International Genomic Data Sharing by Health Technologies Industries*: Points to Consider

v 17 August 2023

Industry Core Group – Members

Shane Chase, Global Head of Policy Governance, Ethics and Privacy, Illumina
Ashley Van Zeeland, Vice President, Corporate and Business Development, Illumina
Holden Williams, Ethics and Privacy Specialist, Illumina
Jacquie Cooke, General Counsel & Privacy Officer, 23andMe
Caitlyn Bruns, Senior Manger, DNA Strategy, Ancestry
Kristin Rand, Director, DNA Strategy & Partnerships, Ancestry
Dr Rowland Illing DM MRCS FRCR, Director & Chief Medical Officer, International Public Sector Health, Amazon Web Services
Erla Th. Petursdottir, DeCODE Genetics
Grant Stapleton, Director of Assurance, Genomics England
Nick Maltby, General Counsel, Company Secretary and Data Protection Officer, Genomics England
Deven McGraw, Lead for Data Stewardship and Data Sharing, Invitae
Geralyn Miller, Senior Director, Microsoft Health AI, Microsoft
Donna Salerno, Vice President, Privacy, Myriad Genetics
Brandon Martin, Senior Corporate Counsel, Roche Molecular Systems, Inc.
Garth Rosengren, Head of Legal, Molecular Labs Customer Area, Roche Molecular Systems, Inc.
Michael Lisi, Head of Legal, Roche Diagnostics Solutions, Roche Diagnostics Solutions
Fiona Laird Hyland, Thermo Fisher Scientific

Centre of Genomics and Policy, McGill University

Bartha Maria Knoppers, Founding Director
Yann Joly, Director
Ma'n Zawati, Research Director
Adrian Thorogood, External Collaborator

*** Health Technologies Industries (HTI):** includes companies from diverse sectors such as pharmaceutics, sequencing platforms, clinical genomics, direct-to-consumer genetic testing, digital health data management, cloud computing, advanced analytics and AI, and digital health and engagement platforms.

International Genomic Data Sharing by HTI: Points to Consider

Introduction

This document outlines Points to Consider (PtC) for the responsible sharing of human genomic and health data internationally by Health Technologies Industries (HTI). HTI can contribute unique resources, technologies, and expertise to translating genomic discoveries into improvements in human health. International data sharing can further accelerate research and innovation. It can strengthen statistical power and reproducibility, facilitate collaboration and creative re-use of real-world evidence, increase the representativeness of precision medicine databases, and power AI approaches (including machine-learning, deep learning and predictive modeling) that support genomic interpretation and clinical decision-making.

Yet, research, innovation, and data sharing to advance precision medicine also raise important ethical issues, which include risks to the welfare and privacy of sequenced individuals, their families, and communities. The legal and policy landscape relating to data sharing is rapidly evolving. Relevant norms apply in areas of data privacy and protection law; AI law, governance, and ethical principles; research ethics regulations; and data sharing policies.

This PtC tailored for HTI builds on the GA4GH's **Framework for Responsible Sharing of Genomic and Health-related Data (2014, re-approved 2019)** and subsequent policies. The Framework is founded on human rights, aiming in particular to activate the right of everyone to share in scientific advancement and its benefits. Relevant Core Elements include: Transparency; Accountability; Data Quality and Security; Privacy, Data Protection and Confidentiality; Risk-Benefit Analysis; and Recognition and Attribution. Implementation of this PtC requires careful attention to the particular context – including the relevant jurisdictions, applicable laws and policies, sectors, companies, data sharing activities, and types of health and genomic data.[1] The PtC is accompanied by **Explanatory Notes** (Appendix A) and issue-driven **Briefs** (Appendix B) to set the international context.

Enforceability: This PtC is intended for voluntary adoption by HTI. It can be customized into more prescriptive sectoral best practices or enforceable codes of conduct. A number of points reflect legal requirements in certain jurisdictions, and will thus already be binding and enforceable. **Local laws take precedence in case of conflict.**

[1] The PtC focuses on human genomic and health data, and does not address the sharing of other genomic data, e.g,. from pathogens, animals, or plants. Human genomic data may variously include raw data, processed data, test/analysis results, predictive models, or predictions. Different types of genomic and health data may have different levels of identifiability and sensitivity, and implicate different rights and interests.

POINTS TO CONSIDER

Consent

Consent for the processing and sharing of genomic and health data is both an ethical principle and often a legal requirement. Consent should generally be freely given, informed, and ongoing. National laws and policies may define different circumstances where consent is or is not required, as well as the elements of a valid consent. An organization seeking consent should consider:

- Defining the scope of consent to data sharing, namely the types/categories of data (and their level of identifiability), the categories of users (e.g., industry and international users), and the scope of further research purposes (consent to general research purposes, or certain research areas, may be forseen by applicable law and policy).
- Seeking regular review and approval of consent materials and processes by a research ethics committee or similar body where appropriate.
- Supporting consent with accompanying safeguards (e.g., oversight, transparency, and a right to withdraw consent).
- Notifying participants of the right to withdraw consent to data sharing, the mechanisms to do so, and any limitations (e.g., where data has already been shared or results have already been published).
- Adapting consent to the target population(s) (e.g., rare disease patients, children, marginalized groups, ...).
- Offering meaningful choices and obtaining prior consent to any return of results of clinical importance to individuals or their families, where return is legally required or otherwise planned.
- Using accessible and intuitive e-consent interfaces – in digital environments – that facilitate meaningful choices and dynamic information exchange.
- Documenting of consents (and any withdrawals), versioning of consents, and capturing consents in a machine-readable form.

Transparency

It is an ethical principle and often a legal requirement to notify research participants about the processing or sharing of genomic and health data, the use of AI and its potential impact on participants or patients, and the possible return of results of health importance to individuals or their families. Applicable law and policy may permit exceptions and may alternatively require public notices. The following should be considered when providing information to research participants:

- Describing the scope of data collection/generation, use, storage, and sharing, including any possibility of external and international data sharing, and whether their information will first be anonymized or pseudonymized/coded before any sharing.

- Describing any use of AI in health products and services, including the types of data incorporated, the potential affects of outputs on individuals, and an explanation of how the AI makes its predictions, recommendations, or decisions.
- Describing an ethically justified and context-appropriate plan outlining the scope of return of results (if any) to research participants and their families.
- Where return of results is forseen, describing the extent of the company's responsibility for clinical validation, communication, and re-interpretation over time.
- Providing sufficient information to participants in plain language. A layered approach to consent can support comprehension by first providing key information, followed by details in later sections.
- Standardizing information elements provided to participants by using voluntary or binding codes of conduct.

