Global Alliance for Genomics and Health: Data Privacy and Security Policy

Preamble

This document is the Global Alliance for Genomics and Health's (GA4GH's) Data Privacy and Security Policy. It applies to data sharing in the context of both clinical practice and health research involving humans. This policy builds on the mission of the GA4GH and provides greater detail to the GA4GH *Framework for Responsible Sharing of Genomic and Health-Related Data* (the "*Framework*").

Compliance with this policy facilitates the processing of data in a way that promotes and protects privacy and security in a proportionate manner. It also facilitates compliance with the obligations and norms set by international and national laws (including tribal, indigenous, and aboriginal laws), regulations, policies, and interoperable standards. Guidance covering specific aspects such as withdrawal of consent for processing data, benefit sharing, and return of individual results are outside the scope of this policy.

I. Context

1. <u>Purpose</u>. Supplementing the *Framework* and complementing the GA4GH Security Technology Infrastructure, the purpose of this policy is to provide principled and practical guidance for processing data in a way that protects and promotes the security, integrity, and availability of data and services, and the privacy of individuals, families, and communities whose data are processed.

   To achieve this purpose, there is a need to**:**

   i. Ensure data subjects are informed, as far as is practicable, about the manner in which their data are being used and for what purposes (which may include broad purposes); data protection and security measures and risks; and their rights in relation to their data;

   ii. Respect data subjects' expectations, interests, and rights in relation to their data;

   iii. Protect against the risk of unauthorized access, use, change, disclosure, or destruction of data; and

   iv. Promote the availability of and access to data.

2. <u>Interpretation</u>.

   2.1 This policy is intended to be flexible enough to adapt to different contexts, cultures, and countries, as well as different technologies that can impact the privacy and security of data.

This policy also should be interpreted in a manner that acknowledges different levels of risk tolerance and community cultural practices and, where appropriate, different contexts of data processing.

2.2 To encourage broad application, this policy adopts the normative term "should" in most instances of the procedural guidance in section II. However, it is acknowledged that depending on the context, certain parts of the procedural guidance may be required or in contrast, be discretionary under law, regulation, or another instrument.

2.3 This policy distinguishes privacy from security. Privacy is treated as a fundamental value and right that protects all aspects of the lives of individuals, families, and communities, and that establishes reasonable limits to processing data. Security refers to the process of protecting data from unauthorized access, use, change, disclosure, and destruction. It also includes the protection from data corruption throughout its lifecycle by considering appropriate network security, physical security, and file security. While the privacy and security aspects of this policy might be addressed to different persons and organizations with diverse responsibilities, this policy should be read as a whole and with the overarching objective of ensuring that these crucial elements work together to deliver responsible sharing and processing of data.

3. Definitions.

3.1 The following definitions are intended to align with the *Framework*, GA4GH Security Technology Infrastructure, and other GA4GH policies. They are not intended as a substitute for definitions found in relevant laws or regulations.

"*anonymized data*" means data that are rendered anonymous in such a way that the data subject is not or is no longer identifiable.

"*controlled access*" means a data access model whereby qualified researchers apply for data access and their research plans are reviewed, often by a committee. Also known as managed or restricted access.

"*data*" means genomic and health-related data. These include data on the health status of individuals and data on non-medical determinants of health, such as health behaviors, living and working conditions, personal resources, and environmental factors. These also include data relating to the genetic characteristics of an individual which have been either inherited or acquired during prenatal development, as they result from an analysis of a biological sample from the individual concerned, in particular chromosomal, DNA, or RNA analysis, or analysis of any other element enabling equivalent information to be obtained.

"*data breach*" means a security incident that has affected the confidentiality, integrity, or availability of data, including accidental or unlawful destruction, loss, alteration, unauthorized disclosure of, or access to, data.

"*data privacy impact assessment*" means a formal process designed to help data stewards systematically analyze, identify, and minimize the data privacy risks of a project or plan.

"*data protection officer*" means an expert in data protection within an organization who ensures, in an independent manner, that an organization applies relevant laws, regulations, and guidelines protecting individuals' data.

