Vol. 28
N° 01

Lex Electronica

2023

CENTRE
DE RECHERCHE
EN DROIT
PUBLIC

Palmira Granados Moreno, Hanshi Liu, Sebastian Ballesteros Ramirez, David Bujold, Ksenia Zaytseva, Guillaume Bourque et Yann Joly

D-PATH (Data Privacy Assessment Tool For Health) for Biomedical Data Sharing

D-PATH (DATA PRIVACY ASSESSMENT TOOL FOR HEALTH) FOR BIOMEDICAL DATA SHARING

Palmira Granados MORENO[1], Hanshi LIU[2], Sebastian Ballesteros RAMIREZ[3], David BUJOLD[4], Ksenia ZAYTSEVA[5], Guillaume BOURQUE[6], Yann JOLY[7]

1 Department of Human Genetics, Faculty of Medicine and Health Sciences, McGill University, Montreal, QC, Canada; Centre of Genomics and Policy, McGill University, Montreal, QC, Canada (premier auteur)

2 Department of Human Genetics, Faculty of Medicine and Health Sciences, McGill University, Montreal, QC, Canada; Centre of Genomics and Policy, McGill University, Montreal, QC, Canada (premier auteur)

3 Department of Human Genetics, Faculty of Medicine and Health Sciences, McGill University, Montreal, QC, Canada, Canadian Center for Computational Genomics, McGill University, Montreal, QC, Canada

4 Department of Human Genetics, Faculty of Medicine and Health Sciences, McGill University, Montreal, QC, Canada, Canadian Center for Computational Genomics, McGill University, Montreal, QC, Canada

5 Department of Human Genetics, Faculty of Medicine and Health Sciences, McGill University, Montreal, QC, Canada, Canadian Center for Computational Genomics, McGill University, Montreal, QC, Canada

6 Department of Human Genetics, Faculty of Medicine and Health Sciences, McGill University, Montreal, QC, Canada, Canadian Center for Computational Genomics, McGill University, Montreal, QC, Canada

7 Department of Human Genetics, Faculty of Medicine and Health Sciences, McGill University, Montreal, QC, Canada; Centre of Genomics and Policy, McGill University, Montreal, QC, Canada

ABSTRACT

The Data Privacy Assessment Tool for Health (D-PATH) is a proof-of-concept online tool designed to help users intending to share biomedical data identify applicable legal obligations and relevant best practices. D-PATH provides a series of simple questions to assess important aspects of the data sharing task, such as the user's legal jurisdiction and the types of entities involved. Based on the combination of answers that the user provides, D-PATH will generate a list of privacy obligations and security-best practices, categorized into themes of 1) accountability, 2) lawfulness of storage, transfer, and protection, and 3) security and safeguards that will likely apply in the user's scenario. Currently, the D-PATH focuses on Canadian and European privacy laws and various global best-practice policies, but there are plans to extend this in later iterations of the tool. D-PATH was developed specifically to inform users about their legal privacy obligations and best practices and was written to facilitate compliant and ethical data sharing. As a proof-of-concept, D-PATH demonstrates the potential value of a tool in simplifying and translating complex concepts into more accessible formats. Such a tool can be adapted and valuable in many different contexts, such as training core researchers in data sharing laws and practices.

RÉSUMÉ

L'outil d'évaluation de la confidentialité des données dans le domaine de la santé (D-PATH) est un outil de preuve de concept en ligne conçu pour aider les utilisateurs, ayant l'intention de partager des données biomédicales, à identifier les obligations juridiques applicables et les meilleures pratiques pertinentes. D-PATH propose une série de questions simples afin d'évaluer les aspects importants du partage de données, comme la juridiction de l'utilisateur et les types d'entités concernées. En fonction de la combinaison de réponses fournies par l'utilisateur, D-PATH génère une liste d'obligations relatives à la protection de la vie privée et de pratiques exemplaires en matière de sécurité, classées selon les thèmes suivants : 1) responsabilité, 2) légalité du stockage, du transfert et de la protection, et 3) sécurités et mesures de protection qui s'appliqueront vraisemblablement au cas de l'utilisateur. Présentement, D-PATH se concentre sur les lois canadiennes et européennes en matière de protection de la vie privée, ainsi que sur diverses politiques mondiales de pratiques exemplaires, mais il est prévu d'étendre sa zone d'application dans les versions ultérieures de l'outil. D-PATH a été rédigé pour faciliter un partage des données conforme aux normes juridiques, éthiques et aux pratiques exemplaires. En tant que preuve de concept, D-PATH démontre la valeur potentielle d'un outil pour simplifier des recherches complexes dans des formats plus accessibles. Un tel outil peut être utilisé dans de nombreux contextes, incluant celui de la formation des chercheurs aux lois et pratiques exemplaires du domaine du partage des données.

INTRODUCTION

[1] The scientific and technological progress of the past few decades has created a reality where data generation is at an unprecedented rate. In this context, it is no surprise that biomedical research is increasingly data-centric (Leonelli, 2016, pp. 13-20 ; Kaye, 2012, pp. 415-431). Biomedical research is also more globalized, with research increasingly being conducted internationally and more reflective of human diversity (Middleton et al., 2020, pp. 743-752 ; Gurdasani et al., 2020, pp. 184-186). The volumes of data generated have the potential to help discover more effective ways to diagnose, manage, treat, and even prevent diseases. Under these circumstances, the advancement of biomedical and health research requires policies and practices that allow researchers and companies to share their data responsibly (Kalaitzopoulos, Patel & Younesi, 2016, p. 36).

[2] Responsible data sharing elicits many benefits. Data sharing has the potential to significantly strengthen academic research, clinical medicine, and public health as it can promote more transparent practices, facilitate collaboration, enable data reuse, reduce effort duplication, contribute to reproducibility and replicability, inform public health policies, reduce research timelines, promote diversity in research, and overall further research discovery (Stark et al., 2020 ; Chawinga & Zinn, 2019, pp. 109-122; Levenstein & J Lyle, 2018, pp. 95–103 ; LeBel, Campbell & Loving, 2017, pp. 230-243 ; Hulsen, 2020 ; Fidler & Wilcox, 2018 ; Granados Moreno, Joly & Roskams-Edris, 2020, pp. 119–180). Likewise, data sharing is conducive to realizing the human right to science and is considered a social good (Knoppers & Beauvais, 2021). Indeed, when data sharing policies enable more universal access to research resources (including raw data), they allow a more equitable circulation of the benefits of science and a more equal distribution of scientific progress (Knoppers & Beauvais, 2021 ; Harris & Wyndham, 2015, pp. 334–337 ; Knoppers et al., 2014, pp. 895–903).

[3] Ultimately, many benefits of data sharing can depend on the research participants' willingness to provide their data for broad sharing. This trust depends on the research community to show that data is managed, shared, and used responsibly. Responsible data sharing requires respecting the interests and fundamental human rights of those whose data are shared. Studies show significant public concern about the potential for unauthorized data access and misuse (Trinidad, Platt & Kardia, 2020, pp. 1-10). Accordingly, entities such as genomic databases abide by robust privacy and security policies and mechanisms to protect the personal information of participants. Data privacy laws and ethics guidelines also impose this obligation on data processors. However, privacy can be a very complex matter because a plethora of intricate norms and agreements at the international, regional, and national levels govern it (Saulnier et al., 2019). Although there may appear to be an irreconcilable tension between participant rights (such as privacy) and facilitating broad accessibility in data sharing, this is not the case. Data sharing and privacy are not mutually exclusive endeavours. For example, one mechanism that allows for a favourable balance is controlled access (Joly et al., 2016, pp. 1150–1154). In controlled access, researchers can submit data access requests to a Data Access Committee (DAC). The DAC may then verify the researcher's credentials and the proposed project protocol, and formally impose limits on how the data may be used. Leveraging such mechanisms, data consortiums such as

the International Cancer Genomic Consortium (ICGC) have found continued success in facilitating responsible data sharing (Joly et al., 2012).

[4] Although there are channels for balancing data sharing and privacy, the legal, ethical, and policy aspects of biomedical data sharing are not always apparent. Navigating through the applicable regulatory systems is nontrivial. With this in focus, the Data Privacy Assessment Tool for Health (D-PATH) was devised to help data processors[8] to navigate through complex data privacy regulatory systems and provide them with a concise, accessible overview of the applicable obligations and standards that they need to meet to engage in responsible data sharing (see Figure 1 for a diagram of laws and best practices considered by the tool.). This article outlines the motivations that led to the development of D-PATH, its objectives, features, and our plans for future developments.

THE NEED FOR PRIVACY-COMPLIANT DATA SHARING

[5] While there is a growing agreement on the beneficial impact of FAIR (Findable, Accessible, Interoperable and Reusable) data sharing on scientific progress and the human right to science, the importance of protecting the privacy of those whose data is being shared is undeniable, and as a fundamental right, it needs to be upheld and emphasized (Health Canada, 2006). Privacy protection contributes to safeguarding people's autonomy and dignity, and it helps prevent the misuse of personal information, thus, protecting people against undue exclusion or interference (Knoppers & Beauvais, 2020, pp.454-457 ; McFarland, 2012). This is particularly pressing when the shared data is health-related and genetic (Knoppers & Beauvais, 2020 ; Knoppers & Beauvais, 2021). Countries have enacted privacy[9] laws and policies to regulate the sharing of personal data to protect individuals' privacy (Office of the Privacy Commissioner of Canada, 2018 ; European Parliament, 2016). Institutionally, different policies have also been created to promote research practices that protect data privacy[10]. Compliance with these practices, mechanisms, and policies enables the safe and legally compliant use of data (Stark et al., 2019). Privacy also serves an essential role in data sharing by contributing to participants trust in the research projects sharing data (Kaye, 2012 ; Health Canada, 2006 ; Kalaitzopoulos, Patel & Younesi, 2016).

[6] Despite the critical role that existing privacy laws and policies play in promoting responsible and safe data sharing, the process can sometimes hinder data sharing practices (Kalaitzopoulos, Patel & Younesi, 2016). Compliance with all relevant privacy laws is complex, particularly when data is shared across national and international borders, as these laws vary with each jurisdiction (Chawinga & Zinn, 2019). A similar situation can occur with respect to institutional and funding agency policies that may impose obligations regarding data sharing, data security, and privacy (*Idid.*) On the one hand, there can be confusion and uncertainty on which policies to follow and how to

8 For purposes of this article, the term data processor refers to an individual, entity or organization that either personally or through a project collects, records, organizes, structures, stores, adapts or alters, retrieves, consults, uses, discloses, transmits, disseminates, or otherwise makes available, aligns, combines, restricts, erases, or destroys data.

9 While some jurisdictions distinguish between Privacy and Data Protection, data protection is generally seen as falling within the broader concept of privacy. For the purposes of this paper, in the context of legislation, the broader term of privacy will be used to describe both privacy and data protection laws.

10 These various laws and policies can vary widely in their scope and content. For example, some of these laws and policies specifically reference using technological tools such as firewalls, encryption, and data de-identification to better protect individual privacy (e.g., see the GA4GH Data Security Infrastructure Policy (2019)

achieve coherence and interoperability among them. In other cases, the effect can be even more serious, as data sharing and data privacy policies can seem contradictory. Such a lack of uniformity can create compliance challenges, as researchers can encounter considerable uncertainty concerning which path they should take (Saulnier et al., 2019).

[7] Bearing this in mind, more work needs to be done to simplify and harmonize global data governance frameworks. It is also possible to ameliorate these data sharing challenges by developing computational tools that guide researchers and data processors, helping them navigate the privacy laws across systems and countries. This is the central purpose behind D-PATH's development as a proof-of-concept tool.

RESULTS

DATA PRIVACY ASSESSMENT TOOL FOR HEALTH (D-PATH)

[8] The Data Privacy Assessment Tool for Health (D-PATH) is a first-of-its-kind, proof-of-concept, online tool whose purpose is to facilitate data sharing activities in the context of biomedical and health research to meet the applicable ethical, legal, and professional requirements associated with privacy. While not equivalent to formal legal advice, D-PATH aims explicitly to assist researchers, data hosts, service providers, and other relevant stakeholders in protecting the privacy of the health-related[11] (National Cancer Institute, 2020) datasets they process in a responsible, accessible, and compliant manner.

[9] D-PATH was developed in the context of the EpiShare project, based in Montreal, Canada. EpiShare is working on a web-based platform to make epigenetic data[12] more easily discoverable and accessible. The project started as a collaboration with the International Human Epigenome Consortium (IHEC) and the Encyclopedia of DNA Elements (ENCODE) and was selected as a collaborative GA4GH driver project (EpiShare, 2019). Within EpiShare, datasets are processed to generate searchable metadata on epigenomic features. Each EpiShare implementation (repository of epigenomic data) shares metadata on available datasets following GA4GH metadata specifications. It allows for the exploration of genome/epigenome interactions by showing the effect of specific genomic variants on a set of epigenetic experiments, such as RNA-Seq, ChIP-Seq and ATAC-Seq (Bourque & Joly, 2017). While the process described here makes epigenomic data more efficient and secure, it remains essential to carefully assess the privacy and confidentiality implications of this innovative data sharing process.

[10] D-PATH was devised to address these types of concerns by enabling responsible data-sharing practices to access and visualize large epigenomics datasets and launch

11 While D-PATH and EpiShare handle both, epigenetic and epigenomic data and, therefore, we use these terms interchangeably, it is important to be aware of their specific focus. Epigenetics "focuses on processes that regulate how and when certain genes are turned on and turned off; while epigenomics pertains to the analysis of epigenetic changes across many genes in a cell or entire organism."

12 Epigenetics is the study of reversible modifications on the genetic material of cells, affecting gene expressions mechanisms. These modifications are partly inherited and partly imputable to environment and life habits.

multi-omics analyses in those datasets in the EpiShare Portal and potentially other similar platforms (*Ibid*.) Given the context in which D-PATH was created, the initial version of the tool focuses primarily on Canadian law and, more specifically, on the privacy laws of the province of Quebec. However, positive comments received during the development and pilot phases led us to broaden the scope of the proof-of-concept tool to encompass more of Canada's privacy laws, best practices, and some key European and U.S. norms[13]. D-PATH is built based on a decision tree that navigates through all these complex conditions, scenarios, and exceptions. Please see Figure 4 regarding how D-PATH classifies the information in question based on the user's inputs. Please also see Supplementary Table 2 to reinforce the logic of the information classification being used throughout D-PATH in the Canadian legislative context.

A CONVIVIAL USER EXPERIENCE

[11] D-PATH begins by asking users to respond to a first set of simple, lay format, queries about the type of activities they engage in with respect to the data in question. For instance, the tool asks about the relationship the D-PATH user has with the data: whether it is an individual, entity or organization that, either personally or through a project, collects, records, organizes, structures, stores, uses, discloses, makes available or destroys (jointly referred to as processing[14]) data and their role (data steward/user) or whether they are individuals who have contributed their own personal health-related data to a project (data donor)._Depending on their answers, users will be firstly categorized as data stewards/users[15] data donors[16]. Those who fall in the data stewards/users category can continue using the tool to know their main privacy and data protection-related obligations and best practices. Others will be re-directed toward general resources more appropriate to their situation.

[12] The next set of questions focuses on 1) the country or region where the project/ organization is established, 2) whether the study monitors the behaviour of individuals, and if that is the case, the location of those individuals whose behaviour is monitored, and 3) whether the study returns individual results, and if that is the case, the location of those individuals whose results are being returned. The answers to these questions provisionally determine the possible applicable jurisdiction(s). The tool then asks an additional set of queries to assess the nature of the data being processed, which is determined by the identifiability level of the data. These additional queries aim to determine whether the processed data is considered personal information[17]. With these answers, the tool will determine whether privacy laws, in general, apply to the data. These two sets of questions (i.e., regarding geographical locations and nature of the data) are designed to establish the first level of applicable jurisdiction to determine the

13 See Figure 1 for an overview list of legislation considered in the D-PATH tool.

14 The term "processing" has two slightly different meanings depending on whether it is in the context of Canada or Europe. Processing of personal and personal health information in the Canadian context comprises collection, storage, use, disclosure, and/or communication. The European GDPR defines processing as "any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction." D-PATH aligns with the latter.

15 Within D-PATH, data steward/user refers to an individual, entity or organization that either personally or through a project collects, records, organizes, structures, stores, adapts or alters, retrieves, consults, uses, discloses, transmits, disseminates, or otherwise makes available, aligns, combines, restricts, erases, or destroys data.

16 Data donor or data subject refers to the individual whom the data is about. In other words, it is the research or study participant.