Identifiability

The level of identifiability is a key determinant of data's privacy risk profile and regulatory status (e.g., anonymized vs. pseudonymized/coded data). While there are differences in terminology and standards relating to identifiability across jurisdictions, international data sharing norms promote a common and proportionate approach to privacy protection. Data sharing plans should consider:

- Conducting regular privacy risk assessment(s) to identify data types, assess their legal status and risk profile in context, and define appropriate controls over time (e.g., privacy, security, and contractual).
- Documenting and periodically reviewing privacy risk assessments and controls as technology and data contexts evolve, with appropriate involvement of experts.
- Sharing data in the least identifiable form necessary (e.g., anonymized, pseudonymized/coded) to achieve the intended purpose.
- Establishing complementary privacy and security safeguards proportionate to privacy risk for better data protection and use.
- Managing trade-offs between protecting privacy and preserving the utility, quality, and representativeness of data.
- Taking into account the practical and legal challenges in some jurisdictions to anonymizing/ pseudonymizing genomic data.

Privacy and Data Protection

The processing and sharing of human genomic and health data implicates the privacy and data protection rights of individuals. Data sharing plans should consider:

- Ensuring personal data are shared in a lawful, fair, and transparent manner, for explicit and legitimate purposes.
- Limiting collection, storage, and sharing of personal data to what is necessary to achieve the intended purposes.
- Assessing and addressing the risks that cross-border transfers of data may result in a lower level of legal privacy protection for participants.
- Employing data localization strategies in order to limit or reduce the risks of cross-border transfers and support compliance with local norms, while still permitting virtual pooling of data internationally.
- Establishing data breach notification procedures in accordance with applicable law.
- Enabling individuals to meaningfully exercise their data protection rights (e.g., the right to access, to rectification, or to not be subjected to a decision based solely on automated processing) where applicable and subject to local limitations.
- If providing individuals access to data is forseen, ensuring that data are provided in a common, machine-readable format suitable for analysis and portability, along with readable summary reports and data quality disclaimers.
- Capturing all of the above elements in a data privacy notice that is written in plain language, accurate, complete, and publicly accessible.

Data Security and Quality

Data security involves protecting the confidentiality, availability, and integrity of genomic and health data. Data quality is a key enabler for realizing the opportunities of data sharing. Data sharing plans should consider:

- Implementing data governance safeguards appropriate for the particular context, including the purpose of processing, the nature of the data, and the associated risks (e.g., data access committees, data access agreements, and privacy enhancing technologies).
- Using data sharing platforms and computing environments that implement cybersecurity standards (e.g., the ability to restrict access to authorized persons, identity authentication and authorization management, audit logs, encryption of data at-rest and in-transit, regular risk assessments and security incident response protocols).
- Regularly auditing and testing security measures, and periodically reviewing them as technology evolves and new vulnerabilities emerge.

- Adopting data breach plans covering breach detection, response, and recovery.
- Following the FAIR Data principles (that data be findable, accessible, interoperable, and reusable), including adopting standard ontologies and schemas facilitating data pooling and linkage.
- Demonstrating the effectiveness of AI models across the diversity of target populations, including demonstrating that training datasets are representative.

Responsible Use

Responsible use involves balancing data control for legitimate commercial interests with data sharing for the public good. Data governance approaches should consider:

- Maximizing the commercial, scientific and societal value of genomic and health data while respecting legal, privacy and ethical requirements.
- Limiting exclusive control over genomic and health data to the time necessary to protect IP rights, preserve commercialization opportunities, or ensure the integrity of studies (e.g., clinical trials).
- Reflecting the particular context of a company's activities (e.g., academic, commercial, philanthropic, pre-competitive, ...).
- Supporting authorities and researchers in response to public health emergencies.
- Using open access resources in a manner that does not impede innovation by others.
- Returning derived data and results to data contributors and sharing such data with the scientific community.
- Ensuring responsible and fair use of AI models.

Governance of HTI / Academic Partnerships

Diverse forms of partnerships and data sharing between HTI and academic researchers are key to the translation of scientific findings into innovative health applications. The governance of these partnerships should consider:

- Clarifying the nature of the partnership and HTI's role within it (e.g., as a research partner or platform provider).
- Defining the scope of HTI access to data, the use of data for HTI's own research and IP-development purposes, and any possible or actual onward sharing (e.g., with partners or law enforcement).
- Clearly articulating commercialization goals, and how these ultimately align with the public good (e.g., improved and more cost-effective patient care).
- Establishing commitments to scientific best practices, processes to manage conflicts of interest, and publication policies protecting scientific freedom.

- Ensuring platform providers store data in an interoperable form to enable portability between service providers, while respecting specific customer needs.
- Establishing up-front criteria for estimating the long-run cost of platform services to inform sustainability.
- Establishing contingency plans to safeguard data in case of a termination of platform services.

Community Engagement

Patients and communities should be engaged in the development and governance of research. Patient engagement refers to the meaningful and active collaboration of patients in research governance, priority setting, conduct, and publication. HTI may work with various types of communities (e.g., patient groups, indigenous communities, or developing countries), necessitating different forms of engagement. Patient engagement approaches should consider:

- Encouraging a greater diversity in research participation that reflects the target community of a study or product, to ensure data are representative.
- Promoting community education, awareness, and understanding of data sharing aims and practices.
- Involving community representatives in the design and implementation of data sharing policies and practices.
- Ensuring that the views solicited from communities are both diverse and representative.
- Sharing the benefits of innovation with communities in an equitable manner (arrangements may include where appropriate: acknowledgement of contributions, return of results, local capacity building, fair or open licensing commitments, expanded access, differential pricing, and recognizing indigenous data sovereignty).

APPENDIX A.

International Genomic Data Sharing by HTI: Explanatory Notes

Data Sharing Context

Large-scale precision medicine initiatives are underway around the world aiming to translate genomic technology into clinical trials and routine clinical care, while supporting ongoing research to better understand human biology and disease across diverse, global populations.[2] The unprecedented scale of these initiatives and the challenges of translation call for an increased role for HTI, in partnership with academic researchers, health systems and communities.

Sharing of human genomic and health data is promoted across the spectrum from discovery research, to clinical research, to clinical care. Data sharing can strengthen statistical power and reproducibility, facilitate creative re-use of real-world evidence, increase the representativeness of precision medicine databases, and power AI. It is also the way to realize the full scientific and societal benefits for which participants give their data. HTI may be required to share data by regulators or contracts, or may share genomic and health data voluntarily in support of the public interest, such as in response to public health emergencies.

A number of stakeholders, including funders, governments, international organizations as well as patient groups and publics have high expectations for genomic and health data sharing. UNESCO's **Recommendation on Open Science** (2021) sets forth an international framework for open science policy and practice that outlines common definitions, values, principles, and standards. The Recommendation aims to make knowledge openly available and to make science more collaborative and inclusive for the benefit of society. Clinical trial transparency and data sharing remain high on the agenda of regulators internationally. Scientific funding agencies and policy makers continue to strengthen and expand the scope of data sharing policies. For example, the US National Institutes of Health (NIH) **Data Management and Sharing Policy** now requires scientists across diverse disciplines to establish plans addressing the collection, management, preservation, and sharing of data, and to deposit data in suitable digital repositories with the aim of maximizing the scientific value of data. These plans will be reviewed as part of funding applications and monitored throughout projects.