"*data steward*" means an entity responsible for assuring the quality, integrity, and access arrangements of data from the moment of data collection, and for managing the metadata that preserves context and associated business rules, including privacy and security attributes consistent with applicable law, institutional policy, and individual permissions.

"*data subject*" means the individual whose data have been collected, generated, held, used, or shared.

"*data user*" means individuals or organizations who are authorized by data stewards or other competent persons or organizations (e.g. research ethics committees, data access committees) to access and use data for an authorized, bona fide purpose. Data users are secondary users of data that are distinct from the primary data generating research team.

"*Framework*" means the GA4GH *Framework for Responsible Sharing of Genomic and Health-Related Data*.

"*GA4GH*" means the Global Alliance for Genomics and Health.

"*identifiable data*" means data that may reasonably be expected to identify an individual, alone or in combination with other data.

"*Identity and Access Management (IAM)*" means a set of business processes and supporting technologies that enable the creation, maintenance, use, and revocation of digital identity. IAM includes identity proofing, credential issuance, rights authorization, identity authentication, and privilege revocation. IAM practices make sure that the right people gain access to the right services and data at the right time, as well as making it safe, secure, and simple to change access rights, group memberships, and other key attributes as users and systems grow, change, are added, or are removed.

"*key*" means a piece of data that an encryption algorithm uses to determine exactly how to unscramble pseudonymized data.

"*logical access*" means control measures used for identification, authentication, authorization, and accountability in digital systems, programs, processes, and information.

"*metadata*" means data that provides information about other data.

"*organizational members*" means the organizations that are a member of the GA4GH.

"*policy*" means, unless otherwise specified, the GA4GH Data Privacy and Security Policy.

"*processing*" means any operation or set of operations which is performed on data or on sets of data, such as collection, recording, organization, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure, or destruction.

"*pseudonymized data*" means data which have been processed in such a manner (e.g. by assigning one or more random codes) that the data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organizational measures to ensure that the data are not attributed to an identified or identifiable natural person. Also known as coded data.

"*registered access*" means a data access model whereby qualified researchers apply for data access to one dataset or multiple datasets at once by providing details of their identity for authentication and agreeing to terms and conditions of data use during the registration process.

"*REWS*" means the Regulatory and Ethics Work Stream of the GA4GH.

"*security risk assessment*" means an objective analysis of the effectiveness of the current security controls that protect an organization's data.

"*supervisory authority*" means the public authority (or authorities) in a jurisdiction responsible for monitoring the application of the administrative measures, laws, and regulations adopted within their jurisdiction pursuant to privacy, data protection, and data security.

"*vulnerable persons/populations*" means individuals or groups that have a greater likelihood of being denied adequate satisfaction of some of their legitimate claims to (i) physical integrity, (ii) autonomy, (iii) freedom, (iv) social provision, (v) impartial quality of government, (vi) social bases of self-respect, or (vii) communal belonging.[1]

    3.2 Words imparting the singular number shall include the plural and vice versa.

---

[1] Adopted from Tavaglione N, Martin AK, Mezger N, Durieux-Paillard S, François A, Jackson Y, Hurst SA. Fleshing out vulnerability. Bioethics 2015; 29(2): 98-107.

4. <u>Intended Audience</u>. It is expected that this policy will be useful to all persons and organizations providing, storing, accessing, managing, or otherwise using data, and in particular the organizational members of the GA4GH. These persons and organizations include, but are not limited to, researchers, research institutions, research participants and patient communities, research ethics committees and data access committees, journal editors and publishers, research funding agencies, data protection supervisory authorities, hospitals, clinicians, industry, ministries of health, and public health organizations.

II. Data Privacy and Security Procedural Guidance

Data privacy

Privacy is a fundamental value and right of human societies. It extends to all aspects of the lives of individuals: the social, cultural, religious, political, physical, and the informational. Its protection also promotes other core human values and human rights. However, privacy is not an absolute right. Privacy protection involves the delicate balance of considerations at individual, familial, and societal levels. The following guidance assists in determining such balances relative to the protection of the core interest at stake and the Foundational Principles at the core of the *Framework*.