17 Determining whether data is personal or not can sometimes be a very complex task. The limitations of this are further discussed in the Areas of future improvements section.

legislation relevant to D-PATH's users. At this level, D-PATH gives two types of answers. It determines whether the data being processed is personal information and therefore protected by privacy laws[18]. It also provides the first set of responses regarding the applicable jurisdictions.

[13] In the cases where the applicable jurisdiction is Canada, D-PATH continues with further questions about the institution to which the steward/user is affiliated. The steward/user can be affiliated, for example, with the Canadian federal government, the provincial governmental institution, a private organization, or it could be an individual in the course of commercial activities, a federally regulated business, a member of the MUSH sector (Municipality, University, School, or Hospital) or a healthcare professional. D-PATH then asks the specific province within Canada in which the steward/user's institution is located. Depending on the type of institution to which the steward/user is affiliated, as well as the province selected, D-PATH provides a much more detailed enumeration of the applicable legislation. The responses displayed by D-PATH at this level identify the specific law(s) applicable to the specific situation its users describe. In addition to the names of the applicable privacy laws and the list of legal privacy obligations, D-PATH provides a list of recommended measures or actions compliant with international best practices or standards on privacy and data protection. While compliance with these practices is not legally required, adherence to them can facilitate interoperability and collaboration among projects and can also guide data stewards/users on how to process health-related information in general responsibly (GA4GH, 2014).

[14] D-PATH's legal privacy obligations and recommended measures are organized in sections: a) laws organized by jurisdiction and b) best practices. The obligation and measures are classified into three broad categories: 1) Accountability, 2) Lawfulness of use, storage, transfer, and protection, and 3) Security and safeguards[19]. The categories in each tab aim to help D-PATH's users understand where specific duties fall within the more general categories of obligations. For this same reason, the obligations are drafted in lay terms. Each obligation listed includes references to the documents or laws requiring it. The three categories were chosen for their intuitiveness and simplicity while being sufficiently broad to describe the various obligations from several different privacy and data protection laws and best practices. These categories also aim to group similar obligations together based on content and meaning regardless of their differing jurisdictions. For example, under Accountability, the relevant Quebec laws may require the data-sharing organization to identify and designate a person responsible for complying with said law. In content, this resembles the GDPR's requirement of having a Data Protection Officer to inform, monitor, and advise on data protection obligations. The two are similar in content and grouped together for intuitiveness.

18 Again, the exact point of when data becomes identifiable and therefore personal is a highly complex and contextual situation. There are various approaches for considering this topic from differential privacy and statistical approaches to jurisprudence. This a meaningful dissection of the topic is beyond the scope of the paper. This will be further expanded upon in the Areas of future improvements section.

19 See Table 1 for screenshots of some of the obligations and recommended measures displayed by D-PATH in compliance with Canadian-Quebec privacy laws, EU GDPR, and best practices as an example.

[15] The initial version is focused on Canada and Quebec data privacy laws. Consequently, when the applicable jurisdiction is Quebec, the answer displayed includes a precise list of privacy obligations. However, when the applicable jurisdiction is a Canadian province other than Quebec, D-PATH displays the specific law (not the specific legal obligations) applicable to the situation described in the users' answers. Similarly, when the applicable jurisdiction is the United States or the European Union[20] (European Commission, 2021). D-PATH displays the name of the main applicable privacy laws in those jurisdictions. However, the answer displayed in these cases does not include a list of specific privacy obligations, as it does when the jurisdiction is Quebec, Canada. D-PATH also provides a general list of the main privacy obligations set out in the General Data Protection Regulation (GDPR)[21]. Finally, when the applicable jurisdiction is other than Canada, the European Union, or the United States, D-PATH simply suggests consulting the privacy legislation of that specific country without providing any specific details about the applicable law or privacy obligations. In the future, D-PATH can be expanded to provide increased jurisdictional coverage.

METHODS

[16] D-PATH's interface is written in JavaScript along with the React Library, which is a component-based library used for front-end development, the part of the system that the user interacts within the browser. The main functionality of the tool is implemented as a decision tree. For example, in Figures 2, 3 and 4, the applicable law(s) depends on the users' input regarding the description of information. D-PATH's tree consists of two main components: 1) the geographical location of the data and the individual(s) associated with the data in question and 2) whether the data is personal information. As an application, D-PATH is relatively simple since it is mainly built with React without a back end, meaning it has neither a database nor a server. The application comprises a landing page, forms working in tandem to implement the decision tree, and a page with the final assessment. D-PATH's source code is available at https://github.com/c3g/d-path under the free and open-source software license Apache 2.0 license (GNU, 2010). Furthermore, please see Figure 5 regarding the logic used in the decision-making tree.

The online version of the tool can be found at: https://www.computationalgenomics.ca/tools/d-path

DISCUSSION

SIMILAR TOOLS AND CHALLENGES

[17] Some existing tools share some similarities with D-PATH; however, none allow for a comparative analysis of legal and policy requirements required for data sharing at the level of granularity provided for by D-PATH. For example, DAISY is a software tool that

20 While we use the term European Union for purposes of convenience, technically, we are referring to the European Economic Area, which includes the countries of the European Union and three countries of the European Free Trade Association (Iceland, Liechtenstein, and Norway).

21 Given the extraterritorial applicability of the European GDPR, it may apply to Canadian D-PATH users who work in projects located in the European Union, who monitor the behaviour of individuals located in the European Union, or who return individual results to people located in the European Union.

facilitates compliance with the GDPR accountability requirement (Regina Becker et al., 2019) while the Data Stewardship Wizard (DSW) proposes a dynamic web forms system to help researchers compose data management plans (DMPs) that also meet FAIR requirements (DSW, 2020). Another such tool, the Covered Entity Guidance tool (CMSgov, 2020), was created to help organizations or individuals determine if they are a "covered entity"[22] under the Health Insurance Portability and Accountability Act (HIPAA) and therefore obligated to comply with the Act (OHSU, 2020). As can be gathered from their description, many of these tools are also jurisdiction-specific and of limited use for international data sharing.

[18] The principal challenge encountered in the development of D-PATH arises from the complexity of legal privacy systems, both nationally and internationally, as well as the globalized nature of data-intensive research. In Canada alone, there are over 20 federal and provincial privacy laws, characterized by unique applicability rules, regulated subjects, and provisions. The intricate process of documenting the content of these laws for the efficient organization through a decision tree is a time-consuming endeavour.

[19] Moreover, certain legal nuances are difficult to capture and represent within the tool's format. One notable example is the distinction between information deemed "public information" and information merely "publicly accessible." Legally, there is a significant difference between the two, such that truly public information can be used with few or very few restrictions. However, information that is merely publicly accessible still holds privacy and use restrictions (Office of the Privacy Commissioner of Canada, 2020). In addition to the previously mentioned complexities, privacy laws in different countries often lack alignment despite sharing crucial similarities. This misalignment greatly complicates the integration of legal privacy systems into a comprehensive tool such as D-PATH.

[20] With the ongoing scientific and technological progress surrounding biomedical and health research and accompanying data-sharing infrastructures, the promised benefits of the field appear to be within our reach. The infrastructures and mechanisms to process and share the associated data allow researchers to continuously grow its volume, improve its quality, and better understand its interconnections with other data and other fields of knowledge (Kalaitzopoulos, Patel & Younesi, 2016)

[21] Research participants and the public generally trust scientific projects when their interests and fundamental human rights are respected. As such, it is of prime importance not only to encourage practices of responsible and privacy-compliant data sharing but also to develop policies, guidelines, frameworks, and tools that enable researchers to implement them (Stark et al., 2019). This prompted the creation of EpiShare and D-PATH.

[22] EpiShare aims to uphold the public's trust and maintain a sustainable level of research participation and a constant pace of scientific progress with how its platform

22 The term "covered entity" under HIPAA refers to individuals or entities that transmit health information for transactions such as healthcare claims, payment, and remittance advice, healthcare status, coordination of benefits, enrolment and disenrollment, eligibility checks, healthcare electronic fund transfers, and referral certification and authorization. Steve Alder, "What Are Covered Entities Under HIPAA?", (18 October 2020), online: HIPAA Journal <https://www.hipaajournal.com/covered-entities-under-hipaa/>.

works. However, its users still lack assistance complying with all the different aspects of responsible data sharing, particularly with respect to their specific privacy obligations. D-PATH helps EpiShare's users (and other data processors) navigate through the complex privacy legal systems of a growing number of jurisdictions. The level of complexity of legal privacy systems varies depending on the jurisdiction. Whereas some jurisdictions have one specialized law applicable across the country, others have overlapping laws with different levels of competence that need to interoperate. An example of the latter is Canada, where given the provincial and federal division of competence, data processors must navigate through provincial and federal laws that focus on private or public bodies and even through laws that, despite focusing on matters other than privacy, include provisions that touch on privacy-related issues. When considering the international context, this complexity is naturally amplified.

