grand chamber) of 29 July 2019 (request for a preliminary ruling from the Oberlandesgericht Düsseldorf — Germany) – Fashion ID GmbH & Co. KG v Verbraucherzentrale NRW eV. Eur-Lex. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A62017CA0040&qid=1613634205133 [accessed 2021-02-18]
63. Regulation (EU) 2016/679 of the European parliament and of the council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/EC (General Data Protection Regulation). Eur-Lex. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679&qid=1613633727166 [accessed 2021-02-18]
64. 4A_365/2017. Tribunal fédéral. URL: https://www.bger.ch/ext/eurospider/live/fr/php/aza/http/index.php?lang=fr&type=show_document&highlight_docid=aza://26-02-2018-4A_365-2017 [accessed 2021-02-18]
65. Staiger DN. Swiss data protection law. In: Moura Vicente D, de Vasconcelos Casimiro S, editors. Data Protection in the Internet Ius Comparatum - Global Studies in Comparative Law vol 38. Cham: Springer International Publishing; 2020:397-408.
66. Kaissis GA, Makowski MR, Rückert D, Braren RF. Secure, privacy-preserving and federated machine learning in medical imaging. Nat Mach Intell 2020 Jun 8;2(6):305-311 [FREE Full text] [doi: 10.1038/s42256-020-0186-1]
67. Jagadeesh KA, Wu DJ, Birgmeier JA, Boneh D, Bejerano G. Deriving genomic diagnoses without revealing patient genomes. Science 2017 Aug 18;357(6352):692-695 [FREE Full text] [doi: 10.1126/science.aam9710] [Medline: 28818945]
68. Munn L, Hristova T, Magee L. Clouded data: privacy and the promise of encryption. Big Data Soc 2019 Jun 24;6(1):1-16 [FREE Full text] [doi: 10.1177/2053951719848781]
69. Grishin D, Raisaro J, Troncoso-Pastoriza J, Obbad K, Quinn K, Misbach M, et al. Citizen-centered, auditable, and privacy-preserving population genomics. bioRxiv. Preprint posted online on October 10, 2019. [FREE Full text] [doi: 10.1101/799999]
70. Albrecht M, Chase M, Chen H, Ding J, Goldwasser S, Gorbunov S, et al. Homomorphic encryption standard internet report no. 939. Cryptology ePrint Archive. 2019 Aug 17. URL: https://eprint.iacr.org/2019/939 [accessed 2021-02-12]
71. A new guide regarding security of personal data. Commission Nationale de l'Informatique et des Libertés. 2018 Apr 4. URL: https://www.cnil.fr/en/new-guide-regarding-security-personal-data [accessed 2021-02-12]
72. Encryption and privacy III: homomorphic encryption. Agencia Española de Protección de Datos. 2020. URL: https://www.aepd.es/en/prensa-y-comunicacion/blog/encryption-privacy-iii-homomorphic-encryption [accessed 2021-02-12]
73. Laine K. Homomorphic encryption. In: Jiang X, Tang H, editors. Responsible Genomic Data Sharing Challenges and Approaches. Cambridge: Academic Press; 2019:97-122.
74. Vinterbo SA, Sarwate AD, Boxwala AA. Protecting count queries in study design. J Am Med Inform Assoc 2012;19(5):750-757 [FREE Full text] [doi: 10.1136/amiajnl-2011-000459] [Medline: 22511018]

Abbreviations

CJEU: Court of Justice of the European Union
COVID-19: coronavirus disease 2019
GDPR: (European Union) General Data Protection Regulation
HIV: human immunodeficiency virus

Edited by G Eysenbach; submitted 19.10.20; peer-reviewed by B Knoppers, S Palmdorf, R Hendricks-Sturrup; comments to author 07.12.20; revised version received 06.01.21; accepted 16.01.21; published 25.02.21

Please cite as:
Scheibner J, Raisaro JL, Troncoso-Pastoriza JR, Ienca M, Fellay J, Vayena E, Hubaux JP
Revolutionizing Medical Data Sharing Using Advanced Privacy-Enhancing Technologies: Technical, Legal, and Ethical Synthesis
J Med Internet Res 2021;23(2):e25120
URL: https://www.jmir.org/2021/2/e25120
doi: 10.2196/25120
PMID: 33629963