HeSANDA Mental Health Node

Data Sharing Standard Operating Procedures

Mental Health Node

sheba
Secure Health Data and Biobank Platform

Barwon Health

DEAKIN UNIVERSITY

Version 1: 7/7/23

Authors:
Marko Milicevic, Deakin University
Professor Alison Yung, Deakin University
Dr Lourdes Llorente, Barwon Health
Nemanja Zivanov, Deakin University

Page of Contents

Introduction

Research with human participants generates a wealth of useful data. While the primary outcomes of a research project are published in journals, much of the data is not further used. Important research questions could be answered using previously collected Individual Patient Data (IPD) in secondary analyses, such as through meta-analyses or using other novel approaches.

The HeSANDA (Health Studies Australian National Data Asset) initiative is delivering the infrastructure of a national health data asset that supports the sharing and reuse of health research data across Australia. Researchers can search the HeSANDA catalogue, which contains meta-data from clinical trials available for secondary use upon request. The HeSANDA Mental Health Node (MHN) is one of nine 'nodes' supported nationally by the Australian Research Data Commons (ARDC), which is led by Deakin University. The MHN will work in close collaboration with the Secure Health Data and Biosample platform (SHeBa), a state-of-the-art platform created collaboratively between Deakin University and Barwon Health for discovering, combining, accessing, and analysing health data and biosamples collected by health services, research groups and other custodians. The MHN and SHeBa aim to accelerate research to improve the health and wellbeing of people in Australia and internationally by enabling clinicians and researchers to find, request, and work with unidentifiable health data and biosamples on a single, secure, remotely accessible platform. This will enable the secondary use of data to maximise the outputs and impact of already collected mental health data.

Visit the HeSANDA website to search the catalogue of clinical trials at https://researchdata.edu.au/health/.

Why share research data?

Benefits to you:

- Data sharing increases citations of the original authors, thus improving researcher reputations.
- Funding bodies look favourably at data sharing practices and may increase the likelihood that future research grant proposals will be successful.
- Meta-analyses are easier to produce.
- Combining data improves sample sizes and study power.

Benefits to your institution:

- A University's citation count can increase, thus improving the university's reputation to the public and research community.
- Data sharing promotes collaboration and resource sharing between institutions, which can result in future, collaborative research projects.
- Reuse of data is economically efficient, as it reduces the waste of grant funding.
- Secondary analysis projects are a cost-effective option for Higher Degree Research students.
- Replicability of research findings becomes easier and more transparent.

Benefits to the community:

- Secondary analyses allow the creation of new research questions and extension of original findings, resulting in improved knowledge and understanding of important public health matters, benefiting the community as a whole.
- The Australian Code for Responsible Conduct of Research encourages researchers to make their data available to other researchers.
- Sharing data encourages better transparency in research methods and results, which can improve the public image and perception of scientific research.

- Respects the generosity of human research participants, as it increases the utility of the data they provide and thus the value of their contribution.

Purpose

The Mental Health Node is hopeful that sharing health research data will become a standard practice for mental health research across Australia. This Standard Operating Procedures (SOP) document will guide researchers in preparing their projects for data sharing from the outset of their research. We strongly encourage researchers to follow this SOP as accurately as possible. However, each institution will have its own governance on data management and sharing. Researchers should familiarise themselves with their institution's governance and follow them where appropriate.

Audience

This SOP has been written for researchers who would like to share their data as part of the HeSANDA initiative and on SHeBa. This SOP will guide researchers through the setup phase of a research project, and what is required of them when their project has ended.

1. Data Custodianship

Researchers must assign a Data Custodian from the outset of a research project. The Data Custodian is an individual who is responsible for the management of a project's research data. The Data Custodian is typically the project Principal/Chief Investigator or their delegate.

The Data Custodian is responsible for:

- ensuring their study is registered on the ANZCTR (only applicable for Clinical Trials).
- ensuring data is stored in accordance with their institution and local government policies, Human Research Ethics Committee (HREC) requirements, and guidelines as illustrated in this SOP.
- ensuring their recorded data meets the requirements illustrated in this SOP that enable the safe and efficient sharing of unidentifiable human research data.
- ensuring HREC approvals allow for sharing of the unidentifiable human research data.
- responding to Secondary Use requests.
- ensuring any relevant conditions or restrictions stipulated in an ethics approval and/or Data Sharing Agreement are followed by any Secondary Users.
- creating Data Sharing Agreements for Secondary Users.
- consulting with Secondary Users throughout their project lifecycle where required and within reason.
- seeking approvals with relevant members of their research team prior to both making their data available for sharing and approving secondary use requests (if the Data Custodian is not assigned sole power of approving/rejecting requests).
- appointing a new Data Custodian where the original Data Custodian is no longer involved in the study or institution. If no original research team members remain, the Data Custodian must adhere to their institution's Custodianship Policy (typically, Custodianship will be handed to the Head of the Academic Unit or institution library, refer to your institution's policy on data custodianship).

2. Ethics applications and informed consent

Projects participating in data sharing must acquire the correct level of ethical approval from HRECs and fully inform participants regarding how their data will be used in future studies.

To access the comprehensive guide on how to appropriately fill an ethics application, state-specific module, site-specific module, and wording for your study protocol as it relates to data sharing, please visit the Mental Health Node website (website coming soon, contact Marko Milicevic at m.milicevic@deakin.edu.au for a copy) or MAGNET website at https://magnetctn.com.

For a comprehensive guide regarding concise and detailed participant information and consent forms (PICF) with data sharing considered, please visit the InFORMed project website at https://www.informedpicf.com.au/.

3. Data management

A critical component of data sharing is the accessibility of the data. Research data must be organised clearly and concisely and be accompanied by supporting documents. This includes 1) a data dictionary to describe what the variables of a dataset are and what they mean, 2) the project protocol which explains the study procedures, and

3) a summary of results from the original study. These documents are considered meta-data and will need to be included on the ANZCTR as a DOI (see point 7. DOI minting for more information), uploaded onto the ANZCTR directly, or be available to download via a persistent URL hosted by the Data Custodian's institution. The information provided for data sharing should enable a Secondary User, with no prior knowledge of the project or data, to understand what the data means, and be able to use the data correctly and efficiently for their secondary analyses.

3.1. Registration on the Australia New Zealand Clinical Trials Registry (ANZCTR)

The ANZCTR (www.anzctr.org.au) is a registry of clinical trials in Australia and New Zealand. Clinical trials participating in the HeSANDA initiative *must* be registered on the ANZCTR. This allows the MHN and ARDC to collect the meta-data about the trial that will help populate the trial listing on the HeSANDA federated catalogue and SHeBa. The Data Custodian/Principal Investigator, primary sponsor or delegate are responsible for registering their clinical trial on the ANZCTR before data collection. The ANZCTR listing for the clinical trial must be maintained and up to date.

The registry *must* satisfy the minimum meta-data fields as outlined below in **Table 1**.

Table 1: required meta-data fields within the ANZCTR

Header Name	**Metadata field(s)**
Titles & Identifiers	Public title, Scientific title, Trial acronym
Health Condition	Health condition(s) or problem(s) studied
Intervention/ exposure	Description of intervention(s)/exposure Comparator / control treatment
Outcomes	Primary outcome(s) and timepoint(s)
Eligibