HEALTH DATA GOVERNANCE PRINCIPLES

(Draft for Public Consultation)

Universalizing the benefits of health digitalisation

Protect people | Promote health value | Prioritize equity

Transform
Health

Health for all in the digital age

Introduction

Over the last several decades, the collection and use of health data have grown in complexity and have become increasingly indivisible from the operation of health systems and the provision of health services. With this increase in data use, the need for robust and equitable governance of health data has also increased. More countries and regions around the world are instituting health data governance policies and legislation—but there is not yet a comprehensive, globally unifying set of principles to inform such governance, especially one that focuses on equity.

Universal Health Coverage (UHC) must be a core aspect of the design and development of data-driven health systems. Progress towards UHC has also guided the development of the Health Data Governance Principles outlined here. These Principles bring a human rights and equity lens to the use of data within health systems and are oriented toward supporting sustainable and resilient public health systems.

This document presents a set of health data governance principles as an initial step toward a global framework for health data governance. They are meant to inform, strengthen, and unify governance instruments, models, treaties, regulations, and standards across countries and regions through a shared vision of health data governance.

They recognize and build on existing principles, such as the World Health Organization's data principles and their guidance on the [ethics and governance of artificial intelligence for health](#); the [Principles for Digital Development](#) and [Digital Investment Principles](#) stewarded by the Digital Impact Alliance; the Organisation for Economic Co-operation and Development's [Recommendation on Health Data Governance](#) and their [Principles on Artificial Intelligence](#); and the report of the [Lancet and Financial Times Commission on Governing Health Futures](#).

Objectives of the Health Data Governance Principles

The Principles are framed around three interconnected objectives:

> (1) *protect people* – individuals and communities;

> (2) *promote health value* – through data sharing and innovative uses of data; and

> (3) *prioritize equity* – equitably distribute the benefits that arise from the use of data within health systems.

Most current approaches to data governance take an individualistic view. However, the following Principles balance both individual and collective perspectives within each of the three objectives. **Protect people** considers the importance of group and community data protections. **Promote health value** speaks to the collective needs

and benefits of public health systems. **Prioritize equity** requires equity among groups in addition to individuals.
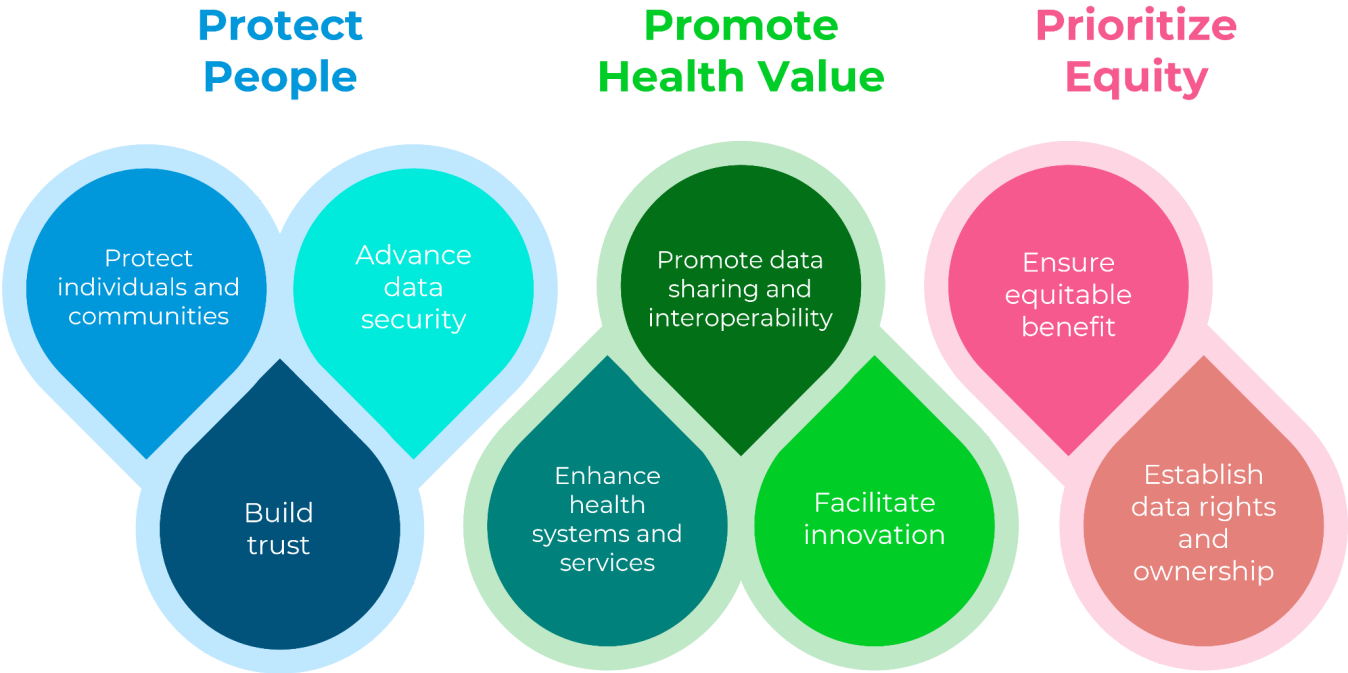
These Health Data Governance Principles were developed in a bottom-up manner, employing an inclusive and consultative process with over 185 thought leaders through global and regional workshops across Sub Saharan Africa; the Middle East and North Africa; South, East, and Central Asia; Latin America and the Caribbean; and Europe, North America, and the Pacific. This process was designed to gather perspectives and expertise from different geographies, sectors, and stakeholders.

