

APPENDIX 5: Impact Assessment

Impact Assessment Checklist

A synthetic data request could include:

- One organisation requests another organisation to generate a synthetic dataset for a specific project
- An organisation wishes to establish a synthetic dataset for multiple potential projects / purposes
- An organisation (or End User) requests to access or use a synthetic dataset that was created for a different purpose or use case (see the section on 'Re-using, re-purposing or re-synthesising synthetic datasets' in the framework above)

The accountable decision-maker must assess the synthetic data request in consideration of the following:



Considerations:	
Public interest	Will the synthetic data use case be in the public interest? This question should be considered in relation to: <ul style="list-style-type: none">• what the objective is when creating a synthetic dataset, as well as• each request for access and use of the synthetic dataset
Resourcing	What will be the resourcing impacts on the organisation, both in expenses and time? Will this project create a high opportunity cost - i.e. will resources be diverted from other projects more closely aligned to the organisation's policy or strategic objectives?
Beneficiaries	Will the synthetic data use case be for the benefit of many or a few? Can we maximise the utility of the data to be shared, to benefit multiple parties or derive useful insights to assist health care delivery and systems?
Community expectations and trust	Will the synthetic data use case meet community expectations about how their data will be handled? Will the data sharing take place in a manner that maintains or builds trust with stakeholders?
Privacy impacts	A Privacy Impact Assessment (PIA) will be needed for high-risk projects involving data

	sharing for linkage prior to creating a synthetic dataset, and where re-identification risk cannot be lowered to a very low level, in order to manage legal compliance risks.
Data ethics	Does the synthetic data use case raise concerns with regard to ethical considerations, in terms of the potential for harmful impacts on individuals, particular cohorts or sections of the community? See further discussion below.
Indigenous data sovereignty	Does the data include data about indigenous status? See further discussion below.


Data ethics

Examples of data types that could raise ethical concerns include:

- Health information (including physical and mental health)
- Disability
- Address or housing status
- Alcohol and drug use
- Sexuality
- Data about children
- Data about family violence / AVO's
- Aboriginality
- Ethnicity
- Language background other than English
- Religion
- Trade union membership
- Text-based fields / unstructured data
- Staff information
- Practice performance
- Data about services to be commissioned, where one of the organisations may be eligible to bid to provide the service during a procurement process

Examples of scenarios that could raise ethical and legal concerns include:

- There is data linkage across multiple data sources and/or platforms
- There is legislation (other than the privacy laws) protecting this specific type of data

- High risk technologies will be applied to the data. For example, generative AI technologies to conduct analysis and generate new information and insights, where the technology has not been assessed as safe 
- The data or results will be published or shared with an external party
- The risk of NOT proceeding with the use case, where meaningful benefits might otherwise flow to health consumers as a result of the use case

The ethics of synthetic data

Even where synthetic data has been robustly re-identified, organisations should be aware that there may still be ethical concerns that can arise from the use of this data and which should be considered. These include:

- Incorrect recognition of patients

Even if they are completely incorrect, if a person with access to a synthetic dataset believes that they recognise an individual in a synthetic dataset, and as a result they 'learn' new facts about that person such as a medical condition (whether that fact turns out to be correct or not), privacy harm can be done to a real person.

- Incorrect inferences

Some methods of controlling for re-identification risk include suppressing some data fields (e.g. country of birth), and/or removing small cell counts (e.g. people with unique combinations of variables). However, this may further entrench disadvantage, if some groups are then excluded from analysis.

Similarly, any inherent biases in the reference data will be carried through to the synthetic data.

If the synthetic dataset is not truly representative of the real population, there is a risk that a person with access to the synthetic dataset could draw inferences or statistical conclusions that are invalid. Whether those inferences are about individuals, populations, medical conditions, medical interventions or the quality of health service providers, invalid conclusions could lead to poor quality policy or operational decisions, or interventions in the health system.

Therefore if the synthetic data is to be used to make policy or operational decisions that could have consequences for individuals, it will be important to detect and correct bias in the generation of synthetic data, and ensure that the synthetic data is representative.⁴¹

In addition, the revelation that inferences drawn from the synthetic dataset were incorrect could in turn undermine confidence in the integrity of the *source* dataset, and the program of data collection that underpins it.

⁴¹ UK Information Commissioner's Office, "Chapter 5: Privacy-enhancing technologies (PETs) – Draft", September 2022, p.35; available from <https://ico.org.uk/about-the-ico/ico-and-stakeholder-consultations/ico-call-for-views-anonymisation-pseudonymisation-and-privacy-enhancing-technologies-guidance/>

- Correct inferences

Even if *correct*, inferences drawn from synthetic data could lead to policy or operational decisions that negatively impact vulnerable populations. Disregard of the risk of such outcomes would be contrary to the principles of ethical design and beneficence.

