Journal of Law and the Biosciences, 1–30
doi:10.1093/jlb/lsaa010
Advance Access Publication 6 May 2020
Original Article

Data protection and ethics requirements for multisite research with health data: a comparative examination of legislative governance frameworks and the role of data protection technologies†

James Scheibner[1], Marcello Ienca[1], Sotiria Kechagia[2], Juan Ramon Troncoso-Pastoriza[3], Jean Louis Raisaro[4], Jean-Pierre Hubaux[3], Jacques Fellay[4,5,6] and Effy Vayena[1,*]

[1]Health Ethics and Policy Laboratory, Department of Health Sciences and Technology, ETH Zürich, Zürich, Switzerland
[2]Centre for Digital Trust, School of Computer and Communication Sciences, EPFL, Lausanne, Switzerland
[3]Laboratory for Data Security, School of Computer and Communication Sciences, EPFL, Lausanne, Switzerland
[4]Unité de Médecine de Précision, CHUV, Lausanne, Switzerland
[5]School of Life Sciences, EPFL, Lausanne, Switzerland
[6]Host-Pathogen Genomics Laboratory, Swiss Institute of Bioinformatics, Lausanne, Switzerland
*Corresponding author. E-mail: effy.vayena@hest.ethz.ch

ABSTRACT

Personalised medicine can improve both public and individual health by providing targeted preventative and therapeutic healthcare. However, patient health data must be shared between institutions and across jurisdictions for the benefits of personalised medicine to be realised. Whilst data protection, privacy, and research ethics laws protect patient confidentiality and safety they also may impede multisite research, particularly across jurisdictions. Accordingly, we compare the concept of data accessibility in data protection and research ethics laws across seven jurisdictions. These jurisdictions include Switzerland, Italy, Spain, the United Kingdom (which have implemented the General Data Protection Regulation), the

† This work has been written and funded as part of the Data Protection for Personalised Health (DPPH) project. This project was supported by the grant #2017-201 of the Strategic Focal Area 'Personalized Health and Related Technologies (PHRT)' of the ETH Domain.

United States, Canada, and Australia. Our paper identifies the requirements for consent, the standards for anonymisation or pseudonymisation, and adequacy of protection between jurisdictions as barriers for sharing. We also identify differences between the European Union and other jurisdictions as a significant barrier for data accessibility in cross jurisdictional multisite research. Our paper concludes by considering solutions to overcome these legislative differences. These solutions include data transfer agreements and organisational collaborations designed to 'front load' the process of ethics approval, so that subsequent research protocols are standardised. We also allude to technical solutions, such as distributed computing, secure multiparty computation and homomorphic encryption.

KEYWORDS: Advanced cryptography, Biomedical data, Data protection, Data sharing, Multisite research, Personalised healthcare

INTRODUCTION

Personalized medicine has the potential to improve healthcare by providing targeted and more effective preventive and therapeutic strategies for individuals. For these benefits to be fully realized, doctors, healthcare providers, researchers, and government agencies must maximize sharing and linkage of patient and administrative data for multisite research.[1] However, data sharing for medical data is subject to privacy and security risks, which may have two contrasting yet equally detrimental effects. On the one hand, privacy and security concerns can undermine public confidence and participation in research, degrading the overall value flowing from data linkages.[2] On the other hand, overreacting to privacy and data security concerns could undermine attempts to share data for research, clinical, and public health purposes.[3] Furthermore, biomedical research is increasingly dependent on multisite research, requiring the transfer of different forms of data between different jurisdictions and accelerating the risks described above.[4]

In this paper, we address two research questions. First, we wish to identify the regulatory obstacles, both ethical and legal, associated with multisite research with health-related data, including data sharing and linkage. Second, we wish to delineate possible organizational and technical solutions to overcome these obstacles, with a special focus on advanced privacy-enhancing technologies.[5] This tight link between ethical, legal, and technical requirements is grounded in the observation that any technical solutions

1 James H. Boyd et al., *Data Linkage Infrastructure for Cross-Jurisdictional Health-Related Research in Australia*, 12 BMC Health Serv. Res. 480, 488 (2012).

2 Pam Carter, Graeme T. Laurie & Mary Dixon-Woods, *The Social Licence for Research: Why Care.Data Ran into Trouble*, 41 J. Med. Ethics 404–409, 406–7 (2015).

3 OECD, *Health Data Governance: Privacy, Monitoring and Research* (2015), 76.

4 Jennifer R. Popovic, *Distributed Data Networks: A Blueprint for Big Data Sharing and Healthcare Analytics*, 1387 Ann. N. Y. Acad. Sci. 105–111 (2017); Rolf H. Weber, *Transborder Data Transfers: Concepts, Regulatory Approaches and New Legislative Initiatives*, 3 Int. Data Priv. Law 117–130 (2013).

5 Our approach is based on the recent Swiss Personalized Health Network (SPHN) initiative. SPHN aims at building an infrastructure an ethical framework to enable the five university hospitals to share medical data for reuse in research. Within this framework, the Data Protection in Personalized Health (DPPH) project explores the synergies between advanced data protection technologies and the existing regulatory and ethical frameworks.

must be reinforced by an ethically and legally compliant governance framework.[6] Accordingly, our paper is organized into three sections. The first section examines existing international frameworks governing research with health data. These include research ethics agreements and declarations, as well as international and supranational agreements governing data protection. Outside of these legal agreements, the first section also considers reports and governance documents that have attempted to define the scope of health data research. These include documents published by research consortiums as well as other entities such as the Organization for Economic Cooperation and Development (OECD). The second section then compares the national data protection and research ethics laws in seven OECD jurisdictions using five criteria for data accessibility published by the OECD.[7] These criteria are used to assess four features of each national regime. The third section identifies the regulatory obstacles for data accessibility on a cross jurisdictional basis. It concludes by illustrating how organizational and technical solutions can be used to overcome these regulatory obstacles.

INTERNATIONAL FRAMEWORKS GOVERNING RESEARCH WITH HEALTH DATA

By examining international research guidelines, a number of principles for research involving health-related data emerge. First, all biomedical and health-related research requires the free and informed consent of research participants.[8] Second, researchers and research organizations must guarantee and be held accountable for the confidentiality of research participant data, as well as ensuring risk mitigation.[9] Third, research participants must be compensated for and allowed to benefit from research involvement.[10] Finally, efforts must be made to ensure the accessibility and maximum reuse of data collected from research.[11] Although research ethics committees play an important role in protecting research participants, they must also balance these principles against each other, despite potential conflicts. For example, the Declaration of Taipei limits the use of biobanking data to where specific consent has been obtained, which may prevent general consent for longitudinal research.[12] These conflicting principles can be resolved by ethics committees requiring researchers to implement a specific study protocol to minimize the risk to participants.

6 Alessandro Blasimme & Effy Vayena, *Becoming Partners, Retaining Autonomy: Ethical Considerations on the Development of Precision Medicine*, 17 BMC Med. Ethics 67, 69 (2016).

7 OECD, *supra* note 11, at 66–68.

8 United Nations Educational Scientific and Cultural Organisation (UNESCO) Universal Declaration on the Human Genome, Art. 5(b) [hereafter UDHG]; UNESCO Universal Declaration on Bioethics and Human Rights, Art. 6, 7 [hereafter UDBHR]; Council for International Organisations of Medical Sciences (CIOMS) International Ethical Guidelines, Guidelines 1, 9, 10, 11, 12; World Medical Association (WMA) Declaration of Taipei, 1, 11, 12, 13, 14, 15, 16.

9 UDHG, Art. 5, 8, 17; UDBHR, Art. 18, 19, 20, 21; International Ethical Guidelines, Guidelines 23, 24; Declaration of Taipei, 19.

10 UDHG, Art. 5, 12, 19; UDBHR, Art. 2, 4, 15; International Ethical Guidelines, Guidelines 3, 4, 7, 8, 13, 14; Declaration of Taipei, 9, 12, 17.

11 UDHG, Art. 12, UDBHR, Art. 2, 14; International Ethical Guidelines, Guidelines 2, 22, 24; Declaration of Taipei, 12, 18s.

12 Evert-Ben van Veen, *Observational Health Research in Europe: Understanding the General Data Protection Regulation and Underlying Debate*, 104 Eur. J. Cancer 70–80, 76 (2018).

Nevertheless, there is significant conflict between protecting privacy and maximizing the use of research materials due to the technical difficulty of anonymization or removal of identifiers, so individuals cannot be re-identified. Ohm argues that true anonymization of personal data is functionally impossible and that releases of data should instead be contextual.[13] Multiple studies, since Ohm's study, have proven the technical ineffectiveness of standard de-identification and anonymization techniques both for omics[14] and clinical data.[15] These questions of confidentiality and security for biomedical research data dovetail into the issue of data protection. The 2013 OECD Privacy Guidelines establish a series of eight data protection principles that underpin data protection laws in OECD jurisdictions.[16] Furthermore, the 2015 OECD report on best practices for health data governance generated six data accessibility criteria.[17] Five of these are relevant to our paper:

(i) Identifiable data are shared with other data custodian or government entities.
(ii) University and non-profit researchers may be approved access to de-identified data.
(iii) Healthcare providers may be approved access to de-identified data.
(iv) For-profit businesses may be approved access to de-identified data.
(v) Foreign government, university, or non-profit researchers may be approved access to de-identified data.

Moreover, the OECD 2015 report defines 'de-identified data' as 'data that cannot be used to identify an individual directly'.[18] The OECD distinguishes 'de-identified data' (used here synonymously with anonymized data) from 'pseudonymization', or techniques where identifying information is converted to meaningless values. Outside the OECD, the Global Alliance for Genomics and Health, a consortium of 580 research organizations, published their 'Framework for the Responsible Sharing of Genomic and Health Related Data'. This framework highlights the importance of the right to free scientific inquiry, including data accessibility.[19] This final principle mandates that

13 Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA Law Rev. 1701–1778, 1706, 1762 (2009).

14 Zhen Lin, Art B. Owen & Russ B. Altman, *Genomic Research and Human Subject Privacy*, 305 Science 183–183 (2004); Nils Homer et al., *Resolving Individuals Contributing Trace Amounts of DNA to Highly Complex Mixtures Using High-Density SNP Genotyping Microarrays*, 4 PLOS Genet. e1000167, 7–9 (2008); Melissa Gymrek et al., *Identifying Personal Genomes by Surname Inference*, 339 Science 321–324, 324 (2013); Christoph Lippert et al., *Identification of Individuals by Trait Prediction Using Whole-Genome Sequencing Data*, 114 Proc. Natl. Acad. Sci. 10166–10171, 10169–10170 (2017).