As a result, the Principles are intended as a resource for a variety of stakeholders involved in the governance of health data, including governments, international organizations, the private sector, nonprofit organizations, research and academic institutions, foundations, civil society, and global coalitions.

Stewarded by [Transform Health](#), the Health Data Governance Principles are primarily driven by civil society to advance equitable health data governance. The support of many partners has been instrumental in their development, including PATH, AeHIN, the BID Learning Network, Mwan Events, RECAINSA, Wilton Park, and Young Experts: Tech 4 Health.

Principles

The Health Data Governance Principles are designed to complement and reinforce one another. As such, the Principles are not weighted or listed in any priority order. Each Principle is supported by *core tenets* that further describe it and how it may be put into practice. As mentioned, the Principles are clustered in three overarching objectives: *protect people*, *promote health value*, and *prioritize equity*.

Protect People

Promote Health Value

Prioritize Equity

Protect individuals and communities

Advance data security

Build trust

Promote data sharing and interoperability

Enhance health systems and services

Facilitate innovation

Ensure equitable benefit

Establish data rights and ownership

Protect People

Health data governance must ensure protection for people and communities against harm. Protection for individuals is often embodied in general data protection laws. However, health data can be very sensitive, and the misuse of such data can cause significant longstanding harm. Further, unprotected individual and aggregate data could expose both individuals and communities to harm. Therefore, health data of all types require special measures of protection against harm.
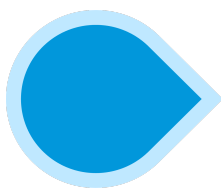
Protect individuals and communities

Health data governance must address the protection of individuals and communities at every stage of the data lifecycle. Data governance should seek to balance the protection of individuals and communities with the societal value of data used for health.

Core Tenets:

Collect data that are purpose-driven
Before any data collection activities, the specific data needs should be defined. This should be communicated to the individual and communities providing their data and to the agencies collecting it. Health data governance should include guidelines on determining data needs and what limits to data collection should be applied (e.g., only collect what is needed, use existing data where possible).

Collect personally identifiable or sensitive data only as appropriate
While health data necessitates collecting personally identifiable information and other types of sensitive data, effort should be made to collect only that which is necessary. Comprehensive, evidence-based guidelines must weigh the consideration, identification, and prompt action related to potential risks against the data needed to provide individual healthcare and other defined data needs.

Sensitive health data must be collected only to achieve the research, policy, advocacy, or medical objective in question (e.g., electronic health records will include sensitive data but are required to improve patient care outcomes). Appropriate security and privacy measures must be ensured when personal data are collected.

Some data use scenarios may warrant added protections to prevent unauthorized use or access to sensitive information.

Use secure data collection, storage, and dissemination mechanisms

Protection of health data should be considered before any data collection—and should include secure methods of collecting the data (e.g., using data collection tools with robust data protection functionality) and secure data storage (e.g., encryption, cloud servers). Consideration should be given to how long data are stored, with guidance on a reasonable timeframe after which data should be deleted or otherwise removed from the system. Because personal health data are "lifetime" data, the data retention policies related to care records should not create gaps in a longitudinal health record.

Secure data collection and storage must be complemented by secure data dissemination and transfer. Comprehensive data security policies must cater to both data storage and transfer approaches used today (e.g., USB drives, external hard drives, routers, servers, databases) and the continuously evolving health innovation ecosystem. Such policies necessitate political commitment to building and strengthening appropriate infrastructure and multi-stakeholder collaboration to implement these policies.

Address individual and collective risk

Health data governance should prioritize the reduction of individual and collective risk, following the doctrine of "do no harm." Particularly when collecting data from marginalized communities, the collection and use of health data must mitigate potential risks an individual may face. Risks to an individual can range from severe (e.g., risks to personal safety, risks of insufficient or incorrect care) to moderate (e.g., breaches in data security). When data are not personally identifiable, health data governance should mitigate collective risks, including those related to a specific group or community (e.g., risks of discrimination) and those relating to the broader society (e.g., risks to public health).

Define inappropriate uses of health data

Health data governance should not allow inappropriate, unethical, or otherwise unlawful collection or use of health data. This may include non-health-related surveillance by state or other actors or discrimination and harassment by public or private stakeholders, especially against marginalized populations. National and global governance frameworks, guidelines, and principles relevant to the digital age may be necessary to respect, protect, and uphold fundamental human rights while collecting, storing, and using health data.

Use de-identification and anonymization

Health data governance should define the level and mechanisms for protecting an individual's identity. This must be applied when data are aggregated or shared but should be considered at every stage of collecting and storing data, particularly for sensitive data and data from marginalized communities. Health data governance

should also address the possibility of reidentification (e.g., by data analysis algorithms or triangulation of data sources).

Provide guidance specific to marginalized populations

Protecting individuals and communities requires data governance practices that are responsive to their unique contexts and data-related needs. Existing recommendations and other guidance specific to marginalized populations must be incorporated into health data governance policies and processes. For example, UNICEF has produced a [manifesto on the better governance of children's data](#).
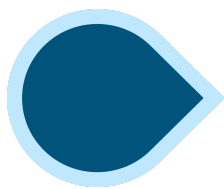
Build trust

Health data governance must reinforce trust in data collection and use and in data systems. Trust requires safeguarding data, preserving privacy, and establishing transparent and inclusive data collection, analysis, and use processes.

Core Tenets:

Align with data protection and privacy best practices

Health data governance should apply existing best practices—and establish new best practices—to protect data and individual privacy. This includes both technical approaches to data collection and storage (e.g., two-factor authentication, encryption, de-identification) and policies and processes related to how data is accessed and used (e.g., security policies, system permissions). Health data governance should align with well-established policies and regulations such as the [General Data Protection Regulation (GDPR) in Europe](#), the [Personal Data Protection Act (PDPA) in Singapore](#), and the [Protection of Personal Information Act (POPI Act) in South Africa](#).

Establish transparent processes and systems

Transparency in health data governance can drive higher-quality insights from data, create buy-in from stakeholders around data processes, and enable greater collaboration on and reuse of data analysis methodologies. The [Data Futures Partnership](#) in New Zealand defines transparent data use with [three dimensions](#): value, protection, and choice. All stakeholders involved should understand how and why data are collected (value); how data are stored, analyzed, and used (protection); and how the systems and processes that support health data governance operate (choice).

Require informed consent and articulate its complexities

When collecting an individual's data, the individual contributing their data has a right to understand what data are collected, why these data are collected, and their rights regarding accessing, changing or removing their data from the system. Individuals should clearly understand how their data inform personal care and whether their data may be reused for additional purposes.

Individuals should also have a reasonable option to accept or decline data collection as appropriate, as well as the option to accept or decline further sharing of their data for purposes other than its initial intended use.

Informed consent is the gold standard for health data governance. However, health data governance policies and processes should acknowledge the complex reality of informed consent, particularly for marginalized populations. For example, an individual may be required to provide their data to receive health services, so may feel compelled to consent regardless of their understanding of or agreement with how their data may be used. In all instances, the protection and agency of an individual or community should be the highest consideration. Health data governance should therefore articulate a nuanced view of consent.

