Version: 26 May 2015

Global Alliance for Genomics and Health: Privacy and Security Policy

Preamble

The Global Alliance for Genomics and Health ("GA4GH") is an international, non-profit coalition of individuals and organizations working in healthcare, research, disease advocacy, life science, and information technology dedicated to improving human health by maximizing the potential of genomic medicine through effective and responsible data sharing. Its mission is "to accelerate progress in human health by helping to establish a common framework of harmonized approaches to enable effective and responsible sharing of genomic and clinical data, and by catalyzing data sharing projects that drive and demonstrate the value of data sharing."

The *Framework for Responsible Sharing of Genomic and Health-Related Data* (the "*Framework*") is a document developed by the GA4GH that sets forth a harmonized and human rights approach to responsible data sharing across the GA4GH in accordance with Foundational Principles and Core Elements. Elaborating on the general principles and guidance offered in the *Framework*, the GA4GH has created Policies that provide specific guidance for responsible sharing of genomic and clinical data. These Policies can help both individuals and organizations that support the mission of the GA4GH make improvements or adopt specific Best Practices[1] for responsible data sharing and governance processes.

This document is the GA4GH's Privacy and Security Policy. The objective of this Policy is to guide the sharing of genomic and health-related data ("Data") in a way that protects and promotes the Confidentiality, integrity, and availability of data and services, and the privacy of individuals, families, and communities whose Data are shared. This Policy builds on the *Framework*'s Core Element: "Privacy, Data Protection and Confidentiality", which states entities and individuals should:

> • "Comply with applicable privacy and data protection regulations at every stage of data sharing, and be in a position to provide assurances to citizens that confidentiality and privacy are appropriately protected when data are collected, stored, processed, and exchanged. Privacy and data protection safeguards should be proportionate to the nature and use of the data, whether identifiable, coded or anonymized."
>
> • "Forego any attempt to re-identify anonymized data unless where expressly authorized by law."

[1] See Appendix 1 for definition of this term. **All capitalized terms in this Policy are defined in Appendix 1.**

I. Context

1. Context. The significant public health imperatives for the translation of the new genomics knowledge, coupled with rapid technological developments, scientific advancements in genomics, and the increasingly global scale of data sharing, have brought new challenges. Particularly when linked to other often publicly available databases of personal data, use and sharing of Data can raise privacy and security concerns. This is because of the Data's inherent connection to identity and kinship, their long-term value and heterogeneous uses, and because of the potential re-identification of individuals from aggregate data. Indeed, privacy and security concerns about Data relate not just to the individuals who have contributed their Data, but also extend to families and communities. As the field is developing, there is uncertainty as to how privacy and security can be maintained. Safeguards are essential to sustain public trust and confidence in science and medicine, and to allow societies and human health to flourish across the globe. Developments in biomedical research require a robust, proportionate, and flexible Privacy and Security Policy, backed by meaningful monitoring and enforcement mechanisms, to ensure appropriate privacy protection of individuals, families and communities, the security of Data, and the trustworthiness of the GA4GH. The level of privacy and security applied to Data should be made compatible across organizations, entities, and countries to the greatest extent possible, while assuring compliance with applicable jurisdictional law. To enable such compatibility to be achieved requires that organizations meet some minimal level of privacy and security protection to be applied to the Data during collection, storage, use, and transmittal, and that Data Stewards communicate in a standard way the rules and constraints under which the Data may be shared and used.

2. Purpose. Building off the *Framework*, the purpose of this Policy is to provide principled and practical guidance for sharing Data in a way that protects and promotes the Confidentiality, integrity, and availability of data and services, and the Privacy of individuals, families, and communities whose Data are shared.

   To achieve this purpose, there is a need to:

   i. Ensure Data Donors are informed about the manner in which their Data are being used and for what purposes, as far as is practicable;

   ii. Effectively manage expectations about privacy protection and security measures, emphasizing both the benefits of responsible data sharing and recognizing that no data sharing use can ever be risk-free, and also positively committing to use Data in a way that is consistent with individual's expectations, and respecting the rights of individuals in relation to their Data;

   iii. Effectively manage the risk of individuals' Data being accessed or used in ways contrary to what has been consented to or otherwise authorized;

iv. Effectively manage the risk of accidental or malicious access, corruption or destruction of Data;

v. Effectively manage the risk of disruption, degradation, non-use, hoarding, and interruption of data and application services supporting availability and access to Data; and

vi. Effectively manage the risk that unethical, illegal or inappropriate actions result in damage or harm, and could cause individuals, families or communities to prohibit or unduly limit use of their Data.

