



## **Genomic Data Governance Glossary of Terms**

**Investigators: Prof Clara Gaff, Prof Wendy Lipworth, Miriam Wiersma, Kimberly Ross, Keeley Reade.**

**Advisor: Prof Ainsley Newson**

**Contact: [lineage.study@sydney.edu.au](mailto:lineage.study@sydney.edu.au)**

**Website: [Lineage Study Glossary - lineagestudy.org.au](http://lineagestudy.org.au)**

Terms frequently utilized within genomic data governance are often inconsistently defined and applied. The LINEAGE Study is a 5-year research project that aims to create a framework for ethical governance of genomic data in Australia. A project of this size and complexity needs to proceed with clear terminology, not least because common terms used by various stakeholders can have inconsistent definitions or ambiguity.

Objective 1.2 within the LINEAGE Study proposed to address the question of terminology prospectively by defining data governance terms used in the genomics context, to promote consistency in terminology throughout the lifespan of the project. The LINEAGE Genomic Data Governance Glossary will assist study Investigators in preparing their publications and research outputs. It will also be of interest to external stakeholders.

The LINEAGE Genomic Data Governance Glossary includes 21 items of data terminology relating to genomics, data sharing and governance. The Glossary was developed following a scan of academic and grey literature and leveraging Investigator expertise. Relevant literature was identified via structured searches, with an emphasis on federal (Australian) documentation to ensure consistency with current regulation and guidelines.

Over 140 peer reviewed and grey literature sources were reviewed, and 439 terms and definitions were extracted. The majority of definitions are already in common use within various contexts. The literature scan and analysis suggested that academic literature often lacked precise definitions, assuming prior knowledge of terms. All Glossary definitions were also reviewed by the LINEAGE Study consortium through surveys and meetings for additional feedback and refinement. Where a definition was modified it has been noted with the reference to the source. A full list of references can be provided upon request. The final list of 21 terms comprises a subset of the initial longlist of terms that were considered to: (1) Be relevant and specific to the LINEAGE Study project; and (2) Could be interpreted in an unambiguous manner.

Definitions will be reassessed annually, and new terms will be added as they become relevant.

## LINEAGE Genomic Data Governance Glossary

### Aggregate genomic data

Aggregate genomic data (also known as Genomic Summary Results) are the output of analysis of genomic data across the many individuals included within a specific dataset.

*Source: Adapted from- [What are Genomic Summary Results.pdf \(nih.gov\)](#)*

### Anonymised data

Data that were related to an identifiable individual when collected, but through a process of removing all direct identifiers prevents the identity of an individual from being readily determined by a reasonably foreseeable method. A key to re-identify the data does not exist.

*Source: Modified from - GA4GH Privacy and security policy, 2015.*

### Data Access

Data Access means the ability to make use of data through retrieval, transmission, viewing, editing, and/or analysis

*Source: [Data Access Definition | Law Insider](#)*

### Data Custodian

The Data Custodian is accountable for how data is managed and governed, including the safe collection, [aggregation](#), storage, use, movement, disclosure or destruction of data.

*Source: Modified from - Australian Genomics Policy on Data Access and Sharing for Secondary Use(Australian Genomics 2022)*

### Data Donor

An individual who has consented to their data being collected, held, used or shared for a research purpose.

*Note: Data collected for patient care is not considered a donation. Patients may be considered data donors if they consent to that data being used for research. When referring to circumstances in which there is a promise of benefit (e.g payment), it may be preferable to use an alternative term such as data contributor or data subject.*

*Source: Modified from -Australian Genomics Policy on Data Access and Sharing for Secondary Use (Australian Genomics 2022)*

## Data Interoperability

Data interoperability is the technical ability to move information efficiently, securely and predictably between organisations and systems using accepted data standards. These standards can vary across research and healthcare settings.

Source: Modified from - *Australian Government Australian Digital Health Agency, Interoperability and digital health standards 2023.*

## Data Sharing

Making information or data available to another agency, organisation or person under agreed conditions.