<u>Lawfulness of data processing</u>

- All data should be processed in accordance with all applicable laws, regulations, norms, and guidelines and should only be disclosed in situations where consent has been provided, or there is a legal or legitimate interest/appropriate need for that disclosure/use.

<u>Data privacy risks and safeguards</u>

- Assessments of data privacy risks should include disclosure risks, and any harms reasonably likely to occur in the event of disclosure. These disclosures may result in individual or group discrimination, stigmatization, profiling or categorization that leads to unfair, unethical or discriminatory treatment contrary to human rights. The reputational risks for persons or organizations of allowing particular uses of data should also be considered.

- Data privacy safeguards should be proportionate to the sensitivity, nature, and possible benefits, risks, and uses of the data. Such safeguards may include controlled access, pseudonymization, and anonymization of data, and quantitative techniques such as differential privacy, k-anonymity, ℓ-diversity, and t-closeness.

- Data processing agreements (also known as data transfer, data use, and data sharing agreements) between persons and/or organizations are an important privacy safeguard.

- Consideration should be given to adopting mechanisms that address compelled disclosure requests by state authorities of identifiable data and that prevent unauthorized access by third parties.

Consent and other lawful bases

- Data should be used strictly in accordance with the data subject's (or their legal representative's) consent for processing, and/or the terms and conditions of authorization for lawful processing by competent bodies or institutions (e.g. terms and conditions set by research ethics committees, waivers of consent), and in compliance with international and national laws (including tribal, indigenous, and aboriginal laws), regulations, general ethical principles, and best practice standards that respect conditions on downstream uses of the data.

Re-identification

- Any attempt to re-identify individuals or to generate information (e.g. facial images or comparable representations) that could allow the identities of research participants to be readily ascertained, should be strictly prohibited (and subject to sanction) unless where expressly authorized by law.
- Reasonable steps should be taken to prevent the identity of data subjects being leaked or determined through indirect means such as metadata, URLs, and message headers.

Data quality

- In order to promote responsible and valuable sharing, data and any associated metadata should be, to the greatest extent reasonably possible, accurate; verifiable; unbiased; current; stored in systems that enhance security, interoperability, and replicability; and in compliance with commonly accepted standards for data and metadata annotation.
- Regular quality assessments of datasets should be conducted.

Identifiable data disclosure to the public

- Subject to any applicable laws and/or the terms and conditions of authorization for lawful processing by competent bodies or institutions (e.g. research ethics committees), identifiable data should only be disclosed publicly in a publication or other format if: (1) data subjects have provided their explicit consent to public disclosure of their identifiable data and have been made aware of any reasonably foreseeable risks associated with the disclosure, and the disclosure is necessary for the purpose concerned; (2) data subjects have knowingly made their identifiable data public by their own explicit actions or permissions; or (3) disclosure serves a public interest, is necessary for the purpose concerned, and adequate safeguards are in place.

Data sustainability

- Where appropriate and in accordance with the data subject's (or their legal representative's) consent for processing, and/or the terms and conditions of authorization for lawful processing by competent bodies or institutions, and subject to appropriate safeguards, data should be retained for future processing through both archiving and using appropriate indexing and retrieval systems.

- A plan should be established for the possible discontinuance of a database or initiative, and in particular should establish, if possible, whether the data will be archived or transferred to another database for use in future initiatives. If such archiving or transfer to another database is foreseen, the plan should make clear that data will continue to be shared with data users subject to ongoing governance oversight through e.g. a research ethics committee and/or data access committee. The lawful basis for the archiving or transferring of data to another database for use in future initiative (e.g. data subject consent) should be verified.

Controlled access and registered access

- Requests by data users for access to data should demonstrate to those managing access requests (e.g. data stewards, research ethics committees, and/or data access committees), at a minimum: (1) legitimate interest in and intended use(s) of the data; (2) accessibility of the data only to authorized individuals; (3) a reasonable and specified time period of data access; and (4) destruction of the data after agreed use.