3. Principles. This Policy is based on the Foundational Principles set forth in the *Framework*, namely:

- Respect Individuals, Families and Communities
- Advance Research and Scientific Knowledge
- Promote Health, Wellbeing and the Fair Distribution of Benefits
- Foster Trust, Integrity and Reciprocity

4. Interpretation.

4.1 Without ascribing legal meaning, this Policy should be interpreted in good faith and is to be understood as a whole. The Best Practices in Section II are to be understood as complementary and interrelated. Their function is to help guide policymakers in decision-making, as appropriate and relevant in different contexts, countries, and cultures.

4.2 This Policy distinguishes Privacy from Security. Privacy is treated as a fundamental value that protects all aspects of the lives of individuals, families, and communities, and that establishes reasonable limits to the use of Data. Security is treated as a measure that establishes safeguards to effectively manage risks to the sensitivity and integrity of Data and the availability of resources and services. While the Privacy and Security aspects of this Policy might be addressed to different actors with diverse responsibilities, the Policy should be read as a whole and with the overarching objective of ensuring that these crucial elements work together to deliver responsible data sharing.

4.3 For the purposes of this Policy, “data sharing” includes, but is not limited to, the use, viewing, transfer, linkage, or exchange of Data between the Data Donor and another party, or between a Data Steward and a third party, either openly or under specified access conditions. Data sharing may occur without having the data move from one place to another.

5. Definitions. Capitalized terms used in this Policy are defined in Appendix 1 (“Definitions and Glossary”). Words imparting the singular number shall include the plural and vice versa.

6. Application. It is expected that this Policy will be useful to all entities or individuals providing, storing, accessing, managing or otherwise using Data, and in particular the Individual and Organizational Members of the GA4GH ("Members"). These entities or individuals include, but are not limited to, researchers, research participants and patient communities, journal editors and publishers, research funding agencies, data protection authorities, hospitals, clinicians, research ethics committees, industry, ministries of health, and public health organizations.

II. Privacy and Security Best Practices

The following Best Practices of this Policy guide and facilitate the sharing of Data in a way that promotes and protects privacy and security in a proportionate manner. They also facilitate compliance with the obligations and norms set by international and national laws, policies and interoperable standards (see Appendix III for examples). Interoperability allows organizations and individuals to adopt these Best Practices (or defensible versions thereof) in a coherent manner that maximizes the opportunity for effective and robust sharing, and that promotes trust therein, both within the GA4GH and externally. These Best Practices should be interpreted in a manner that acknowledges different levels of risk and community cultural practices and, where appropriate, different contexts for data sharing and use.

Privacy

Privacy is a fundamental value 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. However, privacy is not an absolute right. Privacy protection involves the delicate balance of considerations at individual, familial, and societal levels. The following Best Practices assist in determining such balances relative to the protection of the core interest at stake and the Foundational Principles at the core of the *Framework*.

Primary Duty of Data Privacy Protection

- All Data should be safeguarded in accordance with applicable laws, norms, and guidelines, and should not be misused or wrongfully disclosed.

Consent

- Data should be used strictly in accordance with the Data Donor's (or his/her legal representative's) consent for collection, use and sharing, and/or the terms and conditions of authorization for use by competent bodies or institutions, and in compliance with national and international laws, general ethical principles, and best practice standards that respect restrictions on downstream uses.
- Data should be made available in Identifiable form only for specific purposes according to the level of permission of the research participant-Data Donor as well as Data Steward. If

data are Coded or Anonymized, it should take place at the earliest opportunity consistent with use for the authorized purposes. Moreover, Data Stewards should provide a clear summary or description of the coding or anonymization process that was applied, so that prospective Data Users can judge for themselves whether or not they can use the Data by the ethical standards and legal rules of their own jurisdiction, and judge for themselves whether the use of the Data is a responsibility that they can accept or are prepared to accept. Such description should also make clear that if Data are Anonymized, further robust data linkage would not be possible.