Source: Modified from - *NSW Infrastructure Data Management Framework (IDMF) and ACT Government Data Sharing Policy, 2022.*

## Data Sovereignty

In the context of Australia and genomic data, data sovereignty is the authority of a jurisdiction (federal or state) to determine how its data is collected, processed, and shared, as well as enforce its own laws and regulations related to data protection and privacy.

*Please note that an additional term will be provided for Indigenous Data Sovereignty and will be presented to our Indigenous Perspectives Investigators for advice.*

Source: Modified from – *Parliament of Australia, Parliamentary Business. Influence of International Digital Platforms, Chapter 5 – Data.*

## Data Visiting

Data visiting is a form of data access whereby data users view and analyse the data within the computing environment(s) of the data provider(s).

Source: Modified from - *Weise M, Kovacevic F, Popper N, Rauber A (2022)OSSDIP: Open Source Secure Data Infrastructure and Processes Supporting Data Visiting. Data Science Journal21(1):4. 10.5334/dsj-2022-004.*

## Data linkage

The process by which records representing the same entity or individual are connected across two or more sources of data.

Source: Modified from - *GA4GH Global Alliance for Genomics, Data Sharing Lexicon.*

## De-identified data

Data with all associated personal identifiers and other indices removed and replaced with a unique identifier that can be used for re-identification by a person with access to the key.

*Source: Modified from - Australian Genomics Policy on Data Access and Sharing for Secondary Use (Australian Genomics 2022)*

## Genomic data/information

Data or information relating to the inherited or acquired genetic characteristics of a person, their family and/or community.

*Note: there are other types of genomic data, but this project is concerned only with human genomic data*

*Source: Modified from - The GDPR and genomic data - the impact of the GDPR and DPA 2018 on genomic healthcare and research.*

## Health Information/ Data:

A subset of '[sensitive information](#)' and includes:

- Information collected in connection with the provision of a health service
- Information or opinion about the health or disability of an individual
- An individual's expressed wishes about the provision of health services
- Any information about health services provided to an individual.

*Source: Australian Commission on Safety and Quality in Healthcare, Data Governance Framework, 2023.*

## Metadata

Data about stored data. For example: its properties, history, versions etc.

*Source: Modified from - National Microbial Genomics Framework: 2019-2022 (Department of Health, Australia 2019).*

## Open data

Publicly available data that can be freely used, reused and redistributed by anyone.

*Source: Modified from - ACT Government Data Governance and Management Policy Framework August 2020.*

## Personal Information

Information or an opinion, whether true or not, and whether recorded in writing or spoken form, about an individual whose identity is apparent, or can reasonably be ascertained, from the information or opinion.

*Source: Office of the Information Commissioner, Protecting your right to information and privacy – section 12 of the Information Privacy act 2009.*

## Pseudonymization (of data)

RECOMMENDATION: Avoid use of this term. Refer to [anonymisation](#) of data or [de-identification](#) of data as appropriate to the process.

## Raw genomic data

Refers to genomic data that has not been cleaned, organised, reformatted or translated into information. Raw genomic data that has been altered or transformed into a format that is used for analysis and/or visualisation, is processed data.

*Source: Modified from - Australian Commission on Safety and Quality in Healthcare, Data Governance Framework, 2023.*

## Secondary Use

Using data or biological samples in a way that differs from the original purpose for which they were generated or collected.

*Source: Modified from - Global Alliance for Genomics, GA4GH, Data Sharing Lexicon, 2016.*

## Sensitive Information/data:

Sensitive information is a subset of [personal information](#). It means information or an opinion about an individual's racial or ethnic origin, political opinions, membership of a political association, religious beliefs or affiliations, philosophical beliefs, membership of a professional or trade association, membership of a trade union, sexual orientation or practices, or criminal record, or health, genetic or biometric information about an individual.

*Source: Modified from - Australian Government, Australian Law Reform Commission. For Your Information: Australian Privacy Law and Practice (ARLC Report 108)*

## Unstructured Data

Information that either does not have a pre-defined [data model](#) and/or is not organised in a predefined manner. Forms of unstructured data relevant to genomics can include: free text descriptions and summaries of health information, clinician notes, text files (e.g. word documents, PDF documents) patient/research subject videos images and audio files.

*Source: Modified from - ACT Government Data Governance Framework and Management Policy Framework, 2020.*