15 Khaled El Emam et al., *A Systematic Review of Re-Identification Attacks on Health Data*, 6 PLOS ONE e28071, 6–7 (2011).

16 Edward S. Dove & Mark Phillips, *Privacy Law, Data Sharing Policies, and Medical Data: A Comparative Perspective*, *in* Medical Data Privacy Handbook 639–678, 649 (Aris Gkoulalas-Divanis & Grigorios Loukides eds, 2015).

17 OECD, *supra* note 11, at 66–68.

18 *Id.* at 12.

19 Bartha Maria Knoppers, *Framework for Responsible Sharing of Genomic and Health-Related Data*, 8 HUGO J. 3, 4 (2014); Mahsa Shabani, Bartha Maria Knoppers & Pascal Borry, *From the Principles of Genomic Data Sharing to the Practices of Data Access Committees*, 7 EMBO Mol. Med. 507–509, 508 (2015).

researchers make data and results that widely accessible through publishing findings and digitally disseminating results while minimizing obstacles to data sharing.

Therefore, we use the criteria of 'data accessibility' for comparing four features of the regulatory regime in each jurisdiction. These four features are responsibility for processing, the data types available for processing, the consent or approval required for processing, and where health data can be transferred. We restricted our analysis to OECD jurisdictions, as all have modeled their privacy laws around the OECD Privacy Guidelines. We included Switzerland as part of our study, which features historically strong data protection laws with high degrees of data reuse for research.[20] We then extended our analysis to include three jurisdictions compliant with the European Union (EU) General Data Protection Regulation (GDPR): Italy, Spain, and the UK.[21] We chose these jurisdictions as they have implemented specific rules for the processing of biological and health-related data as permitted by the GDPR. We also included the USA and Canada as part of our analysis. We made this decision as these two jurisdictions have historically leaned toward industry self-regulation or more fragmented approaches as opposed to the rights-based approach in Europe.[22] Finally, we included Australia due to the high rates of adoption of electronic health record systems and research data linkage. An overview of the key terms that are considered as part of our analysis is contained in Table 1.

A COMPARISON OF NATIONAL DATA PROTECTION AND RESEARCH ETHICS LAWS

Responsibility for Processing

The *Federal Act on Data Protection* (*FADP*)[23] and the *Ordinance to the FADP* (*OFADP*)[24] govern personal data processing in Switzerland by private persons or federal bodies, in the private or public sector. The *Human Research Act* (*HRA*)[25] and the *Human Research Ordinance* (*HRO*)[26] also govern research involving human beings. Likewise, the GDPR applies to all personal data processing by EU-established controllers or processors.[27] Similarly, Canadian privacy legislation operates on multiple jurisdictional levels. Federal Canadian legislation applies to commercial personal information processing by private-sector organizations, particularly for interprovincial

20 Jillian Oderkirk, Elettra Ronchi & Niek Klazinga, *International Comparisons of Health System Performance Among OECD Countries: Opportunities and Data Privacy Protection Challenges*, 112 Health Policy 9–18, 12 (2013).

21 Although the UK formally left the EU in January 2020, it has still implemented the GDPR into national law via the *Data Protection Act 2018* (UK).

22 Dove and Phillips, *supra* note 24, at 658.

23 Bundesgesetz über den Datenschutz (DSG) [Federal Act on Data Protection, FADP] June 19, 1992, SR 235.1 (Switzerland).

24 Verordnung zum Bundesgesetz über den Datenschutz (VDSG) [Ordinance of June 14, 1993 to the Federal Act on Data Protection, OFADP] June 14, 1993 (Switzerland).

25 Bundesgesetz über die Forschung am Menschen (HFG) [Federal Act on Research Involving Human Beings, HRA] Sept. 30, 2011, SR 810.30 (Switzerland).

26 Verordnung über die Humanforschung mit Ausnahme der klinischen Versuche (HFV) [Ordinance of Sept. 20, 2013 on Human Research with the Exception of Clinical Trials, HRO] Sept. 20, 2013 (Switzerland).

27 EU Regulation 2016/679 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) (Apr. 27, 2016) [GDPR], Art. 3.

Table 1. A summary of commonly used terms in national and supranational data protection laws

Data controller	The natural or legal person who decides how data should be used or processed
Data processor	The natural or legal responsible for processing the data at the request of the data controller
Data custodian	The natural or legal person responsible for handling or holding the data
Data subject	The natural person whose data is collected, stored or otherwise processed
Personal data/information	Any information relating, directly or indirectly, to an identified or identifiable person. Includes information inferred about a person
PHI	Any information about health status, provision of healthcare, or payment of healthcare. Protected under HIPAA in the USA only
Anonymization	The process of rendering data so that it cannot be used to identify an individual directly or indirectly. Not regulated by data protection legislation
Pseudonymization	The process of removing identifiers from data so that data cannot be re-identified without these identifiers. Regulated by data protection legislation
De-identification	The process by which identifiers are removed from the health information, which mitigates privacy risks to individuals and thereby supports the secondary use of data for comparative effectiveness studies, policy assessment, life sciences research, and other endeavors
Specific consent	Explicit consent to use data for a particular purpose or research project. Data obtained under this consent cannot be reused for another purpose
General consent	Explicit consent to use data in a particular field. Data obtained under this consent can be reused for further purposes in the same field
Exchange	To physically move data from one organization to another
Transfer	To physically move personal data from one jurisdiction to another

Note that these terms reflect those used in the OECD, *Health Data Governance: Privacy, Monitoring and Research* (2015) report, as well as other legislation (such as HIPAA).

transfers and where provincial legislation is not substantially similar. In addition, the *Privacy Act 1985* governs data processing by governments and public-sector organizations.[28] Outside these acts, provincial data protection legislation in Canada displaces or supplants the federal legislation on a sectorial basis. All Canadian provinces have separate legislation governing the processing of health-related personal data.

28 *Personal Information Protection and Electronic Documents Act 2000* (PIPEDA) and the *Privacy Act 1985*.

In Alberta, New Brunswick, Newfoundland, Nova Scotia, and Ontario, 'health information custodians', or entities handling health data, are also held responsible for data processing. These custodians include entities that process health data for commercial and non-commercial purposes, therefore supplanting federal legislation.[29] In contrast, in the USA, the *Health Insurance Portability and Accountability Act 1996* (*HIPAA*), the *Privacy Rule*, and the *Security Rule* extend responsibility to three types of covered entities. These entities include healthcare providers, health plan providers, and healthcare clearinghouses (that process non-standard health information).[30] Although not all research laboratories are covered entities, laboratories become covered entities if they process patient billing or insurance information.[31] Likewise, in Australia, 'APP entities' must comply with 'Australian Privacy Principles' (APPs) when processing personal information.[32] APP entities include Australian government agencies or 'organizations' as defined as any natural or legal person making over 3 million dollars a year.[33]

Data Available for Processing

The *FADP* defines personal data as 'information about an identified or identifiable person', with sensitive personal data including 'data on . . . health'.[34] Furthermore, the *HRA* and *HRO* apply to research involving human tissue or/and health-related data.[35] These definitions encompass the processing of biological or medical data for any purpose, research-related, or otherwise and prohibit the processing of health-related data 'without justification'.[36] However, these regulations do not apply to research involving anonymized biological material or health data.[37] The *HRO* describes anonymization as the irreversible masking or the deletion of all items that would enable the data subject to be identified without disproportionate effort.[38] However, the *HRA* and *HRO* still apply to coded biological material and health-related data, which should be considered analogous to pseudonymized data.[39] The GDPR adopts an equivalent definition of personal data and lists 'genetic data', 'biometric data', and 'data concerning health' as special categories of personal data.[40] Furthermore, in a similar fashion to the *FADP* and *HRA*, the GDPR does not apply to anonymized data. Anonymized data are defined as

29 *Health Information Act 2000* (Alberta) [Alberta HIA], §§ 1(f), 1(k); *Personal Health Information Privacy and Access Act 2009* (New Brunswick) [New Brunswick PHIPA], § 1 (definition of 'custodian'); *Personal Health Information Act 2008* (Newfoundland) [Newfoundland PHIA], §§ 3(f), 4(1); *Personal Health Information Act 2013* (Nova Scotia) [Nova Scotia PHIA], § 1(f); *Personal Health Information Protection Act 2004* (Ontario) [Ontario PHIPA], § 3(1).

30 45 C.F.R §§160.103.

31 Barbara J. Evans & Gail P. Jarvik, *Impact of HIPAA's Minimum Necessary Standard on Genomic Data Sharing*, 20 Genet. Med. 531–535, 532 (2018).

32 *Privacy Act 1988* (Cth), § 6.

33 *Privacy Act 1988* (Cth), §§ 6C–6D.