Ensuring Proportionate Safeguards

- Data privacy safeguards should be proportionate to the sensitivity, nature, and possible benefits, risks, and uses of the Data.
- Assessments of privacy risks should involve not only disclosure issues, but also reasonably likely harms, which may include individual or group discrimination or stigmatization. The reputational risks for entities or individuals of allowing particular uses of Data should also be considered.
- Where required, if Data Donors have consented to broad use or sharing of their Data, Data Stewards should conduct Data Privacy Impact Assessments to assess privacy risks before further use or sharing of the Data.
- An important component of fair treatment of Data Donors is adherence to Fair Information Practice Principles throughout the collection, use, storage, and exchange cycle of Data.
- Data Stewards should maintain an inventory that addresses the storage arrangements for Data, as well as the flows of such Data with appropriately defined sensitivity classes of the Data.

Re-identification

- Any attempt to re-identify individuals should be strictly prohibited, except where expressly authorized by the Data Donor or authorized under the law. This obligation follows the Data through the data sharing chain. Data Stewards should monitor data usage on a regular basis to detect any such re-identification attempts.
- Organizations should take reasonable steps to prevent the identity of individuals being leaked or determined through covert means such as metadata, URLs, and message headers.

Data Quality

- In order to promote valuable sharing, Data Stewards should ensure that the Data and any associated metadata held are accurate, verifiable, unbiased and current, and stored in systems that enhance security, interoperability and replicability.
- Data Stewards should conduct regular quality assessments of data sets.

- To ensure that quality controls are kept up-to-date, and to improve interoperability, Data Stewards should develop and maintain simple feedback mechanisms that inform them on the quality of Data and their annotations, and how the quality of the Data might be improved.

Data Disclosure and Publication

- Identifiable Data may only be disclosed publicly in a publication or other format if the Data Steward ensures that either: 1) Data Donors have provided explicit consent to public disclosure of their Identifiable Data, or 2) Data Donors have made their Identifiable Data public by their own actions or permissions. Moreover, no action should unjustifiably interfere with the interests and rights of Data Donors.
- Commitments made to Data Donors, e.g., that their Identifiable Data will not be publicly disclosed, should be respected by Data Stewards even after Data Donors have died, unless the legal representative of the deceased Data Donor provides express proof of the Data Donor's wishes to the contrary.

Collection and Sustainability

- Primary Data collection and aggregation should comply with all legal and ethical requirements of the jurisdiction in which the Data were collected. Identifiable Data should be held and used for only so long as is foreseen necessary for the purposes of their use, unless exemptions apply in applicable laws.
- The collection, use, and sharing of Data should be limited to what is relevant and necessary to accomplish the research purposes of a Data Initiative.
- Where appropriate, Data Stewards should ensure that Data are sustained for future use and sharing, through both archiving and using appropriate identification and retrieval systems, and through critical appraisal of the mechanisms and systems used for sharing Data, whether Identifiable, Coded, or Anonymized.
- Data Stewards, in consultation with relevant entities or individuals, should establish a plan for the possible winding down of a database or Data Initiative, and in particular establish, if possible, that the Data will be archived or transferred to another database for use in future Data Initiatives. Such a policy should make clear that Data will continue to be shared with Data Users without undue restrictions and will remain in conformity with the terms of any original consent or approval and subject to ongoing governance oversight.

Access

- Requests by Data Users to Data Stewards for access to Data should demonstrate, at a minimum: (1) legitimate intended uses that are in the public interest (i.e., securing an objective commonly valued by society) and with regard to established human rights; (2) assurances that Data are being accessed only by authorized individuals, e.g., accredited persons accessing Data that will be held and used only in safe environments; (3) a legitimate

and specified time period of access; and (4) secure disposal or return to the Data Steward of the Data after use and outside of any required retention period.

Data Breach

- A Data Breach by a Data User involving the potential disclosure of Identifiable Data should be disclosed without undue delay to the relevant Supervisory Authority and to the Data Steward. The Data Steward should then disclose the Data Breach to the Data Donors in the affected Data Initiative.
- A Data Breach by a Data Steward involving the potential disclosure of Identifiable Data should be disclosed without undue delay to the relevant Supervisory Authority and to the Data Donors in the affected Data Initiative.
- Mechanisms and procedures should be in place to maximize the likelihood of detection of Data Breaches, and to evaluate the re-identification risks in case such breaches occur. These mechanisms and procedures should be kept under regular review.

Accountability

- Data Stewards should clearly identify the individuals within their organization or entity 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 new regulations, policies, expectations, and best practices, sharing these with responsible individuals within their organization or entity.
- Data Stewards and Data Users must comply with applicable privacy regulations and ethical norms at every stage of data sharing, and be in a position to provide assurances that privacy interests are appropriately protected when Data are collected, stored, processed, and shared.

Transparency

- General information should be made openly available on an ongoing basis to Data Donors as a group about how the Data in a Data Initiative are being used and for what purposes, as far as is practicable.
- Data Stewards should provide individual Data Donors, if they so request, information about how their individual Data are being used and for what purposes, as far as is practicable.
- Data Stewards and Data Users should be open about their policies and practices with respect to the privacy and security management of Data and access arrangements. These policies and practices should be made openly available in a variety of formats (e.g. digital and hard copy) and should be generally understandable.

Complaints or Inquiries

- Data Stewards should put procedures in place to receive and respond to complaints or inquiries about their policies and practices relating to the privacy and security of Data. 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

- Entities or individuals that seek to use and share Data from Vulnerable Populations should consider conducting a Vulnerable Population-specific Data Privacy Impact Assessment regarding the usage and sharing of such Data.
- Data Stewards should consider working with Vulnerable Populations whose Data are proposed to be used and shared to develop a confidentiality agreement to prevent unauthorized disclosure of Data, as well as a data access protocol that governs all requests by third parties for research requiring the use of such Data.
- Research findings that identify Vulnerable Persons or Populations should not be published or disclosed without the consent of the relevant persons or communities or their representatives.

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. The following Best Practices promote safe and effective data sharing environments.

- Each organization should implement a security risk-management program that objectively assesses risks and implements appropriate safeguards to protect the sensitivity and integrity of the Data it holds and accesses, and the availability of its resources and services.
- Organizational, technological and physical measures appropriate to the data use and sharing and its objectives should be implemented in such a way as to protect the interests of the individuals, families and communities whose Data are being contributed, and the interests of the organization and the GA4GH.
- Each organization should implement and maintain security policy, practices, and technical safeguards consistent with applicable laws and current best practices.

Organizational Measures

- There should be ongoing commitment to security and continued emphasis of its importance by all involved in the use and sharing of Data.

- As human errors are among the most difficult errors to control, Data Stewards and their organizations should, with ongoing commitment of adequate resources: (1) develop, monitor and enforce a policy (consistent with this Policy) to secure Data; (2) appoint a security officer responsible for the implementing and enforcing the security policy and practices; (3) implement internal and external security reviews and audits; and (4) implement and require ongoing training and education of personnel on privacy and security policy and best practices.
- Each organization should implement Identity and Access Management (IAM) policy, procedures, and technology 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 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 policy, 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.
- Sharing of Data, whether Identifiable, Coded, or Anonymized, should be limited to legitimate scientific purposes and on a realistic need-to-know basis.
- Consequences for data breaches and breach of Confidentiality should be clearly stipulated and enforced.

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.
- Data should be Coded or Anonymized at the earliest possible opportunity.
- Where Data are Coded, an organization may assign a key to enable Coded Data to be re-identified. The assigned key may 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.
- 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 record 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 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 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 entities or individuals 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, Accountability Policy).

2. The GA4GH Security Working Group will ensure that the technical standards and practices recommended in the Security 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 Working Group (REWG). The REWG shall publicly circulate such amendments for comments and possible inclusion in this Policy.

4. The REWG, in collaboration with Members and other GA4GH Working Groups, will track the adoption of this Policy and its application. The REWG 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.

IV. Acknowledgements

This Policy is the result of the work of many people and committees. Developed under the auspices of the GA4GH's Regulatory and Ethics Working Group, the Policy was formulated by an international committee (Privacy and Security Policy Task Team) representing a wide spectrum of the law, security, bioethics, genomics, life science industry, and clinical communities. Collaborative input was provided from individuals as well as biomedical, patient advocacy, and ethical, policy and legal organizations, committees, and projects from all regions of the world. These include: the Centre of Genomics and Policy (McGill University); Intel Corporation; the National Institutes of Health (NIH, United States); PHG Foundation (United Kingdom); Roche Molecular Systems, Inc.; and the Public Population Project in Genomics and Society (P3G, McGill University).

Appendix 1

Definitions and Glossary

The following definitions shall be considered only for the purpose of this Policy, but are intended to align with the *Framework*.

Anonymized Data	Data that were related to an identifiable individual when collected, but through a process of removing all direct identifiers, thereafter prevents the identity of an individual from being readily determined by a reasonably foreseeable method. Using state-of-the-art techniques, properly anonymized data helps prevent both direct and indirect identification of an individual.
Best Practices	Policies or practices currently in use that successfully meet the goals of facilitating the sharing of Data; protecting the privacy and security of Data and Data Donors; and maintaining the public trust by using data appropriately and demonstrating the social value of the resulting research.
Coded Data	Data that are assigned one or more random codes. Direct identifiers are removed from the dataset(s) and held separately. The Key(s) linking the code(s) back to direct identifiers are available only to a limited number of persons, e.g. a research team, or are held by a third party (such as the data holder or a trusted third party) and are unavailable to the researchers. Also known as pseudonymized data and key-coded data.
Confidentiality	The ethical and legal obligation of an individual or organization to safeguard sensitive data by controlling access to a limited number of persons authorized by law or by the Data Donor. Maintaining Confidentiality in accord with relevant ethical and legal norms respects Data Donors' Privacy.
Data	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.

Data Breach	The wrongful release of Data, whether as a result of accident, negligence or malice.[2]
Data Donor	The individual whose data have been collected, held, used and shared.
Data Initiative	A purposive activity in which data collected for one research purpose are used for a new research purpose, often involving linking with other data sources.[3]
Data Steward	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 Privacy Impact Assessment	A formal process which assists organizations in identifying and minimizing the privacy risks of new projects or policies that make use of Data. The assessment involves working with people within the organization, with partner organizations, and with the people affected to identify and reduce privacy risks.
Data User	Individuals or organizations who are authorized by Data Stewards to access and use Data for the performance of their work. Data Users are secondary users of Data that are distinct from the primary Data generating research team, led by the Data Steward.
De-identified Data	A defensible, repeatable and auditable process that consistently provides assurance, based on proven and repeatable statistical methodologies, including removing identifiers and other indices, such that there is a very small risk of re-identification of any data that are made accessible to researchers.
Fair Information Practice Principles	Guidelines that represent widely accepted concepts concerning fair information practice to ensure that collection and use of data provides adequate data privacy protection. Originally from the United States Federal Trade

[2] Adapted from Nuffield Council on Bioethics, *The collection, linking and use of data in biomedical research and health care: ethical issues* (2015), available at <http://nuffieldbioethics.org/wp-content/uploads/Biological_and_health_data_web.pdf>.

[3] Adapted from Nuffield Council on Bioethics, *The collection, linking and use of data in biomedical research and health care: ethical issues* (2015).

	Commission's Fair Information Practice Principles, they are now reflected in other international documents, including the Council of Europe's Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data, the Organisation for Economic Cooperation and Development's (OECD) Guidelines on the Protection of Privacy and Transborder Flows of Personal Data, and the European Union's Data Protection Directive.
Framework	*Framework for Responsible Sharing of Genomic and Health-Related Data*
GA4GH	Global Alliance for Genomics and Health
Identifiable Data	Data that may reasonably be expected to identify an individual, alone or in combination with other data. One person's genotype data (i.e. the combination of sequence-related data) are, in theory, distinguishable (individuating) from all other people's genotypes, but they may not necessarily be Identifiable.
Identity and Access Management (IAM)	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	A piece of data that an encryption algorithm uses to determine exactly how to unscramble the Data.
Members	Individuals or organizations that are a member of the GA4GH.
Policy	Privacy and Security Policy of the Global Alliance for Genomics and Health
REWG	Regulatory and Ethics Working Group of the Global Alliance for Genomics and Health
Re-identification	The process of linking De-identified Data to an individual.
Privacy	In its informational dimension, a state of affairs whereby information relating to a person is in a

	state of non-access, or controlled access such that the person is able to decide whether and how personal information may be used and shared, and to know how that information is actually used and shared.
Security	The protection of the confidentiality and integrity of Data, and the availability of associated resources and services.
Security Risk Assessment	An objective analysis of the effectiveness of the current security controls that protect an organization's Data and a determination of the probability of losses to those Data.
Supervisory Authority	The public authority (or authorities) in each jurisdiction responsible for monitoring the application of the administrative measures and regulations adopted within their jurisdiction pursuant to data privacy, data protection and data security.
Vulnerable Persons/Populations	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.[4] Such persons or populations may include children, the elderly, pregnant women, prisoners, and those with mental health issues.