Governments are working to promote responsible sharing of health information to support research, innovation, and public health. For example, the EU **Data Governance Act** aims to promote access to sensitive data held by the public sector, data sharing intermediaries, and data altruism. Data altruism organizations support individuals and companies to make their data available voluntarily for the public interest. The entity

[2] By way of example, these include the UK Our Future Health program, the US National Institutes of Health's million-person All of Us Research Program, the Human PanGenome Reference Consortium, Mount Sinai's Million Health Discoveries Program, Biobank Japan and Genome Medical Alliance Japan, Australian Genomics, the EU 1+Million Genomes Initiative and Genomic Data Infrastructure (GDI), FinnGen, UK Biobank, and Canada's All for One initiative, to name but a few.

must be registered under the Act and offer a high level of transparency and security. The draft **European Health Data Space** Regulation intends to compel most EU holders of electronic health data to make their holdings available for innovation, research, and public policy uses in a secure and transparent manner. The draft EU **Data Act** aims to ensure a broader set of stakeholders in the data supply chain can access private sector data and participate in creating value. The OECD **Recommendation on Health Data Governance** (2016) encourages countries to establish frameworks for making health data more available for research and public health, while managing privacy and security risks.

Legal and Regulatory Context

The regulatory landscape affecting HTI involved in genomics and health has been evolving rapidly in recent years. International norms like UNESCO's **Universal Declaration on the Human Genome and Human Rights** (1997) highlight that the individual and collective interest in human genetic data requires that data sharing activities be carried out in a consistent, transparent, and principled manner under proper governance frameworks.

Regional and (sub)national data protection laws generally require data be shared in a secure manner that is respectful of participants' autonomy and privacy. It is predicted that by 2024, **three-quarters of the world's population** will be protected by modern data protection laws. These laws typically include strict protections around health and genetic information collection, processing, and sharing. In the EU, the **General Data Protection Regulation** (GDPR) and related guidance and decisions from regulators and courts have significantly increased the data protection compliance burden of processing personal data, particularly for organizations dealing with sensitive health and genetic data. Japan was an early adopter of data protection laws in Asia, and recently updated its regime to strengthen rules around individuals' rights and breach notification, allowing it to obtain an adequacy decision from the EU. China has adopted comprehensive regulations specifically governing the use and export of human genetic materials and data.

The GDPR has also significantly impacted international transfers by strictly regulating data exports, reinforcing its extra-territorial effect on international data sharing. EU-US data transfers have been a persistent area of uncertainty. The European Commission and US government reached an agreement on a reinvigorated EU-US Data Privacy Framework in 2023 which entered into force with immediate affect. The EU-US Data Privacy Framework supports commercial data transfers for participating organizations who agree to the jurisdiction of US regulators. It aims to ensure that the surveillance practices of US signal intelligence agencies respect data protection principles, and that foreign individuals have their fundamental rights protected and have access to remedies from regulators and courts.

AI Governance Context

HTI are applying AI in genomics to interpret large-scale genomic datasets, annotate sequences, and help clinicians predict the influence of genetic variation on disease, often as part of broader efforts to draw on real-world data (RWD) – data about patient health status or the delivery of health care – to better understand an intervention's usage, effectiveness, and safety.

Laws and policies are evolving around the world to both promote and regulate AI, founded on principles including fairness, accountability, transparency, and explainability. A number of high-level principles have been articulated by international organizations and civil society initiatives, including the OECD's **AI Principles** (2019), UNESCO's **Recommendation on the Ethics of AI** (2021), and the **Montreal Declaration for a Responsible Development of AI** (2018). At the national level, countries are pursuing various strategies for regulating AI ranging from principles-based to prescriptive approaches, and from transversal to sector-specific regulatory approaches. Examples of prominent norms in this area include the proposed EU **AI Act**, UK **pro-innovation approach to AI regulation** (2023), US proposed **Algorithmic Accountibility Act** and **Blueprint for an AI Bill of Rights** (2022) – a set of voluntary guidance, China's **Ethical Norms for New Generation Artificial Intelligence** (2021), Canada's proposed **AI and Data Act**, and Singapore's **AI Governance Framework** (2020).

The Global Alliance for Genomics and Health

The GA4GH is an international, nonprofit alliance with a mission "to accelerate progress in genomic research and human health by cultivating a common framework of standards and harmonized approaches for effective and responsible genomic and health-related data sharing." The GA4GH Regulatory and Ethics Work Stream maintains a toolkit of regulatory and ethics guidance and tools to support responsible data sharing. **Table 1** summarizes the guidance documents and tools, and their application to HTI.

Table 1. Alignment with GA4GH Policies

Policy name	Policy type	Application	Notes
Framework for Responsible Sharing of Genomic and Health-Related Data (2014, re-approved 2019)	Framework	Required	
Model Consent Clauses (varia)	Model clauses	Optional	· Clinical Genomics Consent Clauses · Consent Clauses for Large Scale Initiatives · Familial Consent Clauses · Consent Clauses for Genomic Research · Pediatric Consent Clauses · Model Consent Clauses for Rare Disease Research
Genetic Discrimination: Implications for Data Sharing Projects (GeDI) (2022)	Information brief	Recommended	
Framework for Involving and Engaging Participants, Patients and Publics in Genomics Research and Health Implementation (2021)	Framework	Recommended	
Policy on Clinically Actionable Genomic Research Results (2021)	Policy	Optional	
Data Access Committee Review Standards (DACReS) Policy (2021)	Policy	Optional	
Consent Policy (2015/rev2019)	Policy	Recommended	
Data Privacy and Security Policy (2015/rev2019)	Policy	Recommended	
Ethics Review Recognition Policy (2017/rev2020)	Policy	Optional	
Machine Readable Consent Guidance (2020)	Guidance	Optional	

APPENDIX B.

International Genomic Data Sharing by HTI: Thematic Briefs

Consent

This section provides an overview of best practices for informed consent to international data sharing, according to common regulations and guidelines. It does not aim to harmonize all existing consent rules in genomics.

Consent should be freely given, informed, and ongoing. Freely given means it is clear the choice to participate is voluntary and that it is free of coercion. Informed means supported by the provision of clear and easily accessible information, including the purposes of data sharing and the types of data and of organizations involved. Ongoing generally means research participants can withdraw their consent at any time without negative consequences (and are informed of this right). There may be limits to the right to withdraw consent, e.g., if data has already been shared, analyzed, published, or anonymized. Information sheets and consent forms are typically reviewed and approved by a research ethics committee. The consent process should be documented, versioned, and subject to quality control measures, to make sure consents are accurately communicated and respected throughout data sharing ecosystems. It is also a common requirement that consent be explicit and indicated by an affirmative action such as signing a form or clicking "I agree".