34 Swiss FADP, *supra* note 31, Art. 3.

35 Swiss HRA, *supra* note 33, Art. 2(1)(d)–(e).

36 Swiss FADP, *supra* note 31, Art. 12(2)(c).

37 Swiss HRA, *supra* note 33, Art. 2(2)(b)–(c).

38 Swiss HRA, *supra* note 33, Art. 3(i), Swiss HRO, *supra* note 34, Art. 25.

39 Swiss HRO, *supra* note 34, Art. 26.

40 GDPR, *supra* note 35, Art. 4(1), (13)–(15), Art. 9(1).

information not relating to an identified or identifiable natural person.[41] In contrast, pseudonymized data, or data no longer attributable to an individual without other data, remain personal data and are protected under the GDPR. In a working paper, the former Article 29 Working Party held that no anonymization or pseudonymization techniques completely protects against re-identification.[42] As discussed previously, this conclusion is supported by technical literature.[43] Therefore, the working paper suggests adopting a contextual, relative-based approach to prevent released data from being re-identified. This contextual-based approach is supported by EU Court of Justice (CJEU) precedent that IP addresses are personal data if combined with additional data to re-identify users.[44] While this decision ostensibly extends the relative approach from the former Article 29 Working Group, it retains a relatively broad interpretation of personal data.[45]

In addition, EU member states that implementing GDPR into national law can also introduce further conditions, including limitations on processing biometric, genetic, or health-related data.[46] In both Italy and Spain, relatives of deceased persons can exercise rights with respect to the processing of the deceased's data.[47] Although personal data under the *Data Protection Act 2018* (UK), in practice, the Medical Research Council and the National Health Service (NHS) treat data from deceased persons as personal data. While the Italian Privacy Code is congruent with the GDPR regarding anonymization and pseudonymization, the Spanish legislation requires separation between the pseudonymized data and the entity conducting the pseudonymization.[48] Historically, UK courts have adopted a more flexible definition of anonymization and pseudonymization. For example, in *Common Services Agency v Scottish Information Commissioner*,[49] Baroness Hale held that if the data were delivered to an entity that had no means to re-identify the data, it would remain anonymized.[50] Likewise, the High Court in *R (Department of Health) v Information Commissioner* held that low cell count statistics on abortion were anonymized when released without other identifying information.[51] However, when these cases were decided, the *Data Protection Act 1998* did not define anonymized data, with Baroness Hale referring to the Data Protection Directive.[52] In contrast, the current *Data Protection Act 2018* uses the same definition of anonymized as the GDPR. Furthermore, more recent guidance from the UK Information Commissioner's Office (ICO) has realigned UK standards on anonymization with those of the EU Article 29 Working Party.[53] Accordingly, pseudonymized data trans-

41 GDPR, *supra* note 35, Recital 26.
42 Article 29 Data Protection Working Party, *Opinion 05/2014 on Anonymisation Techniques* 3–4, 23–24 (2014).
43 Lin, Owen, and Altman, *supra* note 22; Homer et al., *supra* note 22; Gymrek et al., *supra* note 22; Emam et al., *supra* note 23.
44 Case C-582/14 *Breyer v Bundesrepublik Deutschland* [2016] ECLI:EU:C:2016:779, 47–49.
45 van Veen, *supra* note 20, at 73.
46 GDPR, *supra* note 35, Art. 9(4).
47 Decree Law 101/2018 amending Decree Law 196/2003 (the Privacy Code) [Italy], Art. 2-terdecies; Organic Law 3/2018 [Spain], Art. 3.
48 Spanish Organic Law 3/2018, Seventeenth Additional Provision, 2(c).
49 [2008] UKHL 47.
50 *Id*. at 92.
51 [2011] EWHC 1430 (Admin), at 68–70.
52 [2008] UKHL 47, at 91.
53 Christopher Graham, *Anonymisation: Managing Data Protection Risk Code of Practice* 14–15 (2012).

ferred to a third party without identifiers will no longer constituted the anonymized data under UK law.

In the USA, *HIPAA* defines 'protected health information' (PHI) to include information regarding health status, provision of healthcare or healthcare payment, or genetic information of an individual. PHI retains this status for up to 50 years after the death of that individual.[54] Nevertheless, this definition has a limited scope compared with the broad definition under European law. First, PHI does not extend to information which may nevertheless be used to draw conclusions about a patient's health indirectly, such as life insurance or purchase data. Furthermore, *HIPAA* does not apply to health data generated about an individual without the use of healthcare, such as data generated using healthcare mobile applications.[55] Finally, *HIPAA* provides three methods for de-identification: by expert determination, by removing 18 identifiers (including genetic information), or by removing 16 identifiers from a dataset for research purposes.[56] These identifiers include any unique identifying number, character, or code.[57] Furthermore, the covered entity de-identifying the data must have no actual knowledge or information that could re-identify an individual in the dataset.[58] However, the covered entity is under no obligation to assess the potential for re-identification for the remaining data when it is placed in another environment. Therefore, the *HIPAA* standard of de-identification is arguably less strict than that found under Swiss and EU law.

Under Canadian law, 'personal information' refers to any information about an identifiable individual.[59] Canadian courts have defined 'personal information' expansively to include information that relates to or concerns the subject.[60] *PIPEDA* also defines 'personal health information' as including physical or mental health information, tissue and information collected incidentally to providing health services, and health information from deceased persons.[61] The question of whether the anonymized data, particularly anonymized genomic data, remain personal data under Canadian law that has not been addressed in federal legislation.[62] However, in *Gordon v Canada (Minister for Health)*,[63] the Federal Court held that the test for anonymization is whether there is a 'serious possibility' of re-identification.[64] This test has been confirmed as the appropriate pan-Canadian test for anonymization by the Office of the Privacy Commissioner of Canada.[65] Nevertheless, some provincial legislation also provides separate standards for anonymization or de-identification. In Alberta, New Brunswick, and Newfound-

54 45 CFR §160.103.

55 W. Nicholson Price & I. Glenn Cohen, *Privacy in the Age of Medical Big Data*, 25 Nat. Med. 37, 39 (2019).

56 45 CFR §164.514(b)(2).

57 45 CFR §164.514(b)(2)(i)(R).

58 45 CFR §164.514(b)(2)(ii).

59 *PIPEDA 2000*, § 2; *Privacy Act 1985*, § 3.

60 *Dagg v Canada (Minister of Finance)* [1997] 2 SCR 403, 68–69.

61 *PIPEDA 2000*, § 2.

62 Anne-Marie Tassé, *A Comparative Analysis of the Legal and Bioethical Frameworks Governing the Secondary Use of Data for Research Purposes,* 14 Biopreservation Biobanking 207–216, 210 (2016).

63 2008 FC 258.

64 *Id.* at 34.

65 *2016–17 Annual Report to Parliament on the Personal Information Protection and Electronic Documents Act and the Privacy Act*, Office of the Privacy Commissioner of Canada (Sept. 21, 2017), https://www.priv.gc.ca/en/opc-actions-and-decisions/ar_index/201617/ar_201617/.

land, health data custodians are empowered to 'strip, encode, or otherwise transform' health information to make it 'non-identifying'.[66] In contrast, the equivalent legislation in Ontario requires the removal of any information that could be used to identify the individual.[67] In Australia, 'personal information' is defined information about an identified individual or an individual who is reasonably identifiable with 'sensitive information' including health, genetic, or biometric information.[68] It is ambiguous as to whether this definition applies to just identifiable data or also aggregate data.[69] Case law has failed to clarify this question with respect to IP addresses that may be re-identified using with other information.[70] If this definition includes aggregate data that could re-identify an individual using other data, aggregate data would remain personal data. Finally, there is no clear definition in any of the jurisdictions under study as to whether encrypted data represent a special category of data or whether it falls into the categories supplied already.

Consent or Approval Required for Processing

In Switzerland, sensitive data cannot be disclosed unless the data subject has given consent or has made their data generally accessible and not prohibited disclosure.[71] A person may be only involved in a research project if they have given consent in writing.[72] Furthermore, a person must receive written information on the nature of the project, the burdens and benefits taken, and data protection strategies.[73] A person must have also been given adequate time to decide whether to participate in the project.[74] If the researcher intends to make further use of the data, they must seek general consent from the person when data are collected. The relevant ethics committee should approve any further use for research purposes in case of lack of consent or information to the right to dissent that has not been obtained.[75] The HRO describes 'further use' as collecting, cataloguing, storing, making accessible, or transferring biological material or health-related data.[76] This broad definition of further use encompasses linkages between different datasets. The *HRA* then specifies the type of consent available for different types of data, which are described in Table 2.

The Swiss Academy of Medical Sciences' general consent form satisfies the requirements under the *HRA* and *HRO*. The General Consent Form sets out the principles for the reuse of materials where patients have provided general consent for their tissue, data, or other samples to be made available for research. Only authorized hospital personnel are permitted to access identifiable records, with researchers restricted to

66 Alberta HIA, *supra* note 37, § 1(1)(r), § 65; New Brunswick PHIPA, *supra* note 37, § 51; Newfoundland PHIA, *supra* note 37, § 21.
67 Ontario PHIPA, *supra* note 37, § 47.
68 *Privacy Act 1988* (Cth), § 6.
69 Joshua Yuvaraj, *How About Me? The Scope of Personal Information Under the Australian Privacy Act 1988*, 34 Comput. Law Secur. Rev. 47–66, 50 (2018).
70 *Privacy Commissioner v Telstra Corporation Limited* (2017) FCAFC 4, 57–66; *Id*. at 48–49.
71 Swiss FADP, *supra* note 31, Art.17, 19.
72 Swiss HRA, *supra* note 33, Art. 16(1).
73 Swiss HRA, *supra* note 33, Art. 16(2).
74 Swiss HRA, *supra* note 33, Art. 16(3).
75 Swiss HRA, *supra* note 33, c. 8, Art. 45(b).
76 Swiss HRO, *supra* note 34, Art. 24.

Table 2. Comparison of different forms of consent that are available for different types of data and data protection measures in Switzerland[77]

	Types of data Further research use of biological material and genetic data (Art. 32 HRA)	Further use of non-genetic health-related personal data (Art. 33 HRA)
Data protection measure	Permissible Forms of Consent	
Personal identifying data	Informed consent for specific research project(s) required (Art. 32(1) HRA, Art. 28 HRO)	General consent required (Art. 33(1) HRA, Art. 31 HRO)
Coded	General consent required (Art. 32(2) HRA, Art. 29 HRO)	General consent required (Art. 33(2) HRA, Art. 32 HRO)
Anonymized	Data subjects have to be informed about anonymization and should not object to it	Not regulated by HRA

coded data.[78] When informed consent requirements have not been met, researchers may use biological material or health-related data if consent is disproportionately difficult to obtain and there is no documented refusal. Furthermore, the interests of research must outweigh the interests of the person concerned in deciding on the further use of his or her biological material and data.[79] Under the *FADP*, researchers can disclose sensitive data for scientific reasons, as well as process it for further purposes. However, researchers must ensure that the data are rendered anonymous as soon as possible, so data subjects cannot be re-identified.[80]

The GDPR requires express consent for the processing of sensitive personal data (Art. 9(2)(a)).[81] Although the GDPR ostensibly recognizes the possibility for general consent via Recital 33, the former Article 29 Working Party held that only general consent for a specific research project is valid.[82] However, the GDPR also creates a number of exceptions for processing sensitive data. First, sensitive data can be processed for

77 Effy Vayena et al, *Responsible Data Processing in Health Research* (2017), https://sphn.ch/wp-content/uploads/2020/01/Report_20170825_Responsible-Data-Processing-in-Health-Research_SPHN-ELSIag.pdf.

78 Unimedsuisse.ch, *Generalkonsent für die Forschung: Unimedsuisse—der Verband der Schweizer Hochschulmedizin*, https://www.unimedsuisse.ch/de/projekte/generalkonsent (accessed Aug. 29, 2019).

79 Swiss HRA, *supra* note 33, Art. 34.

80 Swiss FADP, *supra* note 31, Art. 22.

81 GDPR, *supra* note 35, Art. 9(2)(a).

82 Article 29 Data Protection Working Party, *Working Paper 259 Guidelines on Consent under Regulation 2016/679* (2017).

preventative medicine, medical diagnosis, or providing healthcare services, subject to professional standards of secrecy.[83] Second, sensitive data may be processed for reasons of public interest in the area of public health, such as protecting against threats to public health or ensuring medical device quality.[84] Third, sensitive data may be processed for archiving purposes in the public interest, scientific or historical research purposes, or statistical purposes.[85] However, any statistical or research-related processing requires researchers to first determine whether they can work without personal data or with anonymized data.[86] If anonymized or non-personal data are not sufficient, researchers should determine whether processing can be conducted with pseudonymized data. Finally, under the GDPR, 'scientific research purposes' and 'statistical purposes' are not limited to research or processing in the public interest.[87] Accordingly, privately funded research may be permissible under the exception.[88] In addition, this exception may also permit a significantly broader range of processing on aggregate data on the grounds of 'statistical processing' without adequate ethics approval. This exception may have the effect of allowing larger data processors with technical infrastructure to bypass scrutiny while exposing smaller researchers to heightened compliance costs.[89]

In Italy, the Privacy Code requires processing of genetic, biometric, and health-related data to comply with best practices from European Data Protection Board (EDPB) (the replacement for the Article 29 Working Party), as well as any scientific and technological advances.[90] The Privacy Code also requires a data privacy impact assessment and an ethics committee assessment before any processing of health-related data can occur. Finally, the Privacy Code requires technical measures for security, access control, and data minimization.[91] However, the Privacy Code also permits data to be used for further purposes beyond those initially consented to by data subjects.[92] In Spain, consent alone is not sufficient to lift the prohibition on processing special categories of data mandated by Art. 9(1) of the GDPR. Instead, processing requires the subject's consent and a purpose defined in Arts. 9(2)(g), (h), or (i) of the GDPR, unless for cases of exceptional relevance and seriousness to public health.[93] The reuse of health and biomedical data for scientific research must be 'scientifically integrated' into the initial study and approved by an independent ethics committee.[94] In the UK, reuse of data for research is managed via NHS Digital, previously known as the Health and

83 GDPR, *supra* note 35, Art. 9(2)(h).
84 GDPR, *supra* note 35, Art. 9(2)(i).
85 GDPR, *supra* note 35, Art. 9(2)(j).
86 GDPR, *supra* note 35, Art. 89(1).
87 GDPR, *supra* note 35, Recital 159, 162.
88 Kärt Pormeister, *Genetic Data and the Research Exemption: Is the GDPR Going Too Far?*, 7 Int. Data Priv. Law 137–146, 142 (2017).
89 Effy Vayena et al., *How the General Data Protection Regulation changes the rules for scientific research* (2019), http://www.europarl.europa.eu/RegData/etudes/STUD/2019/634447/EPRS_STU(2019)634447_EN.pdf.
90 Decree Law 101/2018 amending Decree Law 196/2003 (the Privacy Code) [Italy], Art. 2-*septies*, ¶ 2.
91 Guisella Finocchiaro, *Italy: The Legislative Procedure for National Harmonisation with the GDPR Reports: GDPR Implementation Series*, 4 Eur. Data Prot. Law Rev. 496–499, 498 (2018).
92 Decree Law 101/2018 amending Decree Law 196/2003 (the Privacy Code) [Italy], Art. 99.
93 Spanish Organic Law 3/2018, Seventeenth Additional Provision, 2(b).
94 Spanish Organic Law 3/2018, Seventeenth Additional Provision, 2(c).

Social Care Information Centre.[95] NHS Digital has the authority to distribute patient data for public health reasons, including efficient use of National Health System (NHS) resources, or with patient consent.[96] Because NHS Digital has a broad ambit to disclose patient data for particular reasons, this data can be accessed by both private- and public-sector researchers.[97]

In the USA, PHI may only be disclosed to third parties without written consent in three circumstances: for law enforcement, administrative purposes, or to facilitate treatment, payment, or healthcare.[98] In the alternative, data may be disclosed in de-identified form. The 'Common Rule' then requires all human research funded by the US federal government to comply with relevant national laws.[99] The Common Rule also requires all relevant research to be subject to approval by an institutional review board (IRB). In addition, the US National Institute of Health (NIH) has published its own Genomic Data Sharing (GDS) Policy. This policy applies to all NIH funded research that involves generating or using human genomic data, including for secondary purposes or data sharing. The NIH GDS Policy implicitly encourages adopting tiered or dynamic consent and requires institutions that generate large quantities of data to supply it to a data repository. However, the NIH GDS Policy, while permitting such sharing, also makes the data processor and not the owner of the repository liable for any breaches. This displacement of liability might discourage further data linkage with overseas research institutes.

In Canada, under Federal legislation, commercial organizations can only collect personal information with the valid consent of an individual or if an exception applies.[100] However, *PIPEDA* provides that where certain circumstances exist, data controllers can use or disclose information without knowledge or consent of the subject.[101] These include for emergency medical treatment and research purposes where it would be impracticable to obtain consent (provided the research has been disclosed to the Privacy Commissioner).[102] Personal information that is no longer required to fulfil the identified purposes should be destroyed, erased, or made anonymous when held by a private organization.[103] Likewise, under the *Privacy Act*, government agencies can only use personal information consistent with the purpose for which it was collected.[104] This personal information can then only be disclosed for a consistent purpose or if an exception applies.[105] As for PIPEDA, these exceptions include for research related purposes, provided the research could not be completed without identifying information, and no individuals will be re-identified.[106] Finally, the *Tri-Council Policy Statement:*

95 *Health and Social Care Act 2012* (UK), Part 9, c. 2.
96 *Health and Social Care Act 2012* (UK), Part 9, c. 2, § 261.
97 Miranda Jane Mourby et al., *Health Data Linkage for Public Interest Research in the UK: Key Obstacles and Solutions*, 4 Int. J. Popul. Data Sci. 5 (2019).
98 45 CFR §§ 164.502, 164.506, 164.510, 164.512.
99 45 CFR § 46.
100 *PIPEDA 2000*, §§ 6.1, 7(1).
101 *PIPEDA 2000*, §§ 7(2)–(3).
102 *PIPEDA 2000*, §§ 7.1(2)(b).
103 *PIPEDA* 2000, Schedule 1, Clause 4.5, 4.5.3.
104 *Privacy Act 1985*, §§ 7, 8(1).
105 *Privacy Act 1985*, §§ 8(1)–(2).
106 *Privacy Act 1985*, § 8(2)(j).

Ethical Conduct for Research Involving Humans ('TCPS2') mandates that all research involving human participants in Canada require ethics committee approval.[107] Public-sector data do not require ethics approval for use, except where data linkage could result in re-identification.[108] The TCPS2 only permits the use of health-related data without consent for secondary purposes if it is impossible to seek patient's consent and if identifiable information is essential to research.[109] Nevertheless, these requirements will not apply to healthcare institutes or research agencies that are bound by provincial health and freedom of information legislation. These legislative instruments create additional consent or ethics approval requirements for processing or disclosing health-related data for research. For example, legislation in a number of provinces requires an agreement between a researcher and custodian or trustee specifying terms of use before the data transfer occurs.[110]

In Australia, APP entities may only use or disclosure personal health information for research where it is impractical to obtain consent or a permitted health situation applies.[111] Permitted health situations include providing health-related services, public health management, or research defined under Sections 95A and 95A.[112] These sections allow the National Health and Medical Research Council (NHMRC) to establish rules for the processing personal data for research. The NHMRC National Statement on Ethical Conduct in Research then allows researchers to link identifiable patient data for scientific research purposes. In particular, the Statement explicitly distinguishes between individually identifiable data and re-identifiable data (as opposed to 'de-identified' data due to the ambiguity of the latter term).[113] However, the Statement does not define these terms. Furthermore, as part of the informed consent process, participants must be informed of the potential for data linkage.[114] Finally, all research involving genomics must be subject to ethics committee review. However, if no linkage of genomic data is required, the project may proceed at low risk.[115] An ethics committee might also waive the requirements for consent where there are no suitable arrangements for obtaining consent. In light of these more flexible data linkage policies, there are

107 Interagency Advisory Panel on Research Ethics Government of Canada, *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans—TCPS 2 (2018)* Art. 2.1. (2019), https://ethics.gc.ca/eng/policy-politique_tcps2-eptc2_2018.html (accessed Feb. 17, 2020).

108 *Id.* at Art. 2.2.

109 *Id.* at Art. 5.5.

110 Alberta HIA, *supra* note 37, Part 5, Division 3; *Freedom of Information and Protection of Privacy Act 1996* (British Columbia) [British Columbia FIPPA], § 35(1)(d); *E-Health (Personal Health Access and Protection of Privacy) Act 2008* (British Columbia) [British Columbia E-Health Act], § 19; New Brunswick PHIPA, *supra* note 37, §§ 19(1), 28(e), 28(n), 34(m), 34(m.1), 34(m.2), 34(m.3), 37(5), 38(g.2), 38(g.3), 38(h), (h.01), 43, 43.1; Newfoundland PHIA, *supra* note 37, §§ 15, 31, 34, 38, 44; *Health Information Act 2014* (Northwest Territories) [Northwest Territories HIA], §§ 77, 78, 80; Nova Scotia PHIA, *supra* note 37, §§ 57, 60; Ontario PHIPA, *supra* note 37, §§ 12, 36(1)(d), 37(1)(j), 37(3), 37(4), 44; *Health Information Protection Act 1999* (Saskatchewan) [Saskatchewan], § 29(1).

111 *Privacy Act 1988* (Cth), § 16B.

112 *Privacy Act 1988* (Cth), § 16B.

113 National Statement on Ethical Conduct in Human Research, Purpose, scope and limits of this document (2007), https://www.nhmrc.gov.au/about-us/publications/national-statement-ethical-conduct-human-research-2007-updated-2018 (accessed Feb 17, 2020).

114 *Id.* at § 3.1.43.

115 *Id.* at § 3.3.

notable research projects that have been established to support the linkage of public-sector data.[116] These projects will be discussed in further detail in the second and third sections of this paper.