[23] D-PATH differs from the similar tools we identified in its unique focus on privacy and data protection, its concrete and useful output, and its potential to extend to other jurisdictions. Whereas the four tools we covered address privacy issues to a certain extent, only D-PATH and Canada's OPC tool specialize in privacy matters. Nonetheless, the latter's output is very general, as opposed to D-PATH's, which provides very specific answers for some jurisdictions without an equal. Moreover, the scope of each of those tools is limited to one jurisdiction (e.g., European Union, the United States, or Canada) without an easy path toward expansion. Contrastingly, D-PATH was always designed to be expandable. Finally, while DSW provides concrete and practical outputs similar to what D-PATH does, D-PATH focuses more specifically on privacy and data protection aspects of data sharing and provides a unique level of specificity.

AREAS OF FUTURE IMPROVEMENT

[24] D-PATH is currently a proof-of-concept tool. With that, several limitations should be acknowledged. Firstly, in its immature form, D-PATH has limited jurisdictional coverage. It primarily focuses on Quebec (Canada), the E.U.'s GDPR, and several best-practice documents. Given that D-PATH was devised in the context of EpiShare, logically, its first iterations started with Quebec and Canada. However, despite the tool's currently limited scope, a key characteristic of D-PATH is that it can be expanded to incorporate different jurisdictions and best practices. In the future, we envision D-PATH expanding to multiple jurisdictions, best practices, and ethics policies. For example, it has been suggested that D-PATH would benefit from integrating aspects of Indigenous Data Governance best practices such as OCAP (Ownership, Control, Access, Possession) (FNIGC, 2018). At present, D-PATH demonstrates the value and feasibility of an online tool in guiding stakeholders to fulfill their data sharing responsibilities.

[25] It must also be noted that privacy laws and their legal interpretations are actively in flux. Key variables and concepts are highly context-dependent. Definitions will change depending on new developments, such as court cases and official guidance documents. This means a tool like D-PATH must be actively maintained and updated to remain useful. Relatedly, this also means that there are relevant topics that D-PATH cannot hope to comprehensively provide guidance on.

[26] An example of this is determining when data becomes personal and, relatedly, when data is considered de-identified. Identifiability is a significant factor in the decision-

making tree. In many cases, this distinction (of whether data is identifiable or not) will be apparent, but in some cases, additional guidance will be important in contrasting personal and non-personal data. Further complicating matters, these ideas may even differ between jurisdictions. In the future, D-PATH will direct users to informative discussions or expert articles on topics where the state of the law(s) is particularly nascent and under discussion.

[27] Privacy laws and responsible data sharing practices are important for achieving a favourable balance between data sharing and research participants' privacy. However, the complex nature of these tools can be a significant challenge for data sharing. D-PATH makes an important contribution by providing relevant stakeholders with clear and concrete knowledge of the primary privacy and data-protection obligations that data processors must and should respect. By simplifying and organizing these obligations in an intuitive and lay manner, D-PATH has the potential to facilitate responsible data sharing significantly. In the early 2020s, Artificial Intelligence (A.I.) tools have witnessed a remarkable advancement, developing at an astonishing rate. Over the course of just three years, from 2020 to 2023, the emergence of A.I. tools has been dramatic, as they have become increasingly helpful in a variety of settings and have become capable of handling ever-more complex tasks (Salvagno, Taccone & Gerli, 2023). The current proof-of-concept version of D-PATH does not include A.I. support, but it is feasible that future iterations will. If integrated, A.I. could make the tool even more useful and accessible. This would be consistent with D-PATH's purpose to simplify privacy and data-protection obligations to promote responsible data use.

Code Availability

D-PATH's source code is made available at https://github.com/c3g/d-path under the free and open-source software license Apache 2.0 license

Acknowledgements

We would like to acknowledge and express our gratitude to Genome Canada, Genome Quebec, the Canadian Foundation for Innovation, Calcul Québec, Compute Canada and the International Human Epigenomic Consortium for funding and supporting our research.

Authors Contributions

Palmira Granados-Moreno (PGM): Conceptualization, Methodology, Validation, Formal Analysis, Writing-Original Draft, Supervision

Hanshi Liu (H.L.): Communication with Reviewers, Conceptualization, Methodology, Validation, Formal analysis, Investigation, Writing- Original Draft, Writing- Review and Editing,

Sebastian Ballesteros Ramirez (SBR): Software, Validation, Writing- Review and Editing

David Bujold (D.B.): Software, Validation, Writing- Review and Editing

Ksenia Zaytseva (K.Z.): Software, Validation, Writing- Review and Editing,

Guillaume Bourque (G.B.): Funding Acquisition, Supervision, Writing- Review and Editing,

Yann Joly (Y.J.): Conceptualization, Methodology, Validation, Writing- Review and Editing Supervision, Project Administration, Funding Acquisition,

Conflict of Interest

All authors declare that they have no conflicts of interest.

Data Availability Statement

Data sharing not applicable to this article as the research did not generate new datasets.