Consent can be specific, broad, or dynamic. The appropriate model will depend on the nature of the protocol. Specific consent covers participation in a specific research project. Participants would need to be recontacted to provide a new consent for any subsequent data sharing beyond the research project. Broad consent covers a research area. It is appropriate when data may be used for several, future research projects that cannot be fully specified at the time of recruitment. Broad consent should be accompanied with a range of safeguards to ensure future uses respect the initial consent provisions, such as oversight by a data access committee or research ethics committee, ongoing transparency about how data are accessed and used, and offering research participants the right to opt-out from future data sharing. Dynamic consent refers to technology systems allowing continuous communication with research participants about the proposed uses of their data – supporting their decision whether or not to participate in a particular project. Dynamic consent models may also give participants various options for how their data will be used and shared (or not), which can be updated electronically at any time.

The EU General Data Protection Regulation (GDPR), personal data, and in particular special category (e.g., health and genetic data), can only be processed with a lawful basis. Consent is one lawful basis for processing personal data, and specifically for processing special category (health and genetic) data. The GDPR sets a high standard for consent to processing, which can also be withdrawn at any time. As a result, some national Data Protection Authorities advise that organizations should generally prefer alternative lawful bases for processing data in health research, which are more flexible. Consent can still be obtained as an ethical safeguard.

Providing meaningful information is becoming more and more challenging. New technologies such as artificial intelligence present opportunities to re-use data in ways not anticipated at the time of collection. Data sharing and re-use are increasingly complex and hard to meaningfully anticipate and describe, let alone fully understand. Individuals repeatedly asked for consent to share their data are likely to experience consent fatigue. These challenges can make consent a less reliable tool in the future unless organizations adopt new communication approaches and tools. In practice, organizations seeking consent should provide sufficient information to research participants without overwhelming them with details. A layered approach can support comprehension. This is where key information is provided first, with more details provided in later sections. Continued oversight of data sharing is important to compensate for any limitations inherent in the consent process.

Consent should be adapted to the context. In genomic rare diseases research, for example, sharing rare genomic and phenotypic information is imperative to recruit individuals into clinical trials, and to diagnose patients through "matchmaking" services. Rare disease patients may even expect their data to be shared – with appropriate safeguards in place – even where they are at a higher-than-average risk of re-identification. In the case of pediatric consent for genomics, the consent process and language need to be adapted. Depending on age and context, it should be determined if the minor has the capacity to consent, or if the parent or legal guardian must provide substitute consent. Minors lacking legal capacity to consent should still be involved in the decision to an appropriate degree, usually through a process of asking the minor's assent, a complementary form of permission. Language directed towards minors should be clear and comprehensible, and tailored to their age and maturity. For minors, it is important to keep in mind that their best interests are the primary consideration in all decisions concerning their well-being. Similarly, the consent process should support people with diminished decision-making capacity to be involved in decisions that affect them.

There are exceptional instances where data can be shared without consent. Common conditions are that consent would be impossible or impracticable and that appropriate safeguards are in place. There should also be little risk of adverse consequences, no evidence of refusal, and a public or general interest in the research. Transparency may require a general notification about the data sharing on an institutional or commercial website. Often a local research ethics committee or data protection authority must confirm that the conditions are met before granting a consent waiver.

'Notice and consent' is the usual consent model found online or in mobile environments. Participants are provided with terms of use and then required to affirm consent by clicking 'I agree'. Many people habitually tick 'I agree' without reading or understanding the terms. Different data uses are often bundled together as 'take-it-or-leave-it' offers. Information provided online or in mobile environments often fails to consider the realities of how humans make decisions in real-world contexts where they have limited time, insufficient resources, and lack of technical or specialized knowledge. Technology can also offer opportunities to improve consent. E-consent processes involve the use of mobile devices and informative multimedia to make consent more interactive, engaging, and flexible. E-consent is also a highly scalable process, as consent can be given remotely and individual choices can be instantly recorded.

How consent is designed can have an important impact on data sharing. Indeed, organizations intending to collect genomic data based on individual consent and to share data with external research organizations should incorporate standard consent elements. In consortia, partners sometimes agree with each other at the outset to incorporate common core consent elements, so that they may confidently share data with one another. Where consortia bring together already-collected datasets, they often need to check the existing consents cover core elements, such as those provided by the GA4GH. Where different datasets come with different consent-based access and use conditions, these need to be transparently documented and communicated. One data management solution to share heterogenous data is to convert consent conditions into machine-readable consents, allowing for rapid confirmation of data availability by data requestors and data access committees. This depends on agreeing on standard data use terms, an explanation of the meaning of the data use terms, and an optional appendix mapping consent language.

Identifiability

As the volume and variety of personal data as well as the opportunities for data use increase, so does the possibility of re-identification. The environment in which data evolves is in constant transformation and organizations must put in place methods of de-identification that are appropriate and context specific. Identifiability is a moving target and blanket methods have proven to be inadequate and inefficient. The concept of 'identifiability' itself is a matter of ongoing debate within the scientific community. De-identification of data can have an impact on the scientific value of the data set. Indeed, with less metadata attached to the data, it has less scientific value. However, de-identified data may also be exempt from obtaining consent (waiver) or ethics approval.

Identifiable information can be de-identified, which means that information cannot be reasonably used to identify someone. There are different ways to achieve de-identification: 'pseudonymised' information refers to the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information' (GDPR). It is a security measure. One way to pseudonymize data is by encoding it. Then, for data to be decoded, it needs to be combined to a specific key, that is kept separately from the data. Data can also be 'anonymized,' which means that it has been stripped of all information that could reasonably be used to identify someone. Usually, data protection laws distinguish between personal data and anonymized data. The former benefits from certain protections while the latter is subject to fewer restrictions on sharing. For example, 'Personal' information is 'any information concerning an identified or identifiable natural person,' under the GDPR.

Confidentiality, on the other hand, refers to the legal duty of data collectors to keep personal information 'in confidence'. Common security techniques to implement confidentiality include: perturbation (adding statistical noise to the dataset) and small cell suppression (removing/perturbing variables that only apply to a very small proportion of the dataset).

De-identification methods must be documented. To improve their efficiency and to ensure diligent reporting, experts should be involved in their development, review and implementation. Regardless of the chosen method of protection, it should be maintained and ensured throughout every stage of data sharing. Even with current methods available, it often only takes a limited set of data points for re-identification to occur.

Thus, de-identification of data is not infallible and entities should consider relying on robust protections beyond de-identification such as DACs, DUAs, certificates of confidentiality and designated controlled access for certain data. They should also keep up to date on recent developments in this field and perform periodic privacy assessment of their dataset. For instance, de-identified genomic data is often considered inadequately protected as it represents a potential avenue for breach of privacy.

However, in some cases, a lesser degree of de-identification may be sufficient and even desired. In the context of rare diseases, patients often wish and expect re-identification to occur. Considering the potential clinical benefits at stake, rare disease patients often consider the risks associated with re-identification as acceptable given the important benefit that may be obtained. It may also be ethically appropriate to publicly share data with identifiable information in some cases if consent was explicitly obtained from data subjects, if data subjects knowingly made their data publicly available or if disclosure of such data serves public interest, is necessary and safeguards are put in place to minimize foreseeable risks.