Transfers of Personal Data across Borders

Reuse and transfer of data across borders under the GDPR are limited to countries assessed as providing adequate protection.[117] At the time of writing, Switzerland has been assessed as offering adequate protection, with transfers limited to private organizations in Canada and the USA. Alternatively, transfers may be subject to appropriate organizational safeguards, or exceptions such as express consent or limited numbers of transfers. Given that only a handful of jurisdictions has been assessed as offering adequate protection, organizational safeguards are currently preferred as the mechanism of transfer.[118] To this end, the EDPB has prepared guidelines on how to develop codes of conduct for transfer.[119] These guidelines address the requirements for data transfer in medical research, and recommend data controllers install safeguards that allow data subjects to exercise their rights. In addition, these guidelines require data controllers and processors to introduce data minimization, security, and retention measures to protect patient data.[120] *PIPEDA* requires equivalent protection in the target jurisdiction for transfer,[121] although Canadian researchers may also be bound by provincial legislation that applies to public researchers. For example, limits on transfer by custodian may be determined by an ethics committee in Alberta.[122] In contrast, personal health data cannot be disclosed outside New Brunswick without the express consent of the relevant individuals.[123] Any transfer of personal data from Australia to a third-party jurisdiction requires compliance with Australian privacy law.[124] In the alternative, the recipient jurisdiction must offer the same degree of protection as Australia or the data subjects must provide their explicit consent to transfer.[125] The exception to these strict data localization guidelines on transfer exists within the USA. Specifically, PHI may be reused for secondary purposes and transferred to other organizations, although approval from an IRB as required by the Common Rule is necessary.[126] A comparison between each of the key principles discussed in this section is contained in Table 3.

116 Rachel Canaway et al., *Gathering Data for Decisions: Best Practice Use of Primary Care Electronic Records for Research*, 210 Med. J. Aust. S12–S16, 12 (2019).

117 GDPR, *supra* note 35, Art. 44–45.

118 Kostas Glinos, *Global Data Meet EU Rules*, 360 Science 467–467, 467 (2018).

119 European Data Protection Board, *Guidelines 1/2019 on Codes of Conduct and Monitoring Bodies under Regulation 2016/679* (2019).

120 *Id*. at 9.

121 *PIPEDA* 2000, Schedule 1, Clause 4.1.3.

122 Alberta HIA, *supra* note 37, § 54(1)(a)(ii).

123 New Brunswick PHIPA, *supra* note 37, § 19(1)(d).

124 *Privacy Act 1988* (Cth), Schedule 1, Part 3, Clause 8.1.

125 *Privacy Act 1988* (Cth), Schedule 1, Part 3, Clause 8.2(a).

126 § 164.512(i)(1)(ii)–(iii), §§ 164.514(a)–(c), 164.514(e).

Table 3. Impact of data protection laws and research ethics laws on data accessibility

	Identifiable data are shared with other data custodian or government entities	University and non-profit researchers may be approved access to de-identified data	Health care providers may be approved access to de-identified data	For-profit businesses may be approved access to de-identified data	Foreign government, university, or non-profit researchers may be approved access to de-identified data
Switzerland	Can only be shared with other entities with consent (Art. 17, FADP)	May access data if anonymized as soon as practicable and disclosed in anonymous form (Art. 22)	See for university researchers	See for university researchers	Transfer of identifiable data forbidden without consent or appropriate protection (Art. 6)
GDPR	Can only be shared with other entities with (Art. 9(2)(a), GDPR)	May rely on the research exception to access pseudonymized health-related and biomedical data (Art. 9(2)(j)). Note this exception must be implemented into national law and therefore may not be available	May rely on public interest, diagnosis, or public health exceptions to access data with consent and confidentiality (Art. 9(2)(g), (h), (i))	May rely on the research exception to access pseudonymized health-related and biomedical data (Art. 9(2)(j)). Note this exception must be implemented into national law and therefore may not be available	Can only be transferred to a country with an adequacy determination, subject to safeguards or limited exceptions (Art. 45, 46, 48)
Italy	Can only be shared with consent, cannot be disseminated to third parties (Art. 167bis)	As for government analysts	As for government analysts	As for government analysts	As for GDPR (Art. 43–45); subject to criminal sanctions

Continued

Table 3. Continue

	Identifiable data are shared with other data custodian or government entities	University and non-profit researchers may be approved access to de-identified data	Health care providers may be approved access to de-identified data	For-profit businesses may be approved access to de-identified data	Foreign government, university, or non-profit researchers may be approved access to de-identified data
Spain	Personal data cannot be shared without consent and a legitimate reason for exchange	As for GDPR, consent/ethics committee approval is required (Art. 9(2), 17th additional provision)	As for university and non-profit researchers	As for university and non-profit researchers	As for GDPR (Art. 40–43)
UK	Can only be shared with consent or unless a legitimate reason for sharing applies (Schedule 2(1))	As for government analytics	As for government analytics	As for government analytics	As for GDPR (section 18, *Data Protection Act*)
USA	Can only be shared for law enforcement, administrative or procedural purposes	De-identified data must be subject to expert determination, safe harbor or limited data set de-identification	As for university and non-profit researchers	As for university and non-profit researchers	Does not apply overseas (but see research ethics laws below)

Continued

Table 3. Continue

	Identifiable data are shared with other data custodian or government entities	University and non-profit researchers may be approved access to de-identified data	Health care providers may be approved access to de-identified data	For-profit businesses may be approved access to de-identified data	Foreign government, university, or non-profit researchers may be approved access to de-identified data
Canada	Can only be shared with consent unless exception applies, de-identified after processing	Should be de-identified as soon as possible, but accessible without consent and ethics committee approval (note provincial legislation)	As for university researchers	As for university researchers	Requires equivalent level of protection for transfer (Schedule 1, Clause 4.1.3)
Australia	Can only be shared with consent or where a permitted health situation creates an exception (ss16A, 16B)	May access where a permitted health situation applies subject to NHMRC rules (s16B, s95, s95A)	As for university researchers	As for university researchers	Only available if reasonable steps have been taken to ensure compliance, consent for transfer or requesting jurisdiction ensures adequate protection
New Zealand	Can only be used or disclosed with consent (s6, cl. 11 *Privacy Act,* s5 cL 10 and 11, *HIPC*)	Not reasonably identifiable can be used without consent subject to ethics approval (s5 cL 10 and 11, *HIPC)*	See for university researchers	Not reasonably identifiable can be used without consent subject to ethics approval (s6 cL. 10 and 11, *Privacy Act*)	Only available if reasonable steps have been taken to ensure compliance, consent or adequate protection (Part 11A, *Privacy Act* 1993)

POTENTIAL REGULATORY OBSTACLES TO MULTISITE RESEARCH AND APPROPRIATE REGULATORY OR TECHNICAL SOLUTIONS

Regulatory Convergence

The comparison of legislation above demonstrates a number of impediments between each of the four features of each regulatory regime. First, with respect to entities that carry out processing, despite the different legislative approaches to data protection, there is considerable convergence toward a broad approach to processing liability. That is, processors of personal data will still be responsible under data protection law, irrespective of whether they are private individuals or working for public- or private-sector organizations. This responsibility extends to both the recipients and the providers of personal data. Two exceptions exist in the USA and Australia, where researchers that are not covered entities or APP entities, respectively, are not held responsible in either jurisdiction for data processing. The ambiguity regarding regulatory limits could manifest where researchers in the USA or Australia, who assuming they are not liable, link or reuse data without awareness as to their obligations. Likewise, researchers from outside the USA working with US datasets need to be mindful of whether their research is funded by the US Federal government and is therefore protected by the Common Rule. This question of data linkage dovetails into the question of what constitutes personal data and, particularly, whether data from deceased persons or anonymized data are 'personal data'. With respect to data from deceased persons, there appears to be a convergence that personal data include data from deceased persons. The exception exists in Australia, where personal information does not include information from deceased person. However, as discussed previously, the NHMRC National Statement on Ethical Conduct in Research provides a mechanism for using data without obtaining consent where impractical to do so. Accordingly, this consent waiver might provide a pathway for researchers in Australia to reuse health-related data from the records of deceased individuals.[127]

Regulatory Divergence

Nevertheless, a key point of ambiguity identified in each jurisdiction remains the question of what will satisfy the requirements for anonymization. Furthermore, the divergence between each jurisdiction is more severe when considering the question of anonymization and pseudonymization. As alluded to, this divergence is particularly pronounced when comparing the approach adopted in the USA under the *Privacy Rule* and those approaches adopted in Switzerland, Italy, and Spain. In the USA, if a covered entity has removed the requisite set of identifiers from the data, they are no longer under an obligation to identify any privacy risks that emerge from the remaining dataset. Therefore, from a legal perspective, this de-identified data may not meet the requisite standard of anonymization imposed by Swiss or EU data protection law.[128] Furthermore, the standard of anonymization imposed in the UK remains uncertain as well. As discussed above, NHS Digital is responsible for disclosing anonymized or identifying patient information upon request. However, neither the *Health and*

127 Lisa Eckstein et al., *Australia: Regulating Genomic Data Sharing to Promote Public Trust*, 137 Hum. Genet. 583–591, 585 (2018).