Data breach

- Mechanisms and procedures should be in place to maximize the likelihood of detection of data breaches, the handling of which may also be subject to local and national regulatory requirements.
- A data breach involving the disclosure of data by a data user should be reported to the data steward without undue delay. Consideration should be given to reporting a data breach to a competent supervisory authority and, if applicable, the data protection officer and the relevant research ethics committee for the primary study, in addition to the data steward if no such requirement otherwise exists.
- When a data breach involving the disclosure of data is likely to result in a high risk to the rights or interests of data subjects, such a breach should be reported without undue delay by the data steward or a key contact person (e.g. data protection officer, principal investigator, communication or cohort contact officer) to data subjects in the affected database by means of a public communication or similar measure whereby the data subjects have an opportunity to be informed in an equally effective manner.

Accountability

- All persons and organizations are accountable for promoting and protecting data privacy and security, including when data are shared with data users, repositories, and service providers.
- Data stewards should keep track of all whereabouts of the data and the persons and/or organizations with access to the data.
- Data stewards should clearly identify the individuals within their organization who are responsible for data privacy, data management, and reporting procedures (including a contact

person or contact point for complaints). Appropriate and regular training for the identified individuals to discharge these duties should be provided.

- Data stewards should track relevant new laws, regulations, policies, expectations, and best practices, sharing these with responsible individuals within their organization or entity, and with data users as appropriate.

- Where relevant, ongoing communication links should be maintained between data stewards, data users, and research ethics committees and/or data access committees.

Transparency

- Policies and practices with respect to the privacy and security management of data and access arrangements should be made publicly available. Plain language summaries of these policies and practices and access arrangements should also be made public.

- General information should be made openly available on an ongoing basis to data subjects as a group about how their data are being used and for what purposes.

- For data that are not anonymized, a procedure should be established to provide individual data subjects, if they so request, information about how their data are being used and for what purposes.

Complaints or inquiries

- Procedures should be established to receive and respond to complaints or inquiries about policies and practices relating to the privacy and security of data or data access requests. The procedures should be easily accessible and simple to use and should involve a commitment to deal with all complaints in a timely fashion.

Vulnerable populations

- Persons or organizations that seek to process data from vulnerable populations should consider conducting, as far is practicable, a vulnerable population-specific data privacy impact assessment regarding the use and sharing of such data.

- Persons or organizations that seek to process data from vulnerable populations should consider working with them to develop a data access protocol that governs requests by third parties for research requiring the processing of such data, unless there is an established vehicle in place.

Security

Security is concerned with organizational, technical, and physical measures and standards to effectively manage risks to the sensitivity and integrity of data and the availability of resources and services. Due regard should be paid to the GA4GH Security Technology Infrastructure, which complements this policy. The following guidance promotes safe and effective data sharing environments.

Organizational measures

- As human errors are among the most difficult errors to control, organizations should, with ongoing commitment of adequate resources: (1) develop, monitor, and enforce policies (consistent with this policy) to secure data; (2) appoint a security officer responsible for implementing and enforcing security policies and practices, and responsible for monitoring them through standards, procedures, and baselines; (3) implement internal and external security reviews and audits; and (4) implement and require ongoing training and education of personnel on privacy and security policies and best practices.

- The number of copies of data (as backup or otherwise) stored by persons or organizations should be kept to the minimum necessary to ensure adequate protection of the data in the event of primary copy data loss.

- Each organization should implement Identity and Access Management (IAM) policies, procedures, and technologies to verify the identity of each individual to whom access rights are to be granted, and to ensure that each individual is given access to all of (and only) the type and volume of data and services required for a specified period of time. IAM includes identity proofing, credential issuance, rights authorization, identity authentication, and rights revocation. As part of the IAM policies, organizations should maintain a list of persons having access to data and the list should be reviewed regularly and authenticated.

- Organizations that agree to recognize and accept authenticated identities and security attributes issued by other organizations ("federated identity") have the responsibility of assuring the trustworthiness of the issuers, as well as the currency and authenticity of asserted identities. The GA4GH Authentication and Authorization Infrastructure (AAI) standard may be used to federate identity authentication and service authorization.

- Consequences for data breaches should be clearly stipulated and enforced (see also the GA4GH Accountability Policy).