The practices listed above whether it be de-identified information or publicly available information share a common objective of maintaining public trust. This can be accomplished through establishing standards around the processing of personal data. Given the dynamic ecosystem in which data evolves, risks associated with re-identification need to be continuously assessed and managed. If risks of re-identification cannot be appropriately minimized, the organization should consider taking reasonable steps to destroy personal information.

Enforcement of de-identification methods across jurisdictions and across private and public entities is increasingly challenging given that data is internationally shared. Data is available on the internet from a variety of sources making it more vulnerable to breaches and facilitating data combination and, consequently, re-identification through triangulation methods. Diagnostic codes combined with anonymized genomic data and electronic health records can make it possible to uncover individual identities.

All risks associated with de-identification cannot be predicted. It is difficult to know in advance which combination or linkage of datasets may add scientific value. As entities are increasingly focused on worst case scenarios when assessing risks, data custodians sometimes end up avoiding releasing data altogether as a protection measure. As a result, entities often adopt standards exceeding legal requirements and have a tendency to apply them as a one-size fits all approach. Generic approaches to de-identification are usually unsuccessful as each dataset often requires a different combination of legal and technical measures to ensure that the risks of re-identification are minimized.

Privacy and Security

Privacy refers to the fundamental right to be left alone and protected from unwanted interference. Informational privacy refers to the ability to control access to and use of one's personal data. What counts as a reasonable expectation of privacy or a reasonable level of control over one's data will depend on the context. Privacy, while a fundamental human right, is not absolute and must be balanced against the rights of others and important social objectives pursued through legislation (e.g., protection of public health or public safety). 'Privacy' and 'confidentiality' are not interchangeable terms. Confidentiality refers to the duty of a person, professional, or organization to keep (personal) data secret. Data security refers to measures that safeguard against the risk of unauthorized access, use, change, disclosure, or destruction of data. Data privacy laws commonly hold individuals and organizations accountable for minimizing privacy risks and ensuring the confidentiality and security of personal data.

Data minimization is the principle that processing of personal data should be limited to what is necessary to achieve the objective. De-identification processes are an important way to respect data minimization in many research contexts not requiring knowledge of the identity of individual research participants. De-identification can, however, reduce the quality and scientific utility of data. The approach taken should therefore be appropriate to the particular research context. Pseudonymization techniques aim to reduce privacy risk while maintaining a secure mechanism to link back to the individual's identity, in case there is a future need for re-identification (e.g., data linkage). Anonymization techniques aim to irreversibly de-link data from an identifiable individual, and are commonly adopted in situations where data are being publicly released through an open access system. Anonymization should not, however, be approached as a one-off event ('release-and-forget'). Even if direct and indirect identifiers are removed with a high degree of confidence, organizations must remain vigilant to the possibility that data may become re-identifiable in the future as more data becomes publicly available and technology evolves. It is not appropriate to anonymize personal data obtained unlawfully; such data should be completely destroyed.

Determining the appropriate level of security safeguards should take into account the purposes of processing personal data, as well as the nature of the data and associated risks. Data security includes technical, organizational, and physical measures. Technical measures concern systems and technologies within an organization. They include pseudonymization, encryption of data at rest (in storage) and in transit, identity management and access controls restricting access to authorized individuals or entities, configuration management of hardware and software, network security, software patching, and blocking unnecessary functions and ports. Organizational measures include establishing security policies, assigning roles for security, conducting risk assessments, and providing regular training to employees. Physical measures include protections for equipment, hardware, and physical locations. The implemented measures should be ensured across the data lifecycle, and should be continuously reviewed as technology evolves and new vulnerabilities emerge. Safeguards should be routinely tested, including with independent, third-party audits or penetration testing (an authorized, simulated cyberattack).

Beyond identifying and preventing risks, data security is concerned with the detection of, response to, and recovery from security breaches. Data breaches may include unauthorized access to data, inadvertent data release, misuse of data, data theft, or involuntary destruction of data. Breach notification obligations are common in data privacy laws and data sharing agreements, requiring organizations to rapidly report

personal data breaches to various stakeholders including data providers and data privacy authorities, depending on the severity of risk to the affected individuals. Research participants should be informed of breaches likely to result in a high risk to their rights and interests, along with information about the nature of the breach, the corrective measures taken, and steps they can take themselves to mitigate risk.

Data privacy laws go beyond simply ensuring the confidentiality and security of personal data. They also typically grant individuals rights to participate in the processing of their data. These rights may include access, rectification, withdrawal of consent or objection to processing, erasure, portability, opt-out of sale, as well as non-discrimination. Participatory rights may be subject to limitations and justified exceptions e.g., rights to access in research contexts. Research participants must generally be informed of their applicable rights and any limitations. In the EU, data protection is considered a fundamental right in addition to the right to privacy. Data may be processed fairly, lawfully, and transparently. To be lawful, the processing of personal data must have a legal basis (e.g., processing based on consent, processing in the public interest as defined by legislation, and processing for an organization's legitimate interests). Processing of special category (e.g., health and genetic) data is only exceptionally permitted with the individual's explicit consent or under other exceptions (e.g., research subject to appropriate safeguards).

Several privacy challenges are specific to healthcare and health research, in particular the sensitivity of health and genetic data. Healthcare institutions often do not have the legal and technical expertise to confidently share data in a manner compliant with data protection laws. When sharing genomic and related health data with external organizations, best practices include establishing a Data Access Committee (DAC) to manage access to data, and contractual agreements with recipients to ensure privacy and security. DACs typically confirm the trustworthiness of recipient users and organizations, as well as the security of the environments where data will be processed. Recipients may only be granted access to data within designated secure processing environments. Direct access to raw data in these environments may be prohibited. Users may only be permitted to run queries or analysis workflows on the data and to export the summary results. Another technique, homomorphic encryption allows the processing of data by users while it remains in an encrypted format. International data sharing networks often establish common privacy and security frameworks or codes of conduct. This ensures standards are met acceptable to members coming from different jurisdictions and subject to different legal requirements. EU data protection law explicitly encourages the development of codes of conduct, approved by regulators, that help specific sectors demonstrate compliance through detailed interpretations and self-regulation. Various European and national codes are being developed for health research, clinical trials, and cloud computing.

Access

The COVID-19 global pandemic demonstrated the need and urgency for the scientific community to have equitable access to data, in accordance with the FAIR principles (Findable, Accessible, Interoperable and Reusable). The development of scientific data across different platforms and repositories has created a landscape where it is increasingly difficult to federate such data so as to ensure its compatibility with policies and data access/use agreements.