Define concrete exceptions to informed consent requirements

Health data governance policies and processes should clearly and transparently define circumstances where exceptions to informed consent are allowable. Exceptions should be limited to those instances where individual consent requirements pose an acute barrier to public health (e.g., during public health emergencies), when data cannot be considered "owned" by an individual (e.g., metadata), or when data are shared in aggregate (e.g., population-level data). Exceptions to informed consent should not be used to circumvent an individual's control over their data and should only be used when a clearly defined communitarian value exists with limited or no individual risk.

Reinforce health data governance with evidence

Health data governance should be informed and reinforced with evidence of impact, both positive and negative. Governance should be evaluated against best practices and updated when evidence suggests improvements could be made. When gaps in existing knowledge are identified, efforts should be made to address these gaps and contribute to the global evidence base regarding health data governance.

Improve data quality, availability, and accessibility

Trust in data systems also requires trust in data themselves. Health data governance should support the overall improvement in data quality and make data more available and accessible, as appropriate.

Institute feedback and accountability mechanisms

Inclusive, equitable, and accountable health data governance requires mechanisms through which individuals or communities can report data misuse, make inquiries into health data structures and processes, and remove their data from a system. This

requires the implementation of responsible data practices, institutional procedures, and clearly defined roles and responsibilities of all stakeholders.
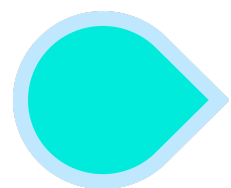
Advance data security

Data security is an essential component of health data governance. Data security is closely linked to the earlier Principle on building public and institutional trust among citizens and organizations. It encapsulates multiple dimensions, including recommending good practices on the collection, storage, use, dissemination, analysis, and disposal of health data. Components of data security are reflected across the Health Data Governance Principles, but data security also serves as a foundational principle itself.

Core Tenets:

Require strong security measures for data collection, storage, and use technologies

Any technology used to collect, store, analyze, or use data should employ robust security mechanisms. This may include password requirements, two-factor authentication, security keys, and data encryption. In addition, health data governance should address common security risks, including phishing and viruses.

Mitigate risks related to security breaches

Health data governance should consider how to minimize the impact of potential security breaches on individuals, communities, and health systems. This may include using unique identifiers in place of an individual's name, placing limits on how long data may be stored, and adding enhanced security measures for personally identifiable or otherwise sensitive data.

Require transparency around data breaches

When data breaches occur, health data governance should require stakeholders to inform individuals and communities affected and report the breach to any concerned regulators. Information should be provided on the nature of the breach, what data may have been exposed, and specific actions that were taken to address and prevent a similar breach in the future.

Consider federated data systems

Data security, data rights, and data ownership may be achieved through federated storage and use of data. Federated data systems bring together multiple, autonomous data sources to allow cross-system sharing and learning while

appropriately adapting good data practices across different sectors. This approach may maximize the value and use of data insights and create new opportunities to generate insights from multiple stakeholders across sectors and mandates.

Promote Health Value

Health governance must maximize the value offered by the use of data to improve health outcomes. Data being a non-rival resource must be shared widely to maximize health value. In any case, much of data's value lies in patterns found in data, which underlines the need for sharing and aggregating health data. Data silos can lead to under-optimal production of health value.
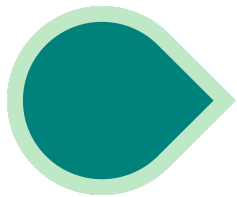
Enhance health systems and services

Health data governance should enhance health system efficiency and resilience and contribute to the improvement and expansion of health services—including ongoing efforts to advance health equity and Universal Health Coverage. The benefits of well-governed health data should especially be felt by individuals and communities who contribute their data.

Core Tenets:

Evaluate the benefits of health data
When defining the potential use of health data, consider how data may benefit the individual and communities contributing data, and the general societal good. For example, data may be used by research institutions, including academia, for research and development purposes. Ensure these stakeholders have appropriate, secure access to data and that individuals and communities who contribute data understand how their data may contribute to research and development.

Enhance health services for individuals and communities
Health data governance should enhance the use of health data to improve health and well-being, including that of individuals and communities who provide their health data. This may be accomplished in various ways (e.g., improving health services and primary health care, providing better diagnostics and predictive analysis, developing precision medicine approaches). Enhancing individual care and ensuring patient safety requires data sharing between health facilities and health providers to support a continuum of care. Data sharing is also needed to support public health informatics and data-led actions. Data sharing and access policies should reflect these needs.