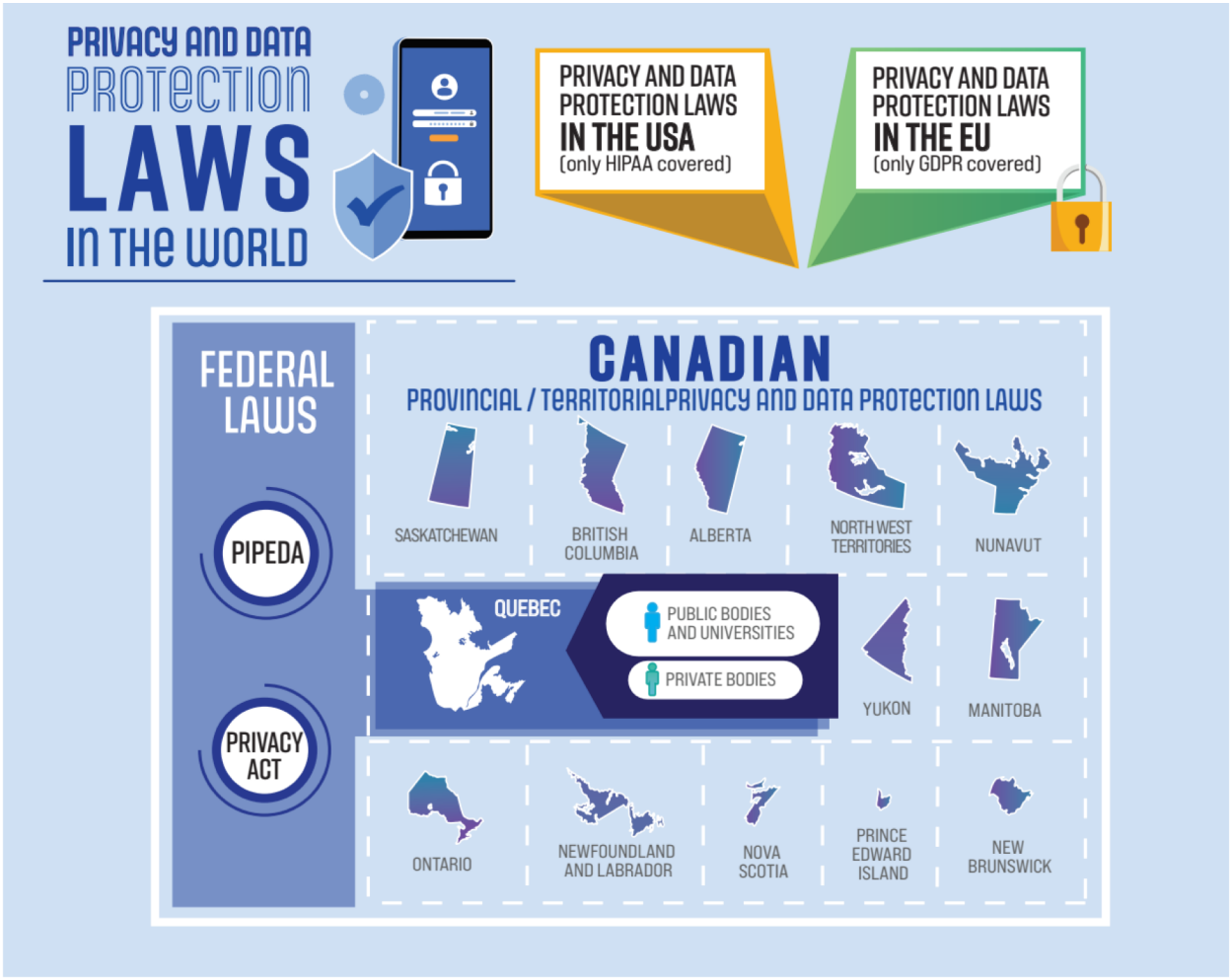
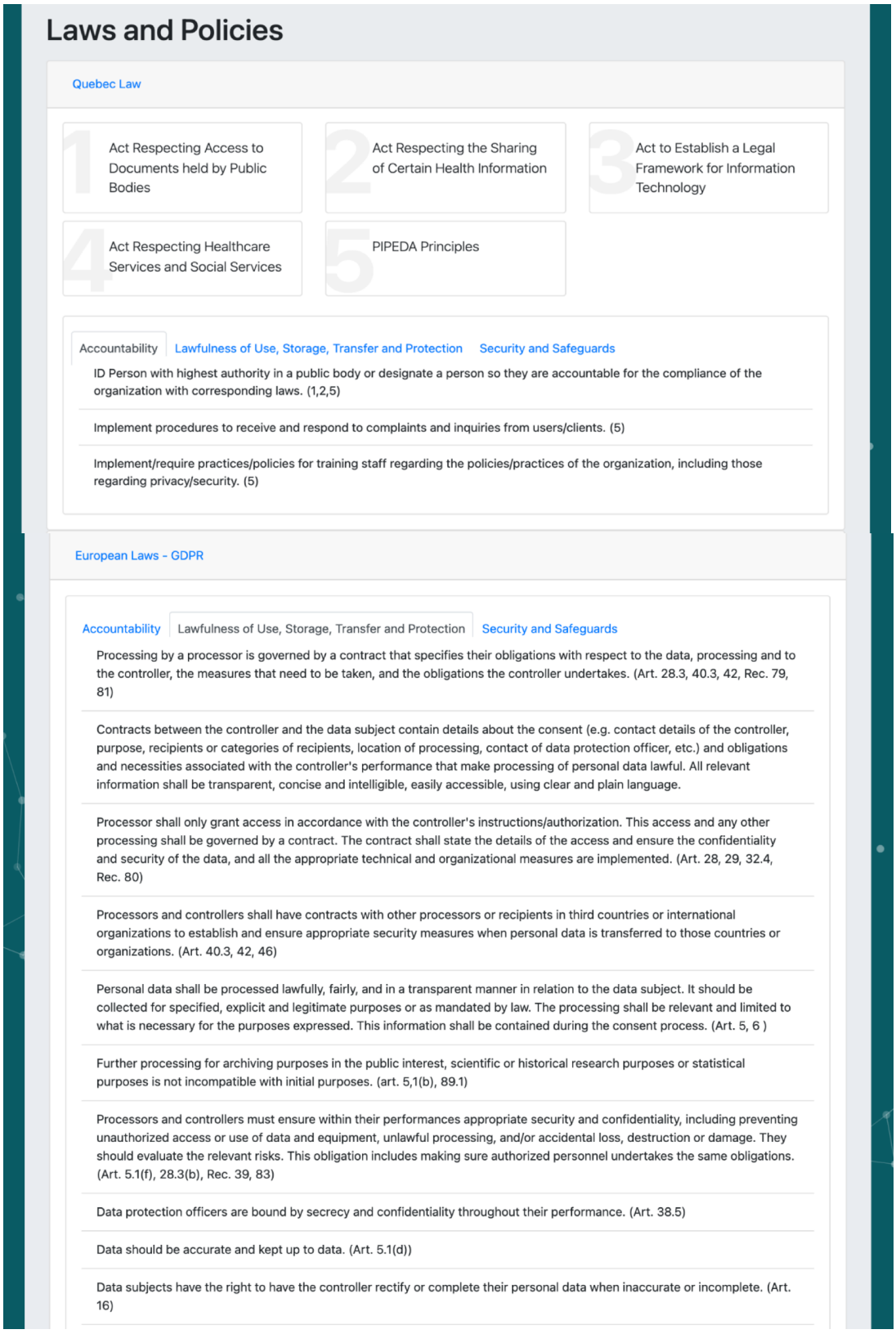
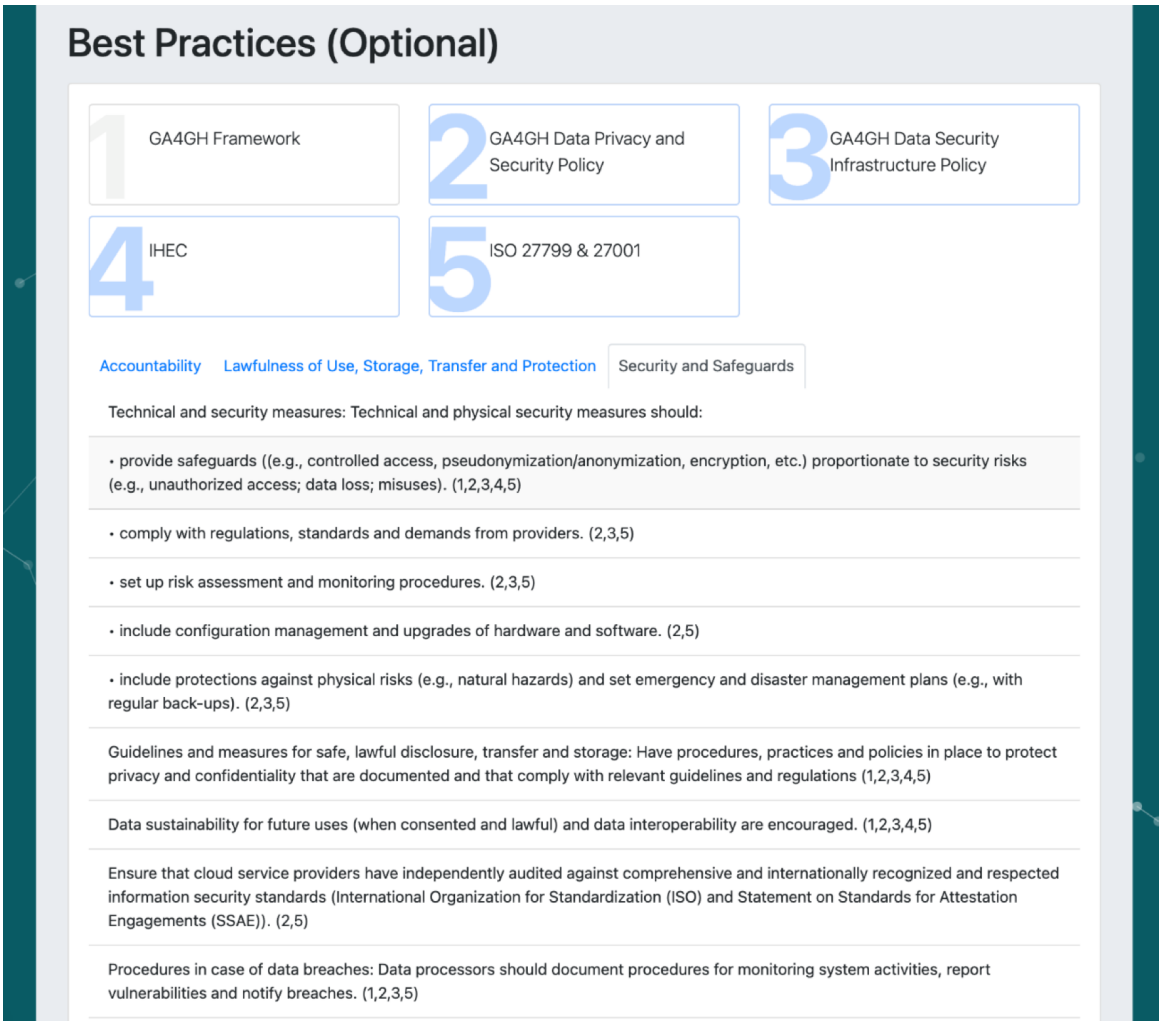