Access to data can be either open or controlled (a synonym of 'controlled' is 'restricted'). Open access to data improves data portability across different organizations and jurisdictions. It increases scientific collaboration and productivity as well as the validation of clinical findings. It reduces duplication. Broader dissemination of scientific outcomes can also generate greater social impact. In the case of controlled access, inclusion of data in this category and access modalities should be proportionate and justified. Examples of justification to limit access may include: protection of human rights, confidentiality, right to privacy, national security, public order, legal process and protection of intellectual property (IP). In cases where IP is at stake, data may be released in phases, according to a plan designed to respect commercial secrets and interests. This plan needs to address the purposes of research and to outline privacy and security best practices. Ideally, to maintain a seamless data flow, access to data should also be sustainable, interoperable and transparent.

One way to accelerate barrier-free access to data is digitalisation. Digitalisation allows members of the public to access, use and manage their data throughout geographical frontiers. Digitalisation democratizes access to data and facilitates data portability. In turn, it enables data subjects to exercise their protection rights over their own data: right to withdraw, right to be forgotten (when reasonably feasible) and right to rectification/explanation. However, by giving more control to data subjects over their data, they simultaneously expose themselves to greater risks of confidentiality breaches.

Controlled access to data offers a certain level of security, by restricting access to certain individuals and entities, agreeing to specific terms of use. Nevertheless, too many restrictions would hinder scientific advancement. To achieve a balance between privacy protection and sharing scientific benefits, tools such as authentication processes or encryption of data can be used. To allow data subjects to exercise their protection rights, an organization needs a clear protocol on data access. This protocol should be easily accessible (e.g., via the organization website) and communicated to data subjects. It needs to be compliant with local laws, well-documented and frequently reviewed by experts. It should be interpreted in a way that facilitates data access through flexible methods rather than restricting it. The protocol should be as uncomplicated and as inexpensive as reasonably possible. If an organization were to reject the access request of a data subject, they must provide their written reasons in a timely manner. A procedure should be in place should the data subject wishes to challenge this decision. Given the above, even in the case of a deletion request, the organization should keep a copy for their deletion records.

Some cases in which an organization may reasonably refuse to grant access to data may include: access would be unlawful, the request is frivolous or vexatious, granting access would be prejudicial to legal proceedings, providing access would have an unreasonable impact on the privacy of another individual, giving access would reveal commercially sensitive information, the organization has reasonable grounds to believe that granting access would pose a serious threat to the life, health or safety of the public or another individual.

Data Governance

Data is one of the most valuable assets for Health Technologies' Industries (HTI). Data governance refers to the ways organizations seek to realize the commercial, scientific, or societal value of data, while managing risks and complying with legal and ethical duties that come with processing genomic and health-related data. Data governance is exercised through mechanisms that control data flows and sharing, including administrative structures, roles, policies, processes, and agreements. Key themes of data (and AI) governance include privacy, protection of human rights, and research ethics. A number of mechanisms can be adopted to ensure that the value of data can be realized while also protecting the privacy of individuals (see Data Privacy & Security Brief). Research ethics concerns often significantly overlap with privacy concerns, but extend to broader questions about appropriate action, such as ensuring that data processing is pursued in alignment with the general or public interest; that risks to individuals, groups, and populations are mitigated; that conflicts of interest are identified, managed, and disclosed; that benefits outweigh risks; and that benefits are equitably distributed.

Data governance includes managing ethical and liability risks associated with analyzing health and genomic data. Inaccurate data or misinterpretation can lead to patient harms or anxiety, and wasted healthcare resources. Genomic analysis can also reveal new or updated findings of potential clinical relevance to research participants or their families, beyond the aims of the analysis, triggering legal and/or ethical duties to inform. Organizations should have a plan in place for handling such findings, including clarity over the scope of analysis services, lists or criteria for reportable findings, and clear communication processes (possibly through an appropriate healthcare professional), with consideration for how the associated costs will be recuperated.

In both data and AI governance contexts, there is increasingly reference to human rights and business ethics principles including non-discrimination, human-centeredness, fairness, accountability, and transparency. An acute fairness concern with commercial genomics is that only the wealthy can access new products and services, thus exacerbating existing health inequalities. There are also concerns about HTIs not sharing genomic data to improve broad knowledge of new advancements or that the data they provide are not representative, limiting benefits to certain groups of participants or patients. Accountability means identifying roles and responsibilities to ensure data are used both effectively and responsibly. Transparency means providing information about what data are being used, for what purposes, and who is benefitting. Privacy and ethics need not be viewed narrowly as risks that need to be managed; they can also be viewed as part of a research organization's core aims. Clinical genomics laboratories, for example, should see themselves as stewards of patient data, with a responsibility to deal faithfully with the data entrusted to their care or oversight. Key governance mechanisms to mitigate privacy and ethical risks include assigning data protection or ethics officers, adopting industry codes of conduct (articulating sector-specific principles and rules and supporting self-regulation), conducting impact assessments, adopting data privacy and security measures, and consulting stakeholders to ensure data governance aligns with their reasonable expectations.

Given that genomic and related-health data have the potential to be used for many different purposes, by many different actors, governance of use and access are key themes in realizing their societal and commercial value, as well as mitigating the risks of dealing with these sensitive data. HTI involvement in genomics should come with clear expectations about purpose limitation – that data only be used for authorized purposes, e.g., delivering care, quality control, research, and/or innovation. Access governance is a major theme for biobanks and databases acting as resources for the scientific community. The aim of access governance is to balance the scientific, societal, and commercial benefits of broad access and re-use of data, while also protecting the rights and interests of data subjects and data providers. Mechanisms include review of access requests by data access committees (DAC) – interdisciplinary committees with a mandate to ensure intended uses respect agreed-upon access policies and consents, data requested are relevant to the intended use, and recipient organizations agree to respect the approved purpose as well as the data providers' and the individuals' rights and interests. Access governance also commonly involves developing and concluding data access and use agreements to ensure recipients manage and use data appropriately.

Special conditions may be established for access by HTI, such as requirements to contribute back enriched data to the resource and by extension the scientific community. As recipients of samples and data, HTI will also be concerned with ensuring the accessed samples and data were legally and ethically collected. They may seek to confirm ethical and legal provenance by documenting and reviewing the applicable consents and approvals. HTI also have their own data assets. They may share data as part of voluntary collaborations or be required to provide access to data for regulatory or public health purposes. HTI access governance must be able to identify opportunities for collaboration as well as reporting requirements. When granting access, HTI must additionally consider and mitigate business risks through secrecy and security safeguards.

In data sharing networks, particularly those aimed at realizing the benefits of Big Data sourced from many data providers, common and coordinated access processes are needed to reduce the time and efforts associated with accessing multiple different resources. Data sharing networks may be on a spectrum from closed partnerships, to partnerships allowing accession of new members, to open databases that are publicly accessible. In membership-based data sharing networks, partners agree to certain constitutive agreements, general terms of access and use, and common access processes. Data providers may impose certain dataset-specific terms of use reflecting local regulatory, consent, or business contexts. Different access models for data sharing networks cover a spectrum from centralized to distributed. Centralized access means a single organization or data access committee (DAC) adjudicates access requests, based on requestor bona fides and the quality of proposed projects, and may also execute a single data use agreement. Distributed access means each data provider has its own DAC adjudicating access requests and enters into its own data use agreement. Different constraints on access by HTI may complicate centralization of access processes.

