APPENDIX 8: Decision tree for complex synthetic health data scenarios

The five-step approach for generating and using synthetic health data outlined in this Framework should be sufficient for managing privacy risks where a synthetic health data request involves a fairly straightforward workflow.

A sample scenario of a straightforward workflow would be as follows:

1. A SynD organisation (e.g. the University of Sydney as the Data Requestor) requests another SynD organisation (e.g. NT Health as the Data Provider) to generate a synthetic health dataset using source data already held by that Data Provider, for an acceptable use case.

2. The source data is fit for purpose, and the resulting synthetic health dataset is considered robustly and effectively de-identified with only a very low risk of re-identification. The data utility is appropriate for the use case.

3. The Data Provider provides the Data Requestor with access to the synthetic health dataset in a secure manner and continues to have an appropriate level of oversight of the handling of the dataset.

However, for scenarios that involve more complex workflows, *or* where any of the steps in this Framework cannot be successfully completed, material privacy and legal compliance risks may emerge. In these cases, additional steps are needed to manage these risks.

The decision tree below anticipates some of these more complex scenarios where additional steps and assessments are required, to ensure the synthetic health data request can proceed lawfully.

Data requestor requests the creation of a synthetic dataset

Responsible data user requests synthetic data from the data provider for an approved use case

Responsible data user requests real data from the data provider, with the intention of using it to generate synthetic data themselves

Responsible data user requests synthetic data for a use case that is not approved under this framework

Data provider assesses the source dataset and generates synthetic dataset

Data provider holds and controls the necessary source data

Data from another organisation needs to be collected and linked to create necessary source data

Data provider generates synthetic data with a very low re-identification risk

Data provider generates synthetic data with *more than* a very low re-identification risk

Synthetic data to be used and shared for the approved use case in a secure manner

Synthetic data shared securely with data user with controls to protect against unauthorised access, use or disclosure

Synthetic data is to be shared publicly and / or with unknown users

PAUSE THE PROJECT

A PIA will be required to determine the lawful privacy pathway, and to identify and mitigate any other privacy risks.

See Appendix 9 for further guidance.