Figure 1. *D-PATH's scope of laws considered*

Table 1. *Examples of obligations and practices displayed after using D-PATH*





Supplementary Material

***Table 2**. Canadian Privacy Laws grouped by their sector*

Province / Territory	Law			
Alberta	Freedom of Information and Protection Privacy Act	Personal Information Protection Act	Health Information Act	
British Columbia	Freedom of Information and Protection of Privacy Act	Personal Information Protection Act	E-Health (Personal Health Information Access and Protection of Privacy) Act	

Province / Territory	Law			
Manitoba	Freedom of Information and Protection of Privacy Act	Personal Health Information Act, Manitoba's privacy law relating to health records		
New Brunswick	Right to Information and Protection of Privacy Act	Personal Health Information Privacy and Access Act		
Newfoundland and Labrador	Access to Information and Protection of Privacy	Personal Health Information Act		
Northwest Territories	Access to Information and Protection of Privacy Act	Health Information Act		
Nova Scotia	Freedom of Information and Protection of Privacy	Personal Health Information Act	Part XX of the Municipal Government Act	Personal Information International Disclosure Act
Nunavut	Access to Information and Protection of Privacy Act			
Ontario	Freedom of Information and Protection of Privacy Act	Municipal Freedom of Information and Protection of Privacy Act	Personal Health Information Protection of Privacy Act	
Prince Edward Island	Freedom of Information and Protection of Privacy Act			
Quebec	Act Respecting Access to Documents Held by Public Bodies and the Protection of Personal Information	Act Respecting the Protection of Personal Information in the Private Sector	Health Insurance Act	Act Respecting the Régie de l'Assurance Maladie du Québec

Province / Territory	Law			
Saskatchewan	Freedom of Information and Protection of Privacy Act	Health Information Protection Act		
Yukon	Access to Information and Protection of Privacy Act	Health Information Privacy and Management Act		

Figure 2. Information Classification Decision Tree : *This decision tree illustrates the logic behind how D-PATH classifies the applicable type of information based on the user's inputted answers. After information classification, D-PATH then identifies the relevant laws or policies applicable*

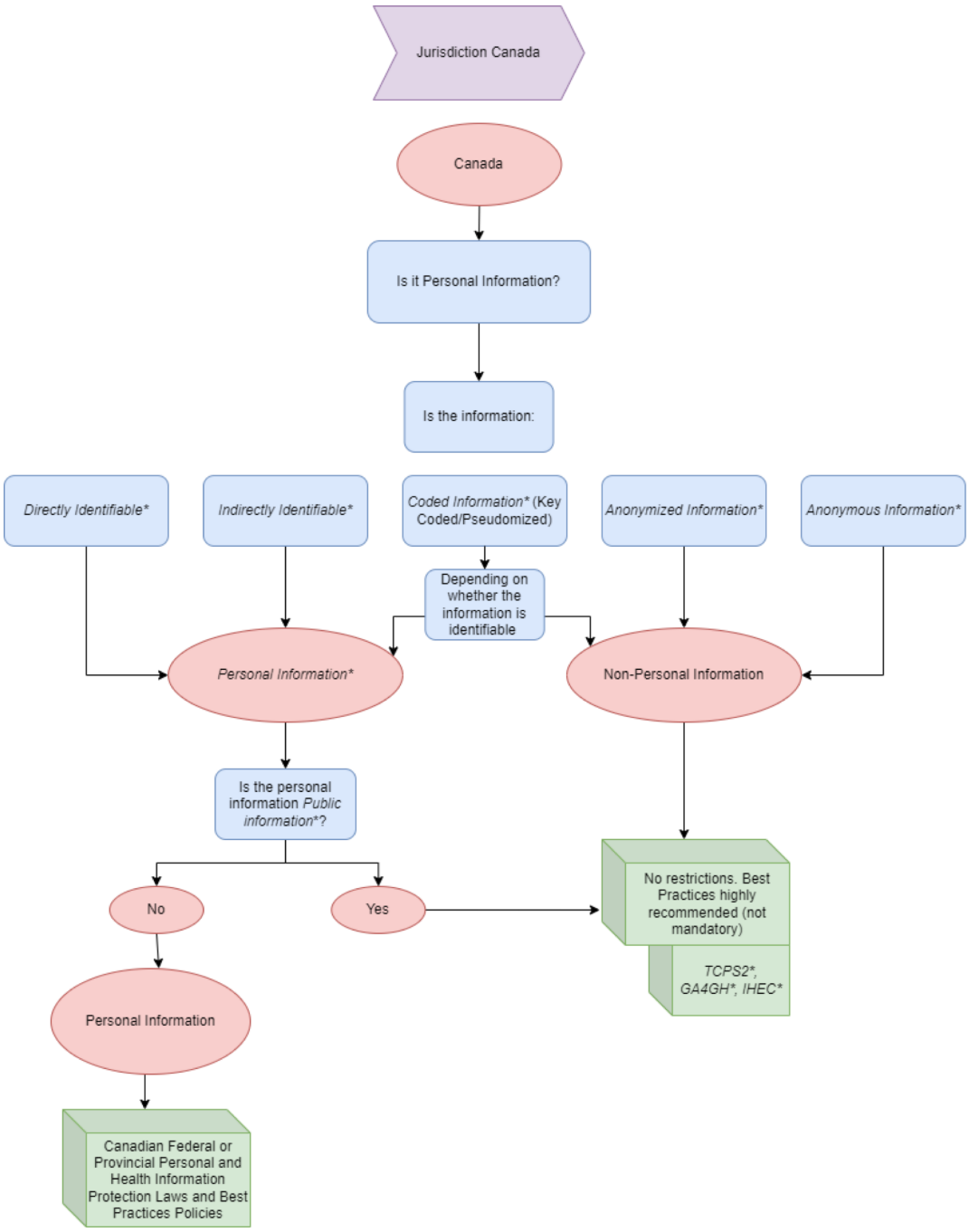


Figure 3. The Legal Decision Tree in Canada : *This diagram provides an overview of how information is also classified based on who performs the information processing*