128 Dove and Phillips, *supra* note 24, at 662.

Social Care Act 2012 nor NHS Digital provides a definition of anonymization, instead relying on the ICO Anonymization Code of Practice. The Anonymization Code of Practice itself does not require any single one technique, but instead, it requires data processors to consider whether there is an 'unacceptable risk of re-identification'.[129] However, the UK Anonymisation Network (UKAN) is critical of this approach constituting 'anonymization', on the grounds that anonymization under UK law also requires considering the environment where data are released.[130] Furthermore, because NHS Digital releases anonymized data where justified, as opposed to specific parties, there is uncertainty as to whether these environmental factors are considered for a release.[131]

Whether the Swiss or EU standard of anonymisation or pseudonymisation is congruent with the equivalent standard under Canadian law is unclear. This uncertainty is primarily due to the overlapping jurisdiction of federal and provincial laws.[132] In particular, the ability to 'strip, encode, or otherwise transform personal health information' to create de-identified information appears equivalent to anonymization. However, guidance documents from these provinces provide varying anonymization. For example, the Alberta Health Services Guidelines distinguish between anonymous, aggregate, and de-identified data (which is roughly analogous to coded or pseudonymized data). The Guidelines then proceed to describe the second method of de-identification under *HIPAA* as the correct form of rendering personal information non-identifying.[133] This approach can be contrasted by the approach adopted in Ontario, which appears to be closer to the contextual standard of anonymization required under Swiss and EU data protection law. The uncertainty about the scope of anonymization within each province has a consequent effect on reusing data or transferring data between jurisdictions. Furthermore, the TCPS2 lacks an express provision on the data sharing requirements, particularly between institutions in Canada. Therefore, this absence of data sharing guidance may undermine attempts to balance data protection against the need to ensure data accessibility and widely disseminate results.[134]

In Australia, the oblique precedent concerning the scope of personal data obscures what constitutes the processing of anonymized data. As discussed previously, linkages of patient data are permissible under the NHMRC National Statement, although the Statement does not define whether non-identifiable data remain personal data. The risks in Australia are increased by the fact that in practice, government agencies decide whether to open their datasets without reference to those with relevant expertise.[135] However, it can be very difficult to preserve confidentiality in the absence of a uni-

129 Graham, *supra* note 61, at 17.

130 Mark Elliot et al, *The Anonymisation Decision-Making Framework* 16, 52–8 (2016).

131 Mourby et al., *supra* note 105, at 5.

132 Adrian Thorogood, *Canada: Will Privacy Rules Continue to Favour Open Science?*, 137 Hum. Genet. 595–602, 597 (2018).

133 Chief Privacy Officer, *Non-Identifying Health Information*, Alberta Health Services (2014), https://extranet.ahsnet.ca/teams/policydocuments/1/clp-ahs-privacy-non-identifying-health-information-standard-ipo-2013-0004.pdf.

134 Dove and Phillips, *supra* note 24, at 672.

135 Keiran Hardy & Alana Maurushat, *Opening Up Government Data for Big Data Analysis and Public Benefit*, 33 Comput. Law Secur. Rev. 30–37, 34 (2017).

versal patient identifier while matching patient data.[136] This heightened risk of re-identifiability is due to the larger range of fields in the final linked dataset.[137] The Office of the Australian Information Commissioner (OIAC) has published additional guidance on what constitutes appropriate anonymization or de-identification. Specifically, the OIAC has published an Australian adaptation of the UKAN Anonymisation Decision-Making Framework as a guide for anonymization and contextual control.[138]

The second relates to the circumstances in which health records can be reused for research-related purposes. As discussed in the introduction, there is a considerable benefit to linking data from electronic health records for generating new insights in public health or personalized medicine. Nevertheless, unless general consent for future has been sought from each individual before their records were processed, researchers may not have lawful grounds to process these data.[139] Australia, Canada, Switzerland, and the USA all feature either federated or code of conduct driven regulations of general consent and data linkage. As discussed previously, in the context of anonymization, this style of regulation provides significant flexibility for both regulators and researchers to respond to technological change. However, it also provides regulators with the capacity to introduce new forms of consent mechanism to permit the reuse of data. This flexibility can be contrasted with the ostensible unavailability of general consent under the GDPR, which may place significant limits on the reuse of personal data. Although Italy, Spain, and the UK have created derogated exceptions for consent in biomedical research, there remain outstanding questions on whether these derogations will remain compatible with the GDPR.[140] Furthermore, it may be impossible to seek general consent for research in countries that do not have these derogated exceptions.

Outside of the risks of re-identification, the regulatory frameworks do not impose significant obstacles on the transfer of data between organizations. However, transfer across jurisdictions constitutes a significantly more severe regulatory to obstacle data sharing and data linkage. Aside from the USA, data cannot be transferred out of any of the jurisdictions discussed above without equivalent protection in the target jurisdiction or safeguards to ensure data protection. Both EU and Swiss law provide a narrow set of criteria under which personal data can be transferred in the absence of adequate legislative, including explicit consent of the subject. For example, the former Article 29 Working Party held that transfers of private data to Australia could only occur if EU equivalent safeguards were provided for this data in Australia. In part, the need for adequate protection was due to the fact that some sectors and activities were excluded from the protection of the act, as alluded to previously.[141] Furthermore, the provincial framework in Canada places significant limitations on transfer even between different

136 Roger S. Magnusson, *Data Linkage, Health Research and Privacy: Regulating Data Flows in Australia's Health Information System*, 24 Syd. Law Rev. 5–56, 27 (2002).

137 Dharmenaan Palamuthusingam et al., *Health Data Linkage Research in Australia Remains Challenging*, 49 Intern. Med. J. 539–544, 540 (2019).

138 Christine O'Keefe et al., *De-identification Decision-Making Framework* (2017), https://www.oaic.gov.au/privacy/guidance-and-advice/de-identification-decision-making-framework/ (accessed Jan. 3, 2020).

139 van Veen, *supra* note 20, at 74.

140 Jiahong Chen, *How the Best-Laid Plans Go Awry: The (Unsolved) Issues of Applicable Law in the General Data Protection Regulation*, 6 Int. Data Priv. Law 310–323, 313–314 (2016).

141 Article 29 Data Protection Working Party, *Opinion 3/2001 on the level of protection of the Australian Privacy Amendment (Private Sector) Act 2000* (2001).

Table 4. Comparison of transfer between jurisdictions and in particular whether data can be transferred without contractual mechanisms

From	To				
	Switzerland	EU-GDPR	US	Canada	Australia
Switzerland	N/A	Adequate	Adequate[a]	Contract	Contract
EU-GDPR	Adequate	N/A	Adequate[a]	Adequate[a]	Contract
US	Adequate[b]	Adequate[b]	N/A	Adequate[c]	Contract
Canada	Contract[d]	Contract[d]	Contract[d]	N/A	Contract[d]
Australia	Contract[e]	Contract[e]	Contract	Contract[e]	N/A

Key: Adequate = can be transferred without contractual mechanisms. Contract = can only be transferred with contractual or corporate governance mechanisms.
[a]Only transferable to private organizations (privacy shield).
[b]Transferrable, but requires compliance with common rule for federally funded research.
[c]Transferrable, but requires compliance with PIPEDA depending on the entity that is sending the data; while commercial entities must comply with PIPEDA, non-profit or government entities may be uncovered.
[d]Depends on provincial privacy laws, may require contractual mechanisms or consent.
[e]Requires contractual or corporate governance mechanisms or equivalence.

provinces in Canada. On the one hand, some provincial legislation provides ethics committees with a broad discretion to permit the exchange of data across borders. Others do not specify when data can be transferred or require appropriate consent for transfer. Furthermore, these frameworks override the operation of *PIPEDA* or the *Privacy Act 1985*. Due to the difficulty in re-obtaining consent for research as previously discussed, these divergences may therefore place significant limits on the exchange of data across jurisdictional borders without appropriate ethics approval.[142] This legal conclusion is supported by survey and interview evidence from the OECD report discussing delays in the authorization of exchange of de-identified health-related data.[143] Therefore, the greatest regulatory obstacles identified in this paper are as follows: inconsistent definitions regarding de-identification, divergences in consent requirements, and the difficulty in ensuring adequate protection for transfer between jurisdictions (Table 4).

APPROPRIATE REGULATORY AND TECHNICAL SOLUTIONS

Before offering appropriate regulatory and technical solutions, it is important to consider that the research ecosystem is in a constant state of regulatory flux. Traditionally, researchers submit a project protocol requesting access to health data stored in hospitals or some other facility. The research ethics committee at that institute approves the research protocol, checks whether consent extends to the purposes requested, and monitors the project. The introduction of electronic patient data storage facilitates not only the accessing but also the reuse of patient data. In the current research landscape, there are multiple internal organizations that often have overlapping roles in managing the release of data for secondary purposes. In universities, data access committees (DACs) play an increasingly important role in designing access request policies for biomedical data, including genomic data.[144] These DACs are complemented by data

142 David B. Hogan et al., *Ethical and Legal Considerations for Canadian Registries*, 40 Can. J. Neurol. Sci. S5–S22, S6 (2013).
143 OECD, *supra* note 11, at 88, 195–197.
144 Shabani, Knoppers and Borry, *supra* note 27, at 507.

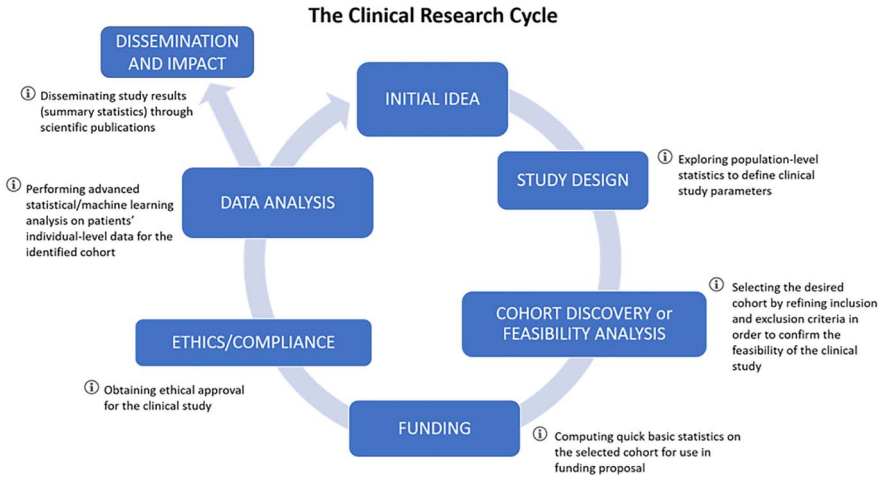

Diagram 1. The clinical research cycle.

access control offices, which were first established as part of the International Cancer Genome Consortium (ICGC) project. The purpose of this office was to standardize the access process for ICGC data across participating institutions.[145] The change in the current research cycle is demonstrated in Diagram 1.

However, this ease of access and reuse raises privacy and confidentiality concerns that warrant organizational solutions and technological advances to guarantee privacy. Furthermore, as Shabani, Knoppers, and Borry notes, for these mechanisms to work, there must be clearly defined organizational responsibilities.[146]

Organizational Solutions

A significant number of the legal and ethical obstacles discussed above result from a lack of focus upon the context for processing, disclosure, and use of health-related data. On one hand, the broad, absolute definition of personal data contained within EU and Swiss data protection law does not ostensibly reflect the potential for re-identification from a particular dataset. In other words, completely preventing any potential re-identification may reduce the usefulness of the data set entirely.[147] In turn, useless data undermines the rationalization for using data for research purposes or may even result in poor conclusions being drawn from data. On the other hand, the narrow definition of PHI contained within *HIPAA* and the *Security Rule* may fail to prevent the disclosure of personally identifying information via inferences. Prima facie, both regimes do not ultimately acknowledge that the risk of re-identification may vary significantly depending on the content of each dataset. Therefore, one strategy to resolve this problem involves the use of contextual disclosure based on organizational agreements between agencies. This strategy reflects the best practices in the UKAN Anonymisation Decision-Making Framework to consider both technical and contex-

145 Yann Joly et al., *Data Sharing in the Post-Genomic World: The Experience of the International Cancer Genome Consortium (ICGC) Data Access Compliance Office (DACO)*, 8 PLOS Comput. Biol. e1002549 (2012).

