APPENDIX 2: Glossary

Accountable decision-maker	In the context of data under this Framework, this is usually the Data Sponsor (Executive Director level) or their delegate, or the Data Custodian. For complex data use and sharing proposals, it could be a Chief Executive or a Deputy Secretary.
Aggregated data	Aggregated data (as distinct from unit record data) is produced by grouping information into categories, typically with a combined count (i.e. numerical value) within each category.
API	Application Programming Interface
APPs	Australian Privacy Principles, found in the Privacy Act
Attribute disclosure	When new facts can be learned or inferred about an individual from a dataset.
Collection	A 'collection' of information occurs when the information comes into the possession or control of an organisation.
Confidentiality Undertaking	A Confidentiality Undertaking is a document containing a number of undertakings made by a data recipient pertaining to the handling of shared data. A Confidentiality Undertaking may be required to be executed by a Data Provider prior to data sharing (for example, if required by organisational data government frameworks and policies).
Data	Any facts, statistics, instructions, concepts or other information in a form that is capable of being communicated, analysed or processed (whether by an individual or by a computer or other automated means). For the purposes of this document, 'information' and 'data' are used interchangeably. Data may or may not include 'personal information' or 'health information, or other 'special category' information.
Data asset or dataset	A data asset or a dataset is a body of information or data, managed as a single unit, which is recognised as having value to the organisation and enables it to perform its business functions.
Data breach	If personal information has been lost, or accessed or disclosed without authority. A data breach will be 'notifiable' if the breach is

	likely to result in serious harm to one or more affected individuals.
Data Custodian	Makes decisions about the management of, access to and release of a data asset, including the definition of quality, and ensuring the asset is registered or catalogued.
Data fidelity	A measure of how accurate, complete, reliable and consistent data is in terms of representing the actual, real-world subject.
Data masking	The process of modifying, obscuring or replacing original data for security or confidentiality purposes.
Data Owner	The person or organisation responsible for the creation of the data, and who exercises authority over the data. The Data Owner may delegate or transfer certain aspects of its authority and its responsibilities to a Data Custodian, including via an agreement. For example, a general practice that collects patient data may (as the Data Owner) provide this data to another organisation, such as a state health department (as the Data Custodian) for specific purposes under an agreement.
Data Provider	The organisation which holds and controls the source health data that is the subject of a synthetic health data request. is disclosing data to one or more of the other organisations
Data Requestor	The organisation that is requesting the generation of synthetic health data from source data held by one of more of the other organisations
Data Sponsor	Undertakes data ownership on behalf of an organisation and ensures appropriate data governance policies are in place. The Data Sponsor may have the authority to approve data sharing.
Data Steward	Has day to day management of a data asset on behalf of the Data Sponsor, including ensuring that data quality and other standards are met. Provides support to Data Custodians and Data Sponsors.
Data utility	A measure of the value or 'usefulness' of data to achieve a goal or objective within a particular context.
De-identified data	'De-identified' data means that a person's identity is no longer apparent, or cannot be reasonably ascertained following the

	successful application of one or more de-identification techniques to 'personal information'[21]
Disclosure	The provision of personal information to another party outside an organisation
DSA	Data Sharing Agreement
DUA	Data Use Agreement
Dummy data	Sometimes described as a 'placeholder' or 'substitute' for real data. Dummy data will typically be fabricated to mimic the structure of real data for software or algorithmic testing purposes, but is non-meaningful and is not suitable for analysis.
Fake data	An umbrella term that means artificially generated data. Dummy data, mock data and synthetic health data can all be described as 'fake data'
Five Safes	A framework for considering how to control two types of privacy risks, when sharing data within a controlled setting
Health consumer	Individuals who use (or will use) health services, including their family and carers
Health information	Personal information that is about a person's: • physical or mental health • disability • current, past or future health services provided to them • wishes about future health services • actual or intended donation of body parts, organs or body substances • genetic information predictive of health • healthcare identifiers, and • all other information collected in the course of providing a health service.
HREC	Human Research Ethics Committee