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

Appendix 2

Table of Concordance of Data Privacy and Security Terms[5]

Spectrum of identifiability

1 (Most identifiable)	2	3	4 (Least identifiable)
"Is or can be fully identifiable to everyone"	"Is unidentifiable to most, but remains re-identifiable to those with access to the key(s)"	"Is likely no longer identifiable to anyone"	"Never was identifiable"
• identified or identifiable • personal • nominative	• coded • key-coded pseudonymized • reversibly de-identified • linked anonymized • masked • encrypted	• anonymized • de-identified • irreversibly de-identified • non-identifiable • unidentifiable • unlinked anonymized	• anonymous[6]

Category 1: Identified/personal/nominative data are labelled with personal identifiers such as name or identification numbers. Data are directly traceable back to the Data Donor.

Category 2: Pseudonymization/coding/key-coding consists of replacing one attribute (typically a unique attribute) in a record by another. Pseudonymized/coded/key-coded data are labelled with at least one specific code and do not carry any personal identifiers, but an individual is still likely to be identified indirectly; accordingly, pseudonymization/coding/key-coding when used alone will not result in an anonymous data. Pseudonymization/coding/key-coding reduces the linkability of a dataset with the original identity of a Data Donor; as such, it is a useful security measure in genomic research, but it is *not* a method of anonymization.

Category 3: Anonymization is intended to prevent re-identification. Data must be processed in such a way that it can no longer be used to identify a Data Donor by using all the means likely reasonably to be used by person or entity. An important factor is that the processing *must* be irreversible to reasonable degree, i.e., anonymized data must not be traceable back to the Data Donor.

Category 4: Anonymous data are never labelled with personal identifiers when originally collected, nor is a coding key generated. Therefore, there is no potential to trace back Data to individual Data Donors. Anonymous data are of extremely limited utility in genomic research.

[5] Adapted from William W. Lowrance, *Learning from Experience Privacy and the Secondary Use of Data in Health Research* (London: Nuffield Trust, 2002) at 34; ICH, *Guidance for Industry: E15 Definitions for Genomic Biomarkers, Pharmacogenomics, Pharmacogenetics, Genomic Data and Sample Coding Categories* (April 2008); Article 29 Data Protection Working Party, Opinion 05/2014 on Anonymisation Techniques; Knoppers BM, Saginur M. The Babel of genetic data terminology. Nature Biotechnology 2005; 23(8): 925-927.

[6] Unlike Lowrance and the Article 29 Working Party, we are of the position that "anonymous" data should be placed in a category separate from "anonymized" data, as the former constitute data that were *never* individually identifying, whereas the latter constitute data that were, prior to an anonymization process, individually identifying.

Appendix 3

Examples of Procedural Guidance

Privacy

- Asia-Pacific Economic Community (APEC), *Privacy Framework* (2004) [link]
- Organisation for Economic Co-operation and Development (OECD), *Guidelines on the Protection of Privacy and Transborder Flows of Personal Data* (2013) [link]
- UK Information Commissioner's Office, *Anonymisation: Managing Data Protection Risk Code of Practice* (2010) [link]
- Australian Privacy Principles (APPs) (2014) [link]
- Government of Western Australia, *Practice Code for the Use of Personal Health Information Provided by the Department of Health* (2014) [link]
- Nuffield Council on Bioethics, *The Collection, Linking and Use of Data in Biomedical Research and Health Care: Ethical Issues* (2015) [link]
- Council of Canadian Academies, *Accessing Health and Health-Related Data in Canada* (2015) [link]
- UK Biobank, *Ethics and Governance Framework* (2007) [link]
- US Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule, 45 CFR Part 164, Subpart E (rev. 2013) [link]

Security

- ICH, *E15 Definitions for Genomic Biomarkers, Pharmacogenomics, Pharmacogenetics, Genomic Data and Sample Coding Categories* (2008) [link]
- Australian Signals Directorate, *Australian Government Information Security Manual (ISM)* [link]
- Government of Western Australia, *Information Security Policy* (2012) [link]
- ISO/IEC 27002:2013 Information technology — Security techniques — Code of practice for information security controls [link]
- ISO/IEC 27001:2013 Information Technology – Security Techniques – Information Security Management Systems [link]
- Standards Australia/Standards New Zealand: HB 231:2004 Information Security Risk Management Guidelines [link]
- HB 174:2003 Information Security management — Implementation Guide for the Health Sector [link]
- US Health Insurance Portability and Accountability Act (HIPAA) Security Rule, 45 CFR Part 164, Subparts A, C and D (rev. 2013) [link]

Policy Revision History

Policy Number/Version	Date Effective	Summary of Revisions
POL 001 / v. 1.0	June 2015	Original document