APPENDIX 6: Technical Assessment

The Data Custodian at the Data Provider typically liaises with the responsible Data Requestor about the suitability of the source data to meet the terms of the request. The Data Custodian at the Data Provider may also bring in specialist expertise, for example on issues around statistical modelling.

Data Requestors should provide context or explanation for why they are asking for certain data for a particular use case, so as to facilitate this assessment. Providing the Data Requestor the metadata or a data dictionary/terminology used by the Data Provider can assist in clarifying the scope of the data request, as well as progress discussions about the feasibility of the proposed methodology.

Technical Assessment Checklist

The accountable decision-maker must assess the request and the relevant source data in consideration of the following:

Considerations:	
Data source	Has the most appropriate data source been identified? Is there already an available data asset that could meet this need? Catalogues setting out datasets and repositories held by organisations can assist Data Requestors to understand what data is available.
Availability	Is the data actually available in the requested and appropriately structured format? For complex requests, this may require a detailed specification to be provided, listing all parameters and selection criteria for the data. Data dictionaries/ terminology and metadata can assist Data Requestors to understand what data is available.
Quality	Is the data of high quality? Indicators of data quality are accuracy, timeliness, consistency, validity, uniqueness and completeness.
Reliability	Is the data valid? Is it reasonable to rely on the data requested to provide valid results for the intended purpose?
Context	The assessment of whether data is fit for purpose should reflect the context of the synthetic health data request, including the overall project, expected outcomes and any decision-making that will rely on the data.
Comprehensible	Is the data comprehensible to enable proper understanding of the information? We need to avoid possible misinterpretation or misuse of information.

Granularity	Is the requested level of granularity appropriate? Could the Data Requestor's objectives be achieved by providing aggregate data instead, without having an undue impact on the utility of the data? In order to minimise the risks of re-identification, Data Requestors should be encouraged to practice data minimisation: • only request data that is required to directly answer the key question related to the use case • minimise the number of demographic data fields requested • aim for the broadest geographical level (e.g. LGA rather than postcode) • reduce the timeframe required to hold the requested synthetic health data
Utility	Is the data 'representative enough' for the use case at hand? Will it provide value in terms of the purpose underpinning the proposed use case / the question it is intended to help answer?
Capability	Does the Data Requestor have the appropriate skills to understand and interpret the data?
Methodology	Is the Data Requestor's proposed statistical methodology appropriate?
Clarity	Has a statement of data quality been prepared? What other notes need to be supplied with the synthetic health data to support End Users of the data to interpret it appropriately, including any caveats or statements about data quality? A statement of data quality should be prepared, to assist Data Requestors in determining if the data is fit for their specific use case. The statement of data quality should be made available with the synthetic health data, together with relevant information to assist with interpreting the synthetic health data. Where appropriate, there should be explanatory material to support End Users of the data to interpret it appropriately, including a data dictionary/terminology and any caveats or statements about data quality in relation to the data presented.

Further resources

Relevant organisational policies that may need to form part of an organisation's technical assessment can be set out / linked to here.

E.g. Data quality frameworks and standards, data quality statement templates, guidelines related to the technical creation of synthetic health data, etc.

- WA Department of Health, Synthetic health data: Governance and Technical Guidelines for the Generation and Use of Synthetic health data
- ABS Data Quality Statement Checklist