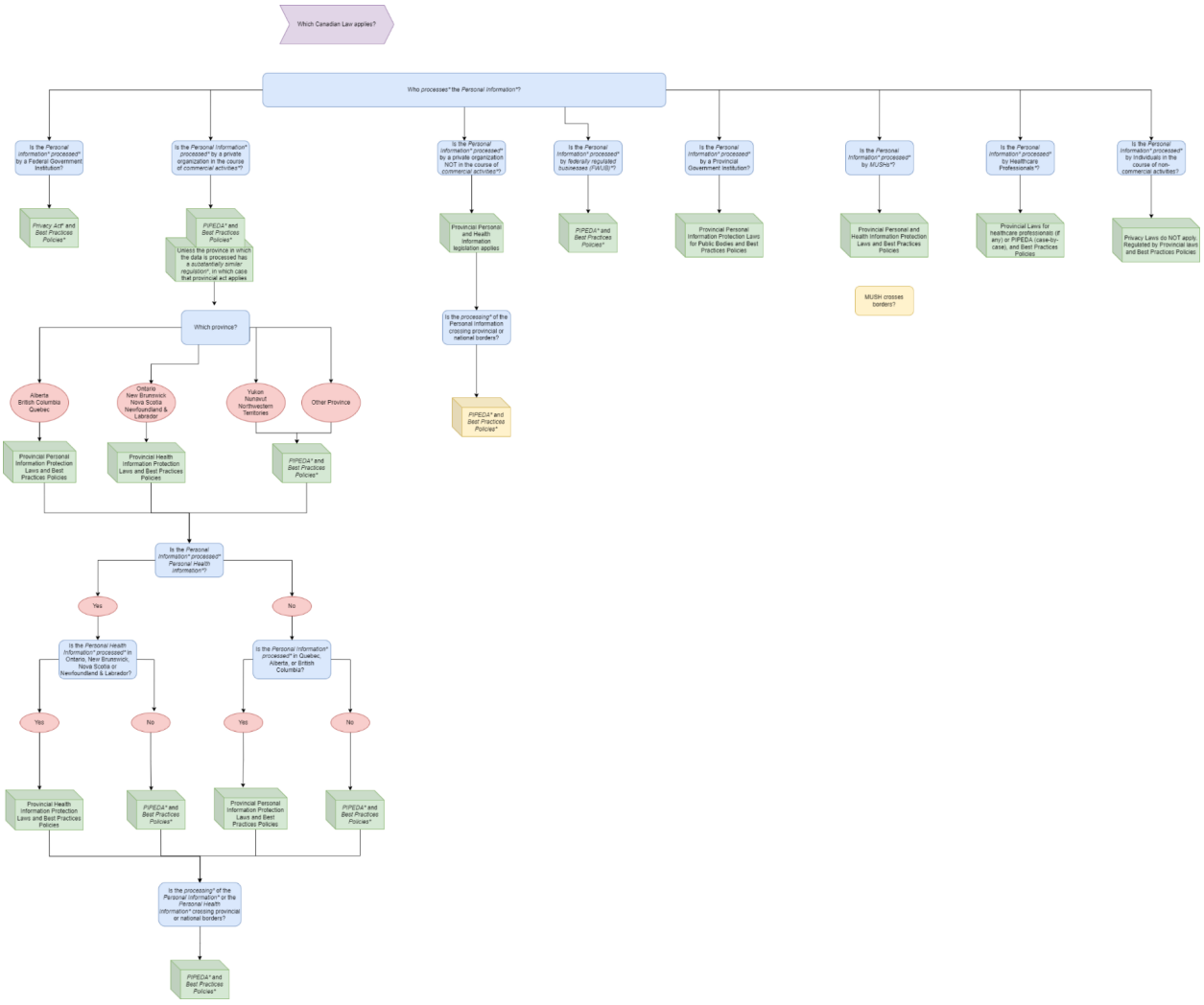


Figure 4 : *This diagram describes how D-PATH classifies information types based on user input. Firstly, the app established the location/jurisdiction applicable, then based on this, the app assesses whether the information is personal or not based on general features relating to data identifiability*

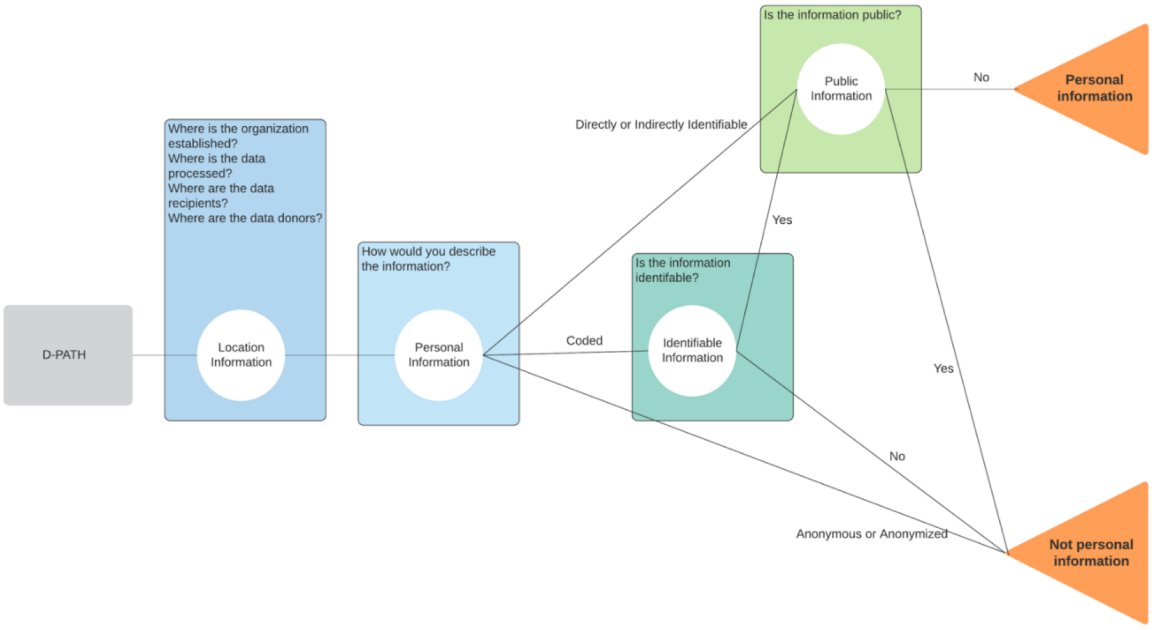
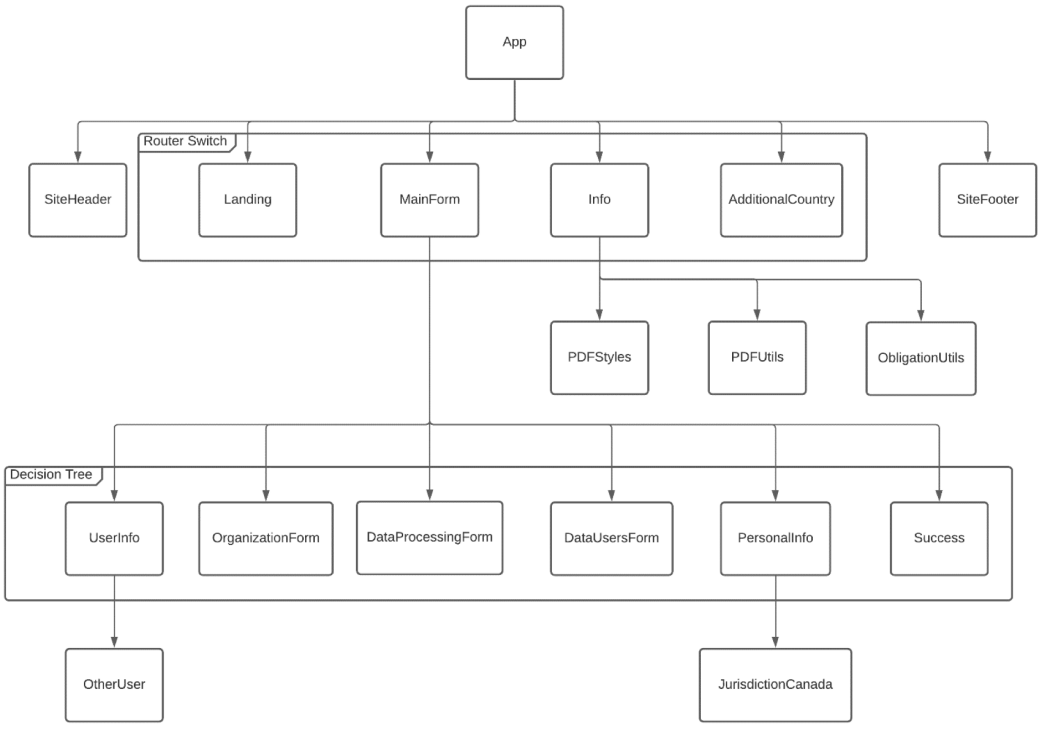


Figure 5 : *The following diagram is a representation of the web application and how it is organized. Most of the logic of the decision tree lies in the main form stage. Here, the users respond to all the questions and a tailored assessment is provided on the Information page according to their answers*



References / Bibliographie

The Personal Information Protection and Electronic Documents Act (PIPEDA), SC 200 c 5 2000.

Leonelli, Sabina, Data-Centric Biology (University of Chicago Press, 2016).

Beauvais, Michael J S & Bartha Maria Knoppers, "Coming Out to Play: Privacy, Data Protection, Children's Health, and COVID-19 Research" (2021) 12 Front Genet 524, online: <https://www.frontiersin.org/article/10.3389/fgene.2021.659027>.

Becker, Regina et al., "DAISY: A Data Information System for accountability under the General Data Protection Regulation" (2019) 8:12 Gigascience, online: <https://academic.oup.com/gigascience/article/8/12/giz140/5652251>.

Ben-Eghan, Chief et al., "Don't ignore genetic data from minority populations" (2020) 585:7824 Nature 184–186, online: <https://www.nature.com/articles/d41586-020-02547-3>.

Bernier, Alexander & Bartha Maria Knoppers, "Pandemics, privacy, and public health research" (2020) 111:4 Can J Public Health 454–457, online: <https://doi.org/10.17269/s41997-020-00368-5>.

Chawinga, Winner Dominic & Sandy Zinn, "Global perspectives of research data sharing: A systematic literature review" (2019) 41:2 Libr Inf Sci Res 109–122, online: <http://www.sciencedirect.com/science/article/pii/S074081881830330X>.

European Parliament, "Regulation (E.U.) 2016/679 of the European Parliament and of the Council of 27 April 2016 (General Data Protection Regulation)" (2016) O.J. 2016 L 119/1, online: <https://gdpr-info.eu/>.

Granados Moreno, Palmira, Yann Joly & Dylan Roskams-Edris, "Could Open Be the Yellow Brick Road to Innovation in Genomics in North America?" (2020) 13:1 McGill J Law Health 119–180, online: <https://mjlh.mcgill.ca/publications/volume-13-issue-1-131-2019/could-open-be-the-yellow-brick-road-to-innovation-in-genomics-in-north-america/>.