146 Shabani, Knoppers & Borry, *supra* note 27, at 509.

147 Emam et al., *supra* note 23, at 2.

tual risks associated with reuse and disclosure. Furthermore, the limited existing case studies of successful linkages of medical data already rely on organizational links for multisite research.

For example, one of the oldest and most successful ongoing projects involving data linkage is the Western Australian Data Linkage System (WADLS), established in 1998. Initially, the WADLS only contained data generated from local hospital records. However, Australia features a hybrid federated system of governance, where both Federal and State governments are responsible for healthcare. Any healthcare provided to patients by state hospitals would nevertheless be funded by Australia's national healthcare System Medicare and the Pharmaceutical Benefits Scheme. Linking this data together could provide rich information for longitudinal health studies, but it could also raise significant risks of re-identification or other breaches of confidentiality. Accordingly, a research protocol was developed and approved by the then Australian Privacy Commissioner to link these datasets to patient datasets from state-funded healthcare providers. First, this protocol involved a memorandum of understanding where the responsibilities of each party were established and ethics approval for creating linked data files was obtained from both State and Federal committees. Additional ethics approval for the use of linked data files in individual research projects then requires minimal bureaucratic delay, as researchers receiving the data only to see the unique identifiers.[148] Although ethics approval was sought to reuse the data without consent, the previous benefits of longitudinal health projects served to engender community trust in health data linkage.[149]

Therefore, similar organizational strategies need to be adopted for health data linkage projects to ensure the successful reuse of data. First, data can only be used or shared between trusted organizations. The identity and responsibilities of these trusted organizations must be determined with the assistance of participating institutions. These include the data custodians such as hospitals and physicians who collect the data from patients, as well as the research institutions responsible for processing data. Furthermore, research ethics committees can play an important role in determining the responsibilities of each institution. In addition, research ethics committees should provide guidance on developing an initial ethics approval to minimize the bureaucratic burden for ongoing uses of data. Where the responsible institutions operate in separate jurisdictions, this process should account for any legislative divergence and can be reinforced by binding contractual mechanisms (as discussed below). Finally, research ethics committees should provide guidance to researchers on mechanisms under which the relevant data can be made available for reuse. These strategies can include promulgating general consent forms to patients to make data available for further use, or determining whether a waiver can be sought for the ongoing use of data.

An additional but equally important step is ensuring that public trust is obtained for the use of the data in this fashion. Although the WADLS has been enormously successful, other projects such as the *care.data* program in the UK have been cancelled due to a lack of outreach and resultant community outrage. In particular, the exchange of patient

148 C. W. Kelman, A. J. Bass & C. D. J. Holman, *Research Use of Linked Health Data—A Best Practice Protocol*, 26 Aust. NZ. J. Public Health 251–255, 253 (2002).

149 C. D'Arcy, J. Holman et al., *A Decade of Data Linkage in Western Australia: Strategic Design, Applications and Benefits of the WA Data Linkage System*, 32 Aust. Health Rev. 766–777, 768 (2008).

data to private companies for commercial purposes can easily result in patients refusing to participate.[150] Accordingly, data linkage projects must reflect the fact that they are ultimately designed to ensure the quality and effectiveness of public health services. There are a number of strategies that can be employed in this context to improve patient trust and reciprocity. The first is the use of dynamic or granular consent, where patients can opt in and out of certain research projects. Although dynamic consent has been considered in-depth within the legal and bioethical literature, Estonia has introduced this mechanism into their national electronic health record system. Via their web accessible record, patients are constantly notified as to whether they wish to continue participating in particular projects.[151] A second important strategy is the use of patient education and awareness campaigns. Robertson et al. noted that expert interviewees identified the lack of public awareness about the *care.data* program as a key reason the program failed (compared with the WADLS).[152] Furthermore, as Gille, Smith, and Mays note, public trust derives from a number of sources, including the existence of regulatory systems to enforce rights. Therefore, any public awareness campaign must be designed so that patients are aware of the ability to enforce their rights, as well as the permissible ethical uses of their data.[153]

Contractual Solutions

Another strategy that has been increasingly used to promote data accessibility, particularly for international research projects, involves contractual solutions to support research and processing. In particular, standard form agreements for data use and transfer help demonstrate satisfactory protection for data transferred between organizations, as well as data transferred across borders. For example, the Swiss, EU, Italian, Spanish, UK, Australian, and some Canadian provincial legislation reference agreements for the transfer of data. These agreements set minimum standards of security and processing that all parties involved in data linkage must comply with. For countries with more flexible data processing standards, such as Australia, contracts can provide more explicit guidance to researchers in those jurisdictions as to how to meet their obligations. Finally, the existence of contracts for processing further reinforces the social license required from members of the public to use their data for processing. This question of public license also raises the related question of data ownership. Although not explicitly explored within the previous comparison of legislative frameworks, collections of health information also involve numerous stakeholders who each hold specific rights over their data. Specifically, any solution requires considering not only data protection laws and research ethics regulations but also intellectual property laws (such as copyright), and data use or material transfer agreements (MTAs).[154] In the USA, there is no

150 Carter, Laurie, & Dixon-Woods, *supra* note 10, at 407.

151 Jaan Priisalu & Rain Ottis, *Personal Control of Privacy and Data: Estonian Experience*, 7 Health Technol. 441–451, 449 (2017).

152 Ann R. R. Robertson et al., *Tightrope Walking Towards Maximising Secondary Uses of Digitised Health Data: A Qualitative Study*, 23 J. Innov. Health Inform. 591–599, 597 (2016).

153 Felix Gille, Sarah Smith & Nicholas Mays, *What is Public Trust in the Healthcare System? A New Conceptual Framework Developed From Qualitative Data in England*, Soc. Theory Health, 6 (2020), https://doi.org/10.1057/s41285-020-00129-x (accessed Mar 3, 2020).

154 Jane Reichel, *Oversight of EU Medical Data Transfers—An Administrative Law Perspective on Cross-Border Biomedical Research Administration*, 7 Health Technol. 389–400, 397 (2017).

copyright protection for raw data, but only products of data.[155] Similar limitations exist on copyright protection for compilations of data in Australia, although Canadian courts have acknowledged copyright for data compilations.[156] In addition, although raw data are unlikely to receive copyright protection, copyright could extend to curated patient records and associated notes.[157]

To resolve these differences in intellectual property protection, standardized contractual mechanisms can be used to determine the rights held by each participating stakeholder. In biomedical research, MTAs are frequently used to govern the exchange of human tissue and data between institutions and to ensure provenance.[158] Therefore, similar agreements have been attempted for data transfer and use agreements (DTUA). However, designing a uniform DTUA carries a number of additional challenges beyond a uniform MTA. The first is resolving the conflict between different data protection and privacy regimes. The analysis above demonstrates that data can only be transferred to jurisdictions where adequate protection is supplied. Therefore, any MTA or DTUA must ensure that arrangements for data sharing are ethically robust and legally compliant.[159] However, one complicating factor is additional intellectual property rights that may vest in data. In the EU (but not Switzerland), copyright protection is further complemented by the *sui generis* regime for databases.[160] Nevertheless, this right only protects the non-technical structure of the database. Furthermore, the Database Directive prevents rights holders from introducing specific contractual terms that would fall outside the exceptions to restricted acts.[161] The effect of this provision is that database rights may act to prevent standardizing contractual terms. Accordingly, when developing collaboration agreements, institutions should consider all applicable intellectual property rights.

Technical Solutions

The previous contractual and organizational measures can support data sharing, but alone can neither entirely prevent intentional or unintentional re-identification. Instead, these solutions must be paired with technical solutions supporting data linkage and data reuse in multisite research while technically minimizing the risk of re-identification. In the example of WADLS above, linkage was achieved by matching together each dataset probabilistically and by replacing all personal details with a unique link ID. These files were created via isolated secure computer systems, with the technicians responsible for linkage not allowed to participate in data analysis and personal demographic data being destroyed afterwards. Furthermore, as

155 *Feist Publications v Rural Telephone Service Co* 499 US 340 (1991).

156 *IceTV Pty Ltd v Nine Network Australia Pty Ltd* (2009); *CCH Canadian Limited v Law Society of Upper Canada* (2002) 4 FC 213.

157 Amanda Levendowski, *How Copyright Law Can Fix Artificial Intelligence's Implicit Bias Problem*, 93 Wash. Law Rev. 579–630, 608 (2018).

158 Dianne Nicol & Jane L. Nielsen, *Patents and Medical Biotechnology: An Empirical Analysis of Issues Facing the Australian Industry* 163 (2003), https://papers.ssrn.com/abstract=2583508.

159 Deborah Mascalzoni et al., *International Charter of Principles for Sharing Bio-Specimens and Data*, 23 Eur. J. Hum. Genet. 721–728, 721 (2015).

160 Directive 96/9/EC of the European Parliament and of the Council on the legal protection of databases (Mar. 11, 1996) [hereafter referred to as the Database Directive].

161 Database Directive, *supra* note 168, Art. 15.

mentioned previously, traditional de-identification and pseudonymization techniques (implemented as the removal of common identifiers) are not enough to properly minimize the re-identification risk. This risk is particularly heightened when dealing with clinical and omics data.[162] However, advanced cryptographic privacy-preserving solutions have the potential to play a major role in the significant reduction of the risk of re-identification. The DPPH project specifically focuses on the practical application of these technologies in medical research. The purpose of cryptography historically has been to hide the content of information by transforming it into ciphertexts through a hard-to-invert or one-way function. However, the greatest challenge with encryption technology reflects that discussed by Ohm regarding anonymization, that is, ensuring anonymized data is both safe and useful.