Data sharing networks also require common infrastructure to facilitate data sharing, such as data catalogs, common APIs, access management tools, and authentication and authorization infrastructure. Consortia must agree what infrastructure is appropriate for their data processing needs, which may entail performance, cost, and security considerations, as well as data localization rules. In federated networks, data providers keep data in their own secure processing environments, and users submit queries or algorithms to the network, only receiving aggregate results. Federated networks may improve control and security of data, but still require governance frameworks – especially where HTI are involved – to address what purposes data may be used for, and who owns the results.

Data Sharing

International data sharing between research laboratories was essential to complete the Human Genome Project, and remains an important norm for community resource projects in genomics. The COVID-19 pandemic has also highlighted the importance of rapid data sharing in response to global pandemics. Data sharing can accelerate scientific progress and improve human health: it increases the scale and statistical power of research, supports scientific coordination and collaboration, improves scientific transparency and reproducibility, and enables creative reuse. Data sharing can also unlock value and promote innovation across sectors of the digital economy. From an ethical perspective, ensuring the integrity of research and maximizing the scientific potential of samples and data is central to respecting the contributions of patients and participants.

Sharing of data and methods is increasingly expected or required by funders, journals, and regulators for clinical trials to support transparency and reproducibility. Industry has proactively addressed transparency by establishing and using clinical data sharing portals, such as ClinicalStudyDataRequest.com and Vivli. These resources are secure, but "open" in the sense that they make anonymized data available to the broader research community. Health technology industries (HTI) may sometimes find these open approaches unacceptable or too expensive in terms of data curation. A compromise may be to limit sharing to restricted collaborations with a limited set of partners. These collaborations include joint ventures, public-private partnerships, and consortia that pool genomic data linked to clinical records to conduct more rigorous and efficient research (see Industry Pharmacogenomics Working Group (I-PWG)). In short, there is a spectrum of data sharing approaches. The approach that can best accelerate scientific discovery and innovation will depend on the context. But in general, shifting towards more open approaches offers the greatest opportunities for scientific advancement and innovation.

Meaningful data sharing requires that data be FAIR: findable, accessible, interoperable and re-useable (See e.g., NIH Data Management and Sharing Policy). Policy-makers in the US, EU, and Canada are all placing greater emphasis on data management obligations in healthcare and research, as a precondition for effective sharing and re-use. This includes proactively identifying relevant data types, standards, and repositories for sharing that are suitable for a particular context. Challenges include selecting appropriate standards, prioritizing high-value datasets, and recouping the significant costs of data curation.

There is an often-stated tension between data openness and the privacy of individuals, particularly where sensitive genomic and related-health data are concerned. The need to address this situation requires substantial expertise in privacy laws, data security and data sharing on the part of HTI companies. Data sharing initiatives seek to mitigate this tension by restricting access to a defined group of collaborators, or through controlled access mechanisms, which ensure sensitive data can be accessed by any qualified and trustworthy data users for legitimate research projects, in a transparent and secure manner (see Access brief). Privacy-enhancing technologies enable analysts to draw useful results from data without actually needing access. Homomorphic encryption, for example, allows analysts to run specific operations on encrypted data. Differential privacy allows the release of statistics alongside mathematical guarantees that privacy risk has been reduced. Federated approaches allow data analysis to scale across networks of databases, without the data leaving a secure local analysis environment (data visitation). Further work

is needed to understand how these technologies impact the cost, speed and scientific utility of analyses. Ultimately, responsible data sharing relies on a combination of regulatory frameworks, appropriate governance frameworks and agreements, robust pseudonymisation/anonymisation techniques, secure infrastructure, and involvement of individuals and communities in governance.

Other barriers to data sharing relate to the lack of incentives of data holders to share data, cases of poor quality data, and commercial data provider concerns over lost competitiveness. Data sharing initiatives address these issues by providing tools and services that facilitate data sharing, as well as frameworks that address data privacy, proper attribution and security.

Where data is destined for wide sharing, the informed consent process should clearly distinguish between any data and sample processing (i.e., collection, sharing and use) directly related to the initial clinical trial, and any processing solely for future research use (e.g., biomarker studies). Furthermore, meaningful information must be provided about the scope of future research and the types of recipient researchers and organizations, as well as opportunities to withdraw (see Consent Brief).

A key ethical and legal challenge for data sharing is clarifying and safeguarding the appropriate bounds of legitimate data re-use. Determining and enforcing these limits can be challenging given the complexity and opacity of many data supply chains. There is also uncertainty about the nature of re-use opportunities that may emerge in the future, over what constitutes personal data (especially across jurisdictions) and over the scope of individual consents. Companies and data sharing initiatives seek to address this challenge through increased attention to robust industry standards around consent and transparency (e.g., Global Alliance for Genomics and Health, **Regulatory and Ethics Toolkit**), supported by a combination of due diligence processes, open licenses, and IT tools aiming to capture, track, and communicate the ethical/legal provenance of data.

Intellectual Property

Human rights instruments highlight not only the right of everyone to benefit from scientific progress, but also the right to the protection of moral and material interests resulting from scientific production (UDHR, Art. 27; ICESCR, Art. 15). These rights are in fact interrelated. Legal protections for intellectual property (IP) aim to foster innovation and creativity, with the ultimate intention of facilitating societal access to new technologies and improvement of human well-being. IP is an important consideration for the health technology industries (HTI), such as patents for pharmaceutical compounds, and various potential protections for algorithms.

IP protection systems aim to strike an appropriate balance between private profit and public access to life-saving products or information. These systems are generally quite balanced, but need to adapt to long term trends in innovation and short term shocks. Access concerns are particularly acute for resource-poor countries in the context of global public health emergencies (e.g., HIV/AIDS or COVID-19). Hence, international trade law includes flexibilities such as compulsory licensing in the face of public health emergencies allowing countries to stimulate local manufacturing. The World Trade Organization recently saw a need to go beyond these flexibilities to respond to COVID-19, and agreed on a partial waiver of

resource-poor countries' international trade law obligations relating to IP. Pharmaceutical firms can also proactively address access barriers through voluntary licensing schemes and investing in local research and biomanufacturing.

Even in wealthier countries, IP protection systems can sometimes function to stifle rather than stimulate innovation, by limiting instead of promoting access to knowledge and technology. One response to this concern is Open science, which aims to stimulate scientific advancement and innovation by promoting collaboration and sharing (see Access brief). Open science relies on the flexibilities within existing IP systems to increase the availability of knowledge for others to build on e.g., through open-licensing schemes. Effective governance can help protect the openness of scientific methods, products and data, and ensure that open access actually leads to equitable benefits. Where industry participates in or builds on Open science, it is recommended to establish shared, transparent expectations about which results of collaborations will be made openly available, and which will be implemented into commercial products. Such transparency over commercial interest and IP is also expected on the consent forms used to collect the data.