For example, while not creating a privacy risk for any individual patients, inferences drawn and publicised about the health outcomes for patients of a particular rural hospital or GP service could lead to a loss of confidence in those local health services. Local communities might face further disadvantage if the inferences were used to cut funding, and/or even poorer health outcomes might result if patients avoided using their local services at all.

Indigenous data sovereignty

Synthetic data use cases, which involve data about Aboriginal and Torres Strait Islander peoples, are subject to the same requirements of this framework as any other synthetic data use case. However some additional considerations will also apply.

SynD organisations are committed to Indigenous data governance, which is the right of Indigenous peoples to autonomously decide what, how and why Indigenous data and information are collected, accessed and used. It ensures that data on or about Indigenous peoples reflects their priorities, values, cultures, worldviews and diversity.

Proponents of Indigenous Data Sovereignty argue that in the past, research has too often been done ‘on’ rather than ‘for’ or ‘by’, Aboriginal and Torres Strait Islander peoples and communities. Thus, Indigenous Data Sovereignty is achieved when projects to support policy development or research:

- are led by Aboriginal or Torres Strait Islander experts
- have teams that include Aboriginal and Torres Strait Islander researchers or policy-makers, and
- collect and use data that reflects Aboriginal and Torres Strait Islander values and frameworks, such as cultural determinants of health.

Proposals to generate synthetic data which include data about Aboriginal and Torres Strait Islander peoples may need to be approved by an appropriate accountable decision-maker at an organisation.

Collaborating and co-designing with Aboriginal and Torres Strait Islander groups (such as Aboriginal Community Controlled Health Organisations (ACCHOs)) on synthetic data projects may provide opportunities to empower and support Indigenous Data Sovereignty, and to support research and data analytics designed to benefit Aboriginal and Torres Strait Islander communities. Synthetic data can provide a pathway for these benefits to be realised in a timely and cost effective manner, and can support research investment and advocacy efforts for Aboriginal and Torres Strait Islander groups.

Where a synthetic dataset is still considered ‘health data’, using or disclosing this data for research about Aboriginal and Torres Strait Islander peoples will also require the approval of a registered Aboriginal HREC, such as the Aboriginal Health & Medical Research Council (AH&MRC), or the [Research Ethics Committee of the Australian Institute of Aboriginal and Torres Strait Islander Studies](#) (AIATSIS).

Further resources

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

E.g. Privacy Impact Assessment guides or frameworks, AI assessment frameworks, data ethics guides, etc.

- Office of the Information Commissioner’s (OAIC’s) [Guide to undertaking Privacy Impact Assessments](#)
- The Lowitja Institute has produced an Information Sheet, ‘Indigenous Data Governance and Sovereignty’, for organisations involved in research.⁴²
- The National Health and Medical Research Council (NHMRC)’s *National Statement on Ethical Conduct in Human Research* outlines key matters for consideration in Chapter 4.7 – Aboriginal and Torres Strait Islander Peoples.⁴³
- NHMRC 2003, *Values and Ethics Guidelines for Ethical Conduct in Aboriginal and Torres Strait Islander Peoples about health research ethics*.⁴⁴
- NHMRC 2018, *Keeping Research on Track II: A companion document to Ethical conduct in research with Aboriginal and Torres Strait Islander Peoples and communities: Guidelines for researchers and stakeholders*⁴⁵
- Australian Institute of Aboriginal and Torres Strait Islander Studies 2012, *Guidelines for Ethical Research in Indigenous Studies*⁴⁶
- Cultural and Indigenous Research Centre Australia 2020, *Aboriginal Privacy Insights Report*, commissioned by the Office of the Victorian Information Commissioner⁴⁷

⁴² The Lowitja Institute, Information sheet: ‘Indigenous Data Governance and Sovereignty’, 2021;

https://www.lowitja.org.au/icms_docs/328550_data-governance-and-sovereignty.pdf

⁴³ <https://www.nhmrc.gov.au/about-us/publications/national-statement-ethical-conduct-human-research-2007-updated-2018>

⁴⁴ <https://www.nhmrc.gov.au/about-us/publications/values-and-ethics-guidelines-ethical-conduct-aboriginal-and-torres-strait-islander-health-research#block-views-block-file-attachments-content-block-1>

⁴⁵ 2018, <https://www.nhmrc.gov.au/about-us/resources/keeping-research-track-ii>

⁴⁶ 2012, <https://aiatsis.gov.au/sites/default/files/2020-09/gerais.pdf>

⁴⁷ <https://ovic.vic.gov.au/wp-content/uploads/2020/10/Cultural-and-Indigenous-Research-Centre-Australia-Presentation-on-Aboriginal-Privacy-Insights-Report.pdf>