Gurdasani, Deepti et al., "Genomics of disease risk in globally diverse populations" (2019) 20:9 Nat Rev Genet 520–535, online: <https://www.nature.com/articles/s41576-019-0144-0>.

Harris, Theresa L & Jessica M Wyndham, "Data Rights and Responsibilities: A Human Rights Perspective on Data Sharing" (2015) 10:3 J Empir Res Hum Res Ethics 334–337.

Hulsen, Tim, "Sharing Is Caring—Data Sharing Initiatives in Healthcare" (2020) 17:9 Int J Environ Res Public Health, online: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7246891/>.

Joly, Yann et al., “Are Data Sharing and Privacy Protection Mutually Exclusive?” (2016) 167:5 Cell 1150–1154.

———, “Data Sharing in the Post-Genomic World: The Experience of the International Cancer Genome Consortium (ICGC) Data Access Compliance Office (DACO)” (2012) 8:7 PLOS Computational Biology e1002549.

Kalaitzopoulos, Dimitris, Ketan Patel & Erfan Younesi, “Advancements in Data Management and Data Mining Approaches” in Aamir Shahzad, ed, Translational Medicine Tools and Techniques (Academic Press, 2016) 36.

Kaye, Jane, “The Tension Between Data Sharing and the Protection of Privacy in Genomics Research” (2012) 13:1 Annu Rev Genom Hum Genet 415–431, online: <https://www.annualreviews.org/doi/10.1146/annurev-genom-082410-101454>.

Knoppers, Bartha M. et al., “A human rights approach to an international code of conduct for genomic and clinical data sharing” (2014) 133:7 Hum Genet 895–903.

Knoppers, Bartha Maria & Michael J S Beauvais, “Three decades of genetic privacy: a metaphoric journey” (2021) 30:R2 Hum Mol Genet R156–R160, online: <https://doi.org/10.1093/hmg/ddab164>.

LeBel, Etienne P, Lorne Campbell & Timothy J Loving, “Benefits of open and high-powered research outweigh costs” (2017) 113:2 J Pers Soc Psychol 230–243.

Levenstein, Margaret C & Jared A Lyle, “Data: Sharing Is Caring” (2018) 1:1 Adv Meth Pract Psychol Sci 95–103, online: <https://doi.org/10.1177/2515245918758319>.

Mecredy, Graham, Roseanne Sutherland & Carmen Jones, “First Nations Data Governance, Privacy, and the Importance of the OCAP® principles” (2018) 3:4 International Journal of Population Data Science, online: <https://ijpds.org/article/view/911>.

Middleton, Anna et al., “Global Public Perceptions of Genomic Data Sharing: What Shapes the Willingness to Donate DNA and Health Data?” (2020) 107:4 Am J Hum Genet 743–752, online: <http://www.sciencedirect.com/science/article/pii/S0002929720302925>.

Pergl, Robert et al. “‘Data Stewardship Wizard’: A Tool Bringing Together Researchers, Data Stewards, and Data Experts around Data Management Planning” (2019) 18:1 Data Sci J 59, online: <http://datascience.codata.org/articles/10.5334/dsj-2019-059/>.

Salvagno, Michele, Fabio Silvio Taccone & Alberto Giovanni Gerli, “Can artificial intelligence help for scientific writing?” (2023) 27:1 Critical Care 75.

Saulnier, Katie M et al., “Benefits and barriers in the design of harmonized access agreements for international data sharing” (2019) 6:1 Sci Data 297.

Stark, Zornitza et al., “Integrating Genomics into Healthcare: A Global Responsibility” (2019) 104:1 Am J Hum Genet 13–20, online: <http://www.sciencedirect.com/science/article/pii/S0002929718304221>.

Trinidad, M Grace, Jodyn Platt & Sharon L R Kardia, “The public’s comfort with sharing health data with third-party commercial companies” (2020) 7:1 Humanit Soc Sci Commun 1–10.

Alder, Steve, “What Are Covered Entities Under HIPAA?”, (18 October 2020), online: HIPAA Journal <https://www.hipaajournal.com/covered-entities-under-hipaa/>.

Bourque, Guillaume & Yann Joly, Epigenomics Secure Data Sharing Platform for Integrative Analysis (EpiShare). Submitted to the 2017 Bioinformatics and Computational Biology Competition of Genome Canada (2017 Bioinformatics and Computational Biology Competition of Genome Canada, 2017).

Bujold, David et al., EpiShare: an open platform to securely share epigenomic data (2020).

Canada, Office of the Privacy Commissioner of, “PIPEDA Findings #2020-004: Joint investigation of the Cadillac Fairview Corporation Limited by the Privacy Commissioner of Canada, the Information and Privacy Commissioner of Alberta, and the Information and Privacy Commissioner for British Columbia paras 113-125”, (29 October 2020), online: <https://www.priv.gc.ca/en/opc-actions-and-decisions/investigations/investigations-into-businesses/2020/pipeda-2020-004/>.

CIHR, NSERCC, & SSHRC, “Tri-Council Policy Statement. Ethical Conduct for Research Involving Humans”, (2018), online: <https://ethics.gc.ca/eng/documents/tcps2-2018-en-interactive-final.pdf>.

CMSgov, “Covered Entity Guidance Tool”, (2020), online: Centers for Medicare & Medicaid Services <https://www.cms.gov/Regulations-and-Guidance/Administrative-Simplification/HIPAA-ACA/Downloads/CoveredEntitiesChart20160617.pdf?fbclid=IwAR02qYceeKBelU-G_XfEXmq5NLQGg2MXKrzeK5McXQYuRmerZrjbK2EqJi8>.

DSW, “Data Stewardship Wizard”, (13 December 2020), online: <https://ds-wizard.org/about.html>.

EpiShare, “EpiShare - About”, (2019), online: EpiShare <https://epishare-project.org/about.html>.

European Commission, “European Economic Area (EEA)”, (10 August 2021), online: <https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:European_Economic_Area_(EEA)>.

Fidler, Fiona & John Wilcox, “Reproducibility of Scientific Results” in Edward N Zalta, ed, The Stanford Encyclopedia of Philosophy, winter 2018 ed (Metaphysics Research Lab, Stanford University, 2018).

FNIGC, “OCAP Education and Training”, online: The First Nations Information Governance Centre <https://fnigc.ca/what-we-do/education-and-training/>.

GA4GH, “Framework for Responsible Sharing of Genomic and Health-Related Data”, (9 December 2014), online: <https://www.ga4gh.org/genomic-data-toolkit/regulatory-ethics-toolkit/framework-for-responsible-sharing-of-genomic-and-health-related-data/>.

———, “Global Alliance for Genomics and Health: Data Privacy and Security Policy”, (August 2019), online: <https://www.ga4gh.org/wp-content/uploads/GA4GH-Data-Privacy-and-Security-Policy_FINAL-August-2019_wPolicyVersions.pdf>.

GNU, “Various Licenses and Comments about them. GNU Operating System”, (May 2020), online: GNU org <https://www.gnu.org/licenses/license-list.en.html>.

Health Canada, Health, “Privacy: a fundamental right in Canada”, (8 December 2006), online: <https://www.canada.ca/en/health-canada/services/environmental-workplace-health/reports-publications/occupational-health-safety/privacy-fundamental-right-canada-national-dosimetry-services.html>.

McFarland, Michael, “Why We Care about Privacy”, (1 June 2012), online: Markkula Center for Applied Ethics Santa Clara University <https://www.scu.edu/ethics/focus-areas/internet-ethics/resources/why-we-care-about-privacy/>.

National Cancer Institute, “Epigenomics and Epigenetics Research”, (24 July 2020), online: National Cancer Institute National Institutes of Health <https://epi.grants.cancer.gov/epigen/>.

Office of the Privacy Commissioner of Canada, “Summary of privacy laws in Canada”, (January 2018), online: <https://www.priv.gc.ca/en/privacy-topics/privacy-laws-in-canada/02_05_d_15/>.

OHSU, “Health Insurance Portability and Accountability Act (HIPAA)”, (2020), online: Oregon Health and Science University <https://www.ohsu.edu/information-technology/health-insurance-portability-and-accountability-act-hippa>.

FIGURES AND TABLES

D-PATH URL: https://www.computationalgenomics.ca/tools/d-path