The sharing of genomic and related-health data involving industry actors can be hampered by concerns over protecting IP interests in data itself or to analysis results. The types of interests may differ depending on the data type (e.g., raw, training, clinical) as well as the context. While IP rights do not typically apply to raw data, this may change as value is added through processing and analysis. During the COVID-19 pandemic, the need for rapid access to data in a pre-competitive space was widely recognized, while strong IP protections were pursued for downstream discoveries. Commercial laboratories sharing data with community databases or granting patients access to their own data may be concerned about losing competitive advantage when giving up trade-secret databases. Some data providers and public funders involved in making data available to the private sector can require the conclusion of benefit-sharing agreements. Such agreements may entail licensing commitments, rights to partial patent ownership, or a share in commercial revenues (see Data Governance brief).

Research consortia and data sharing initiatives involving industry – such as Vivli or ClinicalStudyDataRequest.com that facilitate sharing of anonymized data from completed clinical trials – often establish constitutive agreements, policies, and data access and use agreements to address the concerns of commercial data providers and other stakeholders. Common provisions of such documents include duties of confidentiality for commercially valuable information; protecting data contributors' IP rights and interests in the data they share (including their ability to share with others); and recognizing data users' rights to obtain IP in or commercialize results. IP may be protected by providing priority data access to consortium partners (before community release) or by establishing an embargo during which user rights to publish or establish IP are restricted. As data breaches may affect not only privacy but also IP, cybersecurity and breach response protocols should foresee both types of risks. Private and public sector data contributors may sometimes require access to derived data, results, and IP (in the form of licensing commitments). Overall, data sharing initiatives' governance documents should emphasize transparency concerning issues of commercialization, IP, and benefit-sharing between partners and with external stakeholders.

Just as firms tend to see proprietary data as a competitive advantage, governments are seeking to regulate commercial data access through procurement and even legislation seeking to ensure fair and innovative data supply chains. Governments concerned over insufficient data access under contracts with the private sector may establish procurement rules and contracts to cover both access to IP (licenses) and access to data. As an example of legislation, the draft EU Data Act addresses rights of data access across parties in the data lifecycle, with the aim of allowing more parties to participate in value creation. Also addressed in proposed European legislation are government rights to access private sector data where necessary to respond to public health emergencies. Finally, the draft European Health Data Space legislation will require certain private-sector holders of health data to make it available broadly for secondary use, with appropriate protections in place for commercially confidential information. While these new developments may be a sign that things are changing, for political reasons, governments of wealthy countries have traditionally shown reluctance to limit the exercise of the rights of legitimate patent owners even when allowed by law. Still, firms should consider creative ways to manage IP in order to support both commercial and societal interests.

Partnerships

The health technologies industries (HTI) engage in various partnerships to advance genomic research and precision medicine, including with academic research institutions, scientific consortia, healthcare organizations, other companies, and patient groups. HTI have unique strengths, resources, and expertise essential for research, innovation, and improving healthcare systems. Such partnerships, however, must also adopt strategies to handle different ethical and business practices across sectors.

Transparency is a key principle for these partnerships. HTI can contribute substantial resources, as well as access to cutting-edge sequencing technologies, cloud storage and computing, data analysis tools, and patient/participant engagement platforms (e.g., to return genetic findings). Academic and healthcare organizations can contribute scientific and medical expertise, relationships with patients and participants, as well as access to samples and data. Successful partnerships clearly articulate what HTI are responsible to contribute, as well as what HTI gain in return during the partnership (e.g., access to samples and data) and in the long run (e.g., improved services, commercialization of results).

Academic researchers care about protecting academic freedom and publishing timely, independent, and high-quality science. These interests may sometimes conflict with commercial interests in secrecy to preserve competitiveness. Partnerships can manage these interests through clear publication policies and efficient review committees. They must also ensure the quality of the science is not (perceived to be) biased or tainted by commercial involvement, through a transparent commitment to scientific best practices (e.g., sharing data and other resources with the broader scientific community). Conflicts of interest should also be assiduously identified, managed, and disclosed.

HTI offer expertise in the translation and commercialization of research results. Non-commercial partners may seek assurances that the resulting benefits such as profits and IP are shared equitably. A related concern is that patients and publics have equitable access to commercial health products and services. This is particularly pertinent where HTI benefit from significant contributions from public resources and the participation of patients and participants. Partnerships must be clear about how commercialization will

serve not only private profit motives, but also economic growth and the broader public good (e.g., improved and cost-effective patient care). Benefit sharing may involve making fair or open licensing commitments, or possibly even sharing revenue streams or company equity with contributing communities.

Partnerships involving HTI must also demonstrate how they respect and safeguard the rights and interests of patients and participants. HTI will want to confirm the legal and ethical provenance of scientific inputs (e.g., human samples and data), as inappropriately obtained resources may result in liability, reputational damage, and lost commercialization opportunities. Partnerships require robust governance frameworks to ensure such requirements are respected along the scientific lifecycle (see Data Governance). Public-private research consortia often set out their shared aims and respective obligations in consortia agreements or governance frameworks (e.g., ICGC ARGO's **Ethical Framework for Partnerships with Industry**; **Innovative Medicines Initiative**), supported by strategic committees with cross-sector representation. HTI may also set out their own principles or codes of conduct in areas like use of health data (e.g., **Association of the British Pharmaceutical Industry**).

Some partnerships with HTI will focus on developing infrastructure to support precision medicine research and healthcare, such as cloud services. These partnerships may need to pay close attention to matters of privacy and security, including over cloud provider access to data and use for their own purposes, or access to data by third country law enforcement through laws specific to the commercial cloud sector. Partnerships involving commercial infrastructure also need to ensure long-term sustainability and affordability. Free or low-cost services upfront may be significantly increased once research and healthcare organizations have come to depend on the commercial infrastructure. A lack of data interoperability or competition may make it difficult to change service providers. These issues may be addressed by transparency over criteria for determining service costs in the long run, as well as contractual or regulatory requirements to ensure data is stored in an interoperable form to enable portability between service providers.

HTI generally face a lower level of public trust than academic and healthcare sectors. This trust gap must be addressed at a range of levels. Informed consent documents must clearly explain the value of commercial involvement and the possibility of commercialization. HTI should participate in raising public awareness about the value of its role in precision research, innovation, and healthcare. HTI should also seek to involve patients in the design and oversight of research (or equally participants, members of the public, and members of vulnerable groups). Patient engagement refers to the meaningful and active collaboration of patients in the governance, priority setting, and conduct of research, and the dissemination of results. This includes establishing frameworks for fair benefit sharing. HTI partnerships with patients should strive to ensure that the views solicited are representative of communities, with safeguards in place to ensure transparency about funding and conflicts of interest.