Consider all types of health benefits

In addition to improved health services, health data offers opportunities to improve the efficiency, effectiveness, and sustainability of health systems. Data use also creates opportunities for innovation and advancements in medical sciences. Health data governance should facilitate these and other societal benefits while also protecting individuals and communities.

Encourage data-led insights and action

Health data adds significant value to health systems and services, leading to personal and public health improvements. Health data governance should encourage the uptake and use of data insights to address health system challenges, health inequities, and the improvement of health services. This requires a strong culture of data use at all levels of a health system. Governance should be designed in such a way to build confidence in data users and decision-makers, not only that the data used are of high quality but also that the data are ethically obtained and managed.

Engage the general public around health impact

As part of building inclusive and collaborative data ecosystems, health data governance should encourage greater engagement with individuals and communities to understand how their health data contributed to health impact and social good. This will reinforce trust in data systems and support a more robust data culture.

Reinforce health system resilience

Appropriate health data governance is a prerequisite to a resilient, responsive health system and can improve the efficiency and effectiveness of health services. These benefits extend to all operational components of a health system (e.g., supply chain, health workforce management). Health data governance should include operational improvement as a necessary use case when defining data needs.
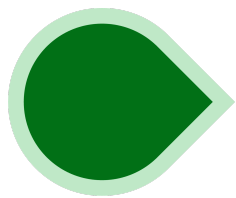
Promote data sharing and interoperability

Data sharing is a prerequisite for creating value from health data. Data sharing at national, regional, and global levels allows for deeper and more significant insights related to health needs and challenges, including preventing and responding to health emergencies.

Core Tenets:

Promote interoperability of data systems

Data, and the digital health systems that support their collection and use, should be designed with interoperability in mind. Such interoperability will make sharing data between systems simpler and more secure while preventing potential errors during manual data transfers. Interoperability is accomplished through technical approaches (e.g., global data standards) and system design (e.g., use of open APIs). Concepts like data portability, open data, community data, data trustees, and data exchanges may also be applied.

Establish data sharing policies

Health data governance should include data sharing policies that address a range of potential data sharing scenarios. This includes data sharing required for individual care delivery, between government ministries within a country, between government systems and the private sector, within the private sector, and between national, regional, and global stakeholders. Data sharing policies should minimize both individual and collective risk and enhance public health equity.

Support multi-sector partnerships

Health data governance should support partnerships between national governments, the private sector, academic institutions, civil society, non-governmental organizations, and other types of stakeholders to create a resilient ecosystem of data collection, curation, and use. Clear policies are needed for engaging with these stakeholders. These policies should prioritize the interests of individuals and communities providing data and societal interests, especially of public health equity.

Validate informed consent before sharing data

Data should only be used for the purpose for which it was collected unless informed consent exists from the individuals who contributed their data. This may require gathering additional consent from individuals who previously contributed data or having clear data-sharing agreements in place before data collection.

Define common data structures across health systems

Common data structures (e.g., specific fields that are collected, the underlying architecture of data systems) will support the use of emerging technologies by allowing for easier consolidation of data from a variety of data systems. Such structures enable interoperability and opportunities for insights that add efficiency to medical sciences and improve health outcomes. This common structure can serve as a foundation for understanding the types of data available for future use.

Use common definitions

Health data governance should build on existing definitions of key terminology and roles. These definitions are provided by various global normative bodies, and efforts should be made to align to these definitions whenever possible.

Use global data standards and coding

Existing data standards and frameworks (including [ISO/TS 22220:2011](#), [HL7 FHIR](#), [OpenHIE](#), [GS1](#), and others) should be applied whenever possible. This promotes greater standardization and comparability of health data, which allows for greater systems interoperability, data sharing, and data hygiene.

Define multiple levels of access to data

Health data governance should outline limitations on data access—identifying which stakeholders will have access to various levels of de-identified or aggregated data. These permissions may exist on a technical level (e.g., systems permissions) or a policy level (e.g., defining "data stewards" vs. "data users"). Each level of access should include clearly defined roles and responsibilities.
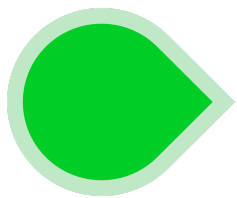
Facilitate innovation

Health data governance should be forward-looking and anticipate (wherever possible) the application of emerging technologies such as artificial intelligence. Leveraging the continual innovation of digital technologies and data systems is key to attaining the Sustainable Development Goals and UHC. This requires supporting a governance environment that can flexibly accommodate or enable innovation and be applied to new digital technologies and new kinds of data uses.

Core Tenets:

Address non-health data that may support public health equity

Many applications of data science and emerging technologies utilize data from sources beyond health systems. Health data governance should consider other types and sources of data that may be combined with health data. When combining health data with data from other sources, the intended use of this data must be clearly defined and uphold the intent of health data governance (e.g., promoting health equity and protecting the individual). Combining data from multiple sources may create additional risks to the individual (e.g., reidentification), which must be mitigated. However, such data may also offer new or more impactful opportunities for health impact.

Support policy innovation

In some instances, policy and legislation may lag behind the technological capacities

of data collection, analysis, and use. Policy innovation may be necessary to support new developments within health systems, particularly around new mechanisms for data analysis (e.g., machine learning) and the use of emerging technologies. Such policy innovation must guide the use of emerging technologies to promote the appropriate use of health data.

New policy approaches may be needed to support innovation (e.g., the development of precision medicine or the application of big data to the development of medical devices or pharmaceuticals). A comprehensive value chain approach may also guide appropriate policy innovation across the data life cycle: from collection to storage to use.

Apply health data governance to emerging technologies

Emerging technologies are not exempt from health data governance. Their application should consider governance policies, legislation, and principles from the design and ideation stages. Mechanisms should be defined to address conflicts between existing health data governance and the needs or impact of emerging technologies. Sandbox facilities for controlled testing of technical and business innovations may be useful in this regard. In the face of rapid technological development, more general guidelines may be needed in addition to specific guidance that addresses current and known technological contexts, such as the WHO Guidance on Ethics and Governance of Artificial Intelligence for Health.

Prioritize Equity

Health value created by the use of data must equitably benefit individuals and communities. Data are contributed by people, whether as individuals or as communities and so people should have an equitable stake in the health value that their data generates.
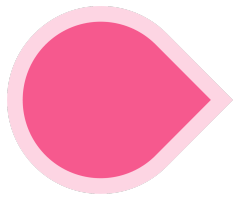
Ensure equitable benefit

Equity must be inherent to health data governance—ensuring equitable representation in data of all individuals and communities, regardless of social or economic characteristics, and equitable access to data-generated health value. Equitable health data governance reinforces population-sensitive applications of health data within health services and systems; it also promotes the equitable sharing of data-led health service improvements and innovation. Further, equity in health data governance must extend beyond policies, processes, and outcomes to include outward public engagement, education, and meaningful involvement of all groups.

Core Tenets:

Represent all populations

Health data should be inclusive of all populations, regardless of social (e.g., age, gender, race, sexual orientation, refugee status) or economic (e.g., income status, profession) characteristics. This requires inclusive data collection methodologies and processes that consider which individuals are asked to provide data; which data categories are collected; and the intended use of collected data.

Consider the unique needs of marginalized populations

Populations that have been traditionally marginalized may have unique needs related to the governance of health data. For example, children may not have the same legal authority to provide consent. Providing information on sexual orientation can place individuals at risk for arrest or violence in some contexts. Considering these unique needs requires marginalized groups' active and meaningful involvement in the development, implementation, and review of health data governance policies and processes.

Reduce and mitigate bias

Bias can be introduced at any point in the collection, analysis, and use of data. Such bias can perpetuate inequities, undermine the veracity of data, and lead to incorrect or incomplete conclusions drawn from data. Bias can also lead to discrimination and exclusion, intentionally or unintentionally. Health data governance should aim to counteract discrimination and identify where bias may be introduced. It must provide mechanisms to address existing biases within data systems and protect against the misuse of health data that reinforces bias.

Implement inclusive data feedback mechanisms

Feedback mechanisms should be established so that individuals and communities involved at every stage of data collection, analysis, and use are aware of how data is used. Particularly, individuals and communities who provide their data and individuals involved in the collection of health data (e.g., frontline health workers) should understand the purpose and outcomes of data use. Individuals must also have agency over their data and be capable of making appropriate decisions for themselves. Feedback loops will require data accessibility and channels of communication at multiple levels.

Use accessible language and address knowledge gaps

Health data governance should be understandable to the general public—including children, individuals with low literacy, and individuals who speak minority languages. While specific legal or technical documents may be needed to legislate or operate health data governance, supporting resources should enhance individuals' and communities' understanding of their rights in a practical, actionable way. Efforts should increase the general public's knowledge of health data governance and how it may impact them on an individual and community level.

Promote equitable impact and benefit

Health data governance should ensure that the benefit of data use and data-based health systems is equitably shared across all populations, regardless of social, economic, or political characteristics. This may be a matter of the design of data-based health systems and their reach and accessibility. Benefits gained from data use should especially be shared fairly and equitably with individuals and communities who contribute data. In addition, health data governance should specifically enable and encourage the use of data to enable equitable and high-quality health services for marginalized populations.
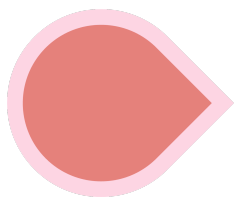
Establish data rights and ownership

Health data governance should be rooted in strong and clear data-related rights. Data-related norms, principles, policies, and laws should be drawn from such overarching rights. This includes consideration of all human rights, including the right to protection and safety, and the right to benefit equitably from data contributed, both at individual and community levels. Data ownership implies that individuals and communities have a right to know, determine, and control how their data are used and to benefit equitably from such data. Such rights extend to products and services derived from data, such as AI. Health data systems, and their governance, should be designed based on such data rights and ownership.

Core Tenets:

Apply a human rights lens to health data governance

Human rights — as expressed in documents such as the [Universal Declaration of Human Rights](#), the [International Covenant on Civil and Political Rights](#), and the [International Covenant on Economic, Social, and Cultural Rights](#) — must be central to the definitions of data rights and ownership. Many rights, including those normally associated with data (e.g., privacy), apply to many data use scenarios. For example, rights or protections for women or workers could be applicable in a specific data use context or process. The rights of marginalized populations should be given particular consideration.

Codify data rights and ownership

Data rights and ownership should be codified in legislation and policy in alignment with national, regional, and global norms, policies, laws, and regulations related to collecting and using data and accessing benefits from data use. This should include definitions of ownership (e.g., health data are owned by the individual or community providing the data) and related rights (e.g., the right to control the use of data, the right to decline participation in data collection, the right to withdraw data from a system). These definitions should be linked with the defined roles and responsibilities of health data stakeholders. Further, health data governance should outline and provide the mechanisms for exercising such rights.

Extend data rights and ownership to products and services

Data rights and ownership extends beyond data themselves to related products and services, such as AI. Because data should not be used to harm individuals or communities, the products and services derived from such data should also not be used to cause harm. Similarly, individuals and community ownership over their data extend to the benefit from the products and services built on their data.

Define clear roles and responsibilities

Roles—including data owner, data custodian, data processor, data steward, and data use beneficiary—should be clearly defined within health data governance. These roles should include clearly defined responsibilities, particularly data privacy and protection and benefit-sharing. The definitions used in existing data governance

guidelines such as GDPR and PDPA can be adapted or used as a framework for other contexts.

Use participatory data governance mechanisms

Once collected, data of individuals and communities are often reused for multiple purposes in a continuing fashion. Therefore, health data governance must provide ongoing mechanisms for individual and community participation. This may include mechanisms for reaffirming consent when new data needs are defined or mechanisms to address long-term concerns or inquiries from individuals and communities.

Connect to broader accountability mechanisms

Health data governance should be integrated into existing public accountability mechanisms to reinforce adherence to health-related policies and laws. In addition, certain types of health data may be useful to the monitoring and accountability efforts of communities and civil society.

Acknowledgments

Heather, Alison Dunn, and Clare Ashurst from Wilton Park; and to Richard Dzikunu from Young Experts: Tech for Health.

3. Transform Health's Data Policy Circle organized two multi-stakeholder consultations: one in December 2020 to unite other global coalitions and collaboratives involved in health data governance; and in September 2021 to engage Transform Health partners not otherwise involved in other convenings.