Traditional encryption is profusely used nowadays to protect data in transit and at rest. Data in transit can be encrypted being communicated between two end-points, protecting it against eavesdroppers. Data at rest can be encrypted while stored in an untrusted environment, to protect it against unauthorized actors. Nevertheless, data reuse for research goes beyond communication and storage, involving a significant re-identification risk during computation, in which traditional encryption cannot be mitigated. For this purpose, more advanced cryptographic techniques can enable the computation of some functions on data without disclosing these data to the processor. These techniques are aligned with the minimization and purpose principles. Specifically, the technology makes it unnecessary to provide the recipient with access to the data, but only to the computation results needed for the established purpose. These techniques can be classified into software- and hardware-based solutions. Software-based solutions comprise homomorphic encryption and secure multi-party computation (SMPC). Hardware-based solutions comprise trusted execution environments (enclaves) where data are decrypted inside a tamper-proof enclave and processed in the clear only inside the enclave. In this section, we focus on software-based solutions. Recent advances in fully homomorphic encryption,[163] which can handle an arbitrary number of encrypted operations, are promising. Unfortunately, their computational overhead is still far from practical.[164] However, the so-called Somewhat Homomorphic Encryption (SHE) or Practical Homomorphic Encryption can efficiently enable a limited set of operations on encrypted data without the need to decrypt them first. This technology protects the data from an untrusted processor (or from attacks happening at the processor premises). Nevertheless, the versatility of these algorithms is limited, and they have to be paired with other techniques in order to develop fully functional end-to-end systems. Conversely, SMPC enables several parties to evaluate a function on private data coming from distinct data sources without aggregating or sharing the input data.[165] At the end of the protocol, the parties learn nothing more but the value

162 Lin, Owen, & Altman, *supra* note 22; Homer et al., *supra* note 22; Gymrek et al., *supra* note 22; Emam et al., *supra* note 23.

163 Ilaria Chillotti et al., *Faster Fully Homomorphic Encryption: Bootstrapping in Less Than 0.1 Seconds*, *in* Advances in Cryptology—ASIACRYPT 2016 3–33 (Jung Hee Cheon & Tsuyoshi Takagi eds, 2016).

164 Andreas Wiebe & Nico Schur, *Protection of Trade Secrets in a Data-Driven, Networked Environment—Is the Update Already Out-Dated?* 14 J. Intellect. Prop. Law Pract. 814–821, 820 (2019).

165 Ronald Cramer, Ivan Bjerre Damgrd & Jesper Buus Nielsen, Secure Multiparty Computation and Secret Sharing (1st ed. 2015).

of the function.[166] Various privacy-preserving supervised machine learning algorithms have been proposed and analyzed in the SMPC setting.[167] In addition, both training and prediction algorithms for deep learning have been developed.[168]

Crucially, the combination of the aforementioned techniques may provide the means to allow computation without needing to transfer data between jurisdictions.[169] These cryptographic techniques minimize the re-identification risk for data at rest, in transit, and during computation. However, the re-identification risk is not decreased for disclosed computation outputs that are disseminated for research, and sufficient information is available to identify the individual. In other words, the same legal and ethical limits that apply to the exchange of data will continue to apply to these research results. Specifically, repeated query requests to these data can be used to conduct inference attacks against individual records. However, statistical protection techniques, such as k-anonymity,[170] t-closeness,[171] and, in particular, differential privacy,[172] can be used to address this issue. These techniques operate by minimizing the re-identification risk based on mathematical guarantees at the cost of a controlled reduction on data utility. Concrete examples of these technologies in multi-site data sharing scenarios have been proposed to enable queries on a set of independent databases while protecting privacy and confidentiality. These solutions also aim at protecting the privacy of each individual storing its data in these shared databases.[173] To compare genomic data among different patients in an honest-but-curious model, one secure two-party computation method involves patients encoding their genomic variants. The system ensures that query outputs only some statistical information about the queried variants and nothing else about the patients.[174] Recently, several frameworks[175] have been described which use homomorphic encryption or multiparty computation in a decentralized way. These frameworks enable statistical queries on distributed databases while avoiding single

166 Andrew C. Yao, *Protocols for Secure Computations*, *in* 23rd Annual Symposium on Foundations of Computer Science (SFCS 1982) 160–164 (1982).

167 Valeria Nikolaenko et al., *Privacy-Preserving Ridge Regression on Hundreds of Millions of Records*, *in* 2013 IEEE Symposium on Security and Privacy 334–348 (2013).

168 Payman Mohassel & Yupeng Zhang, *SecureML: A System for Scalable Privacy-Preserving Machine Learning*, *in* 2017 IEEE Symposium on Security and Privacy (SP) 19–38 (2017).

169 Gerald Spindler & Philipp Schmechel, *Personal Data and Encryption in the European General Data Protection Regulation*, J. Intellect. Prop. Inf. Technol. Electron. Commer. Law [i]-177, 166 (2016).

170 Latanya Sweeney, *k-Anonymity: A Model for Protecting Privacy*, 10 Int. J. Uncertain. Fuzziness Knowl.-Based Syst. 557–570 (2002).

171 Ninghui Li, Tiancheng Li & Suresh Venkatasubramanian, *t-Closeness: Privacy Beyond k-Anonymity and l-Diversity*, *in* 2007 IEEE 23rd International Conference on Data Engineering 106–115 (2007).

172 Cynthia Dwork, *Differential Privacy*, *in* Automata, Languages and Programming 1–12 (Michele Bugliesi et al. eds, 2006).

173 Johes Bater et al., *SMCQL: Secure Querying for Federated Databases*, 10 Proc VLDB Endow 673–684 (2017).

174 Karthik A. Jagadeesh et al., *Deriving Genomic Diagnoses Without Revealing Patient Genomes*, 357 Science 692–695 (2017).

175 David Froelicher et al., *UnLynx: A Decentralized System for Privacy-Conscious Data Sharing*, 2017 Proc. Priv. Enhancing Technol. 232–250 (2017); David Froelicher et al., *Drynx: Decentralized, Secure, Verifiable System for Statistical Queries and Machine Learning on Distributed Datasets*, ArXiv190203785 Cs (2019), http://arxiv.org/abs/1902.03785 (accessed Jan. 3, 2020); Henry Corrigan-Gibbs & Dan Boneh, *Prio: Private, Robust, and Scalable Computation of Aggregate Statistics* 259–282 (2017), https://www.usenix.org/conference/nsdi17/technical-sessions/presentation/corrigan-gibbs (accessed Nov. 22, 2019); Wenting Zheng et al., *Helen: Maliciously Secure Coopetitive Learning for Linear Models*, *in* 2019 IEEE Symposium on Security and Privacy (SP) 724–738 (2019).

points of failure for the security and privacy guarantees. Within the DPPH project, we have also developed an operational system, called MedCo.[176] MedCo protects the confidentiality of the patients while enabling very efficient data discovery and feasibility analyses on clinical and genomic data. Our system achieves this by combining homomorphic encryption, multiparty computation, and differential privacy. At the time of writing, this system is being tested at the Swiss university hospitals as a pilot deployment.[177]

CONCLUSION

This paper highlights the key regulatory jurisdictional divergences in data protection and research ethics laws for health-related data through the lens of data accessibility. Specifically, we identify three key differences between Switzerland, the GDPR (and its national implementation in Italy, Spain, and the UK), the USA, Canada, and Australia. The first key difference concerns inconsistent and technically vague definitions of anonymization, pseudonymization, and de-identification. As a consequence, these definitions obfuscate whether anonymization and de-identification can be satisfied algorithmically or whether changing the environment where the data are stored is also necessary. Furthermore, it is unclear whether encrypted data constitute their own category of data or should be assigned to one of the existing categories. The second key difference concerns the potential for secondary uses of data beyond that for which consent has been obtained. In particular, the GDPR places significant limitations on the use of data for secondary purposes. National implementations of the GDPR explored in this paper permit secondary processing of genetic and other health-related data. Nevertheless, there remain outstanding questions whether these laws comply with the GDPR's consent requirements and whether they will create further regulatory fragmentation in the EU. The third key difference dovetails into this issue and concerns the requirements for transfer of health-related data between jurisdictions. With the exception of transfers within the EU and transfers between the EU and Switzerland, there are conditions on the exchange of data between jurisdictions. For transfers to the USA and Canada, transfer on the grounds of adequacy is limited to private companies. Accordingly, for transfers from the EU to public research institutes in the USA or Canada, either binding contractual and corporate rules or standard data protection clauses are required. Furthermore, these safeguards are required for transfers to and from Australia to all other jurisdictions considered in our paper.

Therefore, in this paper we suggest a number of organizational, contractual approaches which we believe are necessary to increase data accessibility for multisite research consortia. Organizational approaches include clarifying the roles of each participant in the consortia, prioritizing ethics requirements prior to research, and obtaining patient trust for research. Contractual approaches include continuing to utilize and improve standardized DTUA for the exchange of data between institutions. These agreements can provide explicit and uniform protection for patient rights across jurisdictions while clearly allocating intellectual property rights between institutions.

176 Jean Louis Raisaro et al., *MedCo: Enabling Secure and Privacy-Preserving Exploration of Distributed Clinical and Genomic Data*, 16 IEEE/ACM Trans. Comput. Biol. Bioinform. 1328–1341 (2019).

177 https://medco.epfl.ch.

Nevertheless, we concede that these organizational and contractual solutions alone are not sufficient to entirely protect against the risk of re-identification. Accordingly, in this paper, we propose a number of software-based technologies to protect the privacy of patient data while ensuring data accessibility. In particular, these technologies include variants of homomorphic encryption, such as fully and SHE, and secure multiparty computation. Crucially, these solutions diverge from traditional encryption in that they extend beyond protecting data at rest and in transit, and instead only disclose the results of computation. This protection is achieved either by encrypting the inputs of computation or ensuring that each party can process data without the need for access to all the inputs. To decrease the risk of potential re-identification from results, secure computation can be coupled with further privacy guarantees such as differential privacy. Future research should explore, in-depth, the compatibility between these advanced privacy preserving technologies and the existing ethical and legal frameworks, as well as future standards and soft law. Finally, any use of these technologies must be backed by an appropriate education and public awareness campaign. The technological solutions that have been proposed in this paper are conceptually complicated, increasing the difficulty of explaining the benefits of this technology. Therefore, any public education campaign should be carefully crafted so as to reflect the computer literacy of members of the general public. Furthermore, this education campaign should be framed in such a way as explain to users how to exercise their rights with respect to the technology in question.