- In the context of cloud computing, companies providing cloud computing services to store, analyze, or warehouse data should have good management infrastructure and robust data encryption capabilities. The responsibility is on the data user/organization to ensure this infrastructure is compliant with local laws and regulations when uploading data to the cloud. Organizations should ensure that cloud service providers have independently audited against comprehensive and internationally recognized and respected information security standards, such as those promulgated by the International Organization for Standardization (ISO) and Statement on Standards for Attestation Engagements (SSAE). Organizations should also ensure that cloud service providers have up-to-date third party audit certifications and are maintained throughout the duration of the cloud service.

Technical measures

- Physical and logical access to computer systems and networks should be restricted to authorized individuals, and access granted only for those information assets and functions required to perform the user's assigned duties.

- Whenever possible, data should be pseudonymized or anonymized at the earliest possible opportunity.

- Where data are pseudonymized, an organization may assign a key to enable the data to be re-identified. The assigned key should not be derived from or related to the associated individual, should not be used for any other purpose, and should not disclose the mechanism used for re-identification. The direct identifiers associated with keys should be isolated on a separate dedicated server/network without external access. A defined procedure and auditable mechanism for reversing the pseudonymized data to (re)attribute to the data to a specific data subject should be in place.

- Emergency-management and disaster-recovery plans and safeguards should be implemented, including regular back-ups.

- Technical measures to secure data should comply with the relevant guidance and regulations (e.g. for clinical trials) and should aim to be interoperable with data sharing systems and software.

- Every system that accesses, stores, or transmits data should record an audit log of all security-relevant events. Audit trails should be reviewed regularly, and all suspicious events should be investigated. Where possible, automated, enterprise-wide, audit trail monitoring, with alerts for misuse and algorithms to amend or terminate access, should be implemented. Audit logs should be maintained for a minimum of one year, or as otherwise required by applicable law, and carefully protected.

- Configuration management of all hardware and software (including operating systems) should be implemented. Every change should be reviewed for potential privacy and security impacts.

- Organizations should take recommended actions to protect data and services from known and emerging threats, which would include monitoring sources of security threat information and installing security-critical upgrades as soon as they become available and have passed quality assurance testing within the organization.

- Organizations should protect data from new security vulnerabilities in any software used over the lifespan of a project involving the data. Such consideration should include ensuring that security patches to the software are promptly applied and that any vulnerabilities for which security patches cannot be applied in a timely way will be subject to scrutiny regarding alternative security safeguards.

- Organizations should routinely test their security systems, and periodically (e.g. yearly) engage an independent third party to perform security assessment and penetration testing.

Physical measures

- Computers, network equipment, media, and facilities used to collect, access, store, process, transport, or transmit data must be continuously protected using appropriate physical, technical, and procedural safeguards that limit access to authorized individuals.

- Physical security measures should be in place to protect data from natural hazards such as floods, fires, or earthquakes.

- Hardware used for sharing data should be tamper-resistant.

III. Implementation Mechanisms and Amendments

1. All persons and organizations supporting this policy should take all reasonable and appropriate measures, whether of a regulatory, contractual, administrative, or other character, to give effect to this policy and promote its implementation, monitoring, and enforcement. Procedures and policies should be transparent and accessible. Attention should be paid to the interrelation of this policy with other GA4GH policies (e.g. Consent Policy, Ethics Review Recognition Policy, Accountability Policy).

2. The GA4GH Data Security Work Stream will ensure that the technical standards and practices recommended in the GA4GH Security Technology Infrastructure are consistent with, and help enforce, this policy.

3. Any entity or individual supporting this policy may propose one or more amendments to the present policy by communicating the amendments to the GA4GH's Regulatory and Ethics Work Stream (REWS). The REWS shall publicly circulate such amendments for comments and possible inclusion in this policy.

4. The REWS, in collaboration with organizational members and other GA4GH Foundational and Technical Work Streams, will track the adoption of this policy and its application. The REWS will also routinely review the policy's provisions, be aware of advances in basic research and technology, and ethical and legal developments, and attempt to ensure that this policy is fit for purpose.

Acknowledgements

This policy was developed by the Regulatory and Ethics Work Stream of the GA4GH, and is the result of the collaborative work, comments, and input of many individual and organizational contributors.