[21] See the NSW Information & Privacy Commission Fact Sheet 'De-identification of personal information', 2020, https://www.ipc.nsw.gov.au/fact-sheet-de-identification-personal-information and the Office of the Australian Information Commissioner's guide 'De-identification and the Privacy Act', 2018, https://www.oaic.gov.au/privacy/privacy-guidance-for-organisations-and-government-agencies/handling-personal-information/de-identification-and-the-privacy-act

Identity disclosure	When data is re-identified and a person's identity can be assigned to a record. Identity disclosure can arise by one of two ways: by either matching a person to data (such as taking an individual, and finding data that matches them), or matching data to a person (such as starting with the data and finding the individual to whom that data relates).
Information	See 'data'
Insights	Information or details derived from data once it has been processed or analysed; the message conveyed by (or in reference to) data.
Membership disclosure	Membership disclosure occurs if it can be determined if an individual's data was in the source dataset that was used to generate a synthetic health dataset.
Mock data	Simulated or fictious data that is not derived or created from real data. It may be designed to replicate the structure and format of real data but does not contain or relate to real data records.
NHMRC	National Health and Medical Research Council
OAIC	Office of the Australian Information Commissioner
Output	The outcomes resulting from data use. For example, data analysis, results, insights, reports, or other information generated from the data.
Personal information	In the context of this Framework, personal information means any information about a person or that relates to a person who is at least reasonably identifiable. A person may be 'identifiable' if they can be 'distinguished' from all other members of a group. This may not necessarily involve identifying the person by name. Information does not have to be 'private' to be included in this definition. It can be true or false, an opinion or fact, recorded in a material form, or not recorded at all. Personal information can be about any living person, or a person who has died.[22] Personal information includes 'health information' and other special category data.
Perturbation	The process of modifying data for security or confidentiality purposes by making small changes intended to obscure original

[22] This meaning encompasses the elements of 'personal information' as defined in the range of privacy laws described in Appendix 3.

	values without impacting the overall statistical properties of the dataset (e.g. by adding 'noise' to the data).
PIA	Privacy Impact Assessment
Privacy Act	Privacy Act 1988 (Cth)
Real data	'Real world' data that relates to actual people, places, events, etc.
Redaction	The process of permanently removing or concealing data for security or confidentiality purposes.
Secondary purpose / secondary use	Using personal information for a purpose *other than* the primary purpose for which the information was originally collected
Sensitive information	Personal information relating to an individual's ethnic or racial origin, political opinions, religious or philosophical beliefs, trade union membership, sexual orientation activities, criminal record, health or genetic information, and some aspects of biometric information. Sensitive information is subject to additional legal privacy protections
Sharing	Data sharing involves data being provided from one organisation (the Data Provider) to another party at another organisation (known as the Data Requestor or End User)
Source data	The original data collected and held by the Data Provider from which a synthetic health dataset will be generated
Statistical disclosure risk	The risk that the identity of individuals, or new information about known individuals, within a dataset can be revealed. Includes both attribute disclosure and identity disclosure.
Statistical properties	Characteristics of a dataset that can be measured, analysed or interpreted
Synthetic health data	Data generated by a system or model that can mimic and resemble the structure and statistical properties of real health data, and uses real health data as input.[23]
Synthetic health data request	A synthetic health data request could include: One organisation requests another organisation to generate a synthetic health dataset for a specific project

[23] From the IAPP: https://iapp.org/resources/article/key-terms-for-ai-governance/

	An organisation wishes to establish a synthetic health dataset for multiple potential projects / purposes An organisation (or End User) requests to access or use a synthetic health dataset that was created for a different purpose or use case
Unit record data	Also called 'micro' data, this is data at the level of a single observation, for example data items relating to a unique individual, or a particular entity (such as a general practice)
Use	The use of personal information by a person inside an organisation
'Very low risk'	In the context of this Framework, 'very low risk' of re-identification means that even though it may be technically possible to identify an individual from information, doing so is so impractical that there is almost no likelihood of it occurring.[24]

[24] This is the standard of de-identification used by the OAIC for information to no longer be regarded as 'personal information' for the purposes of the Privacy Act. See: https://www.oaic.gov.au/privacy/privacy-guidance-for-organisations-and-government-agencies/handling-personal-information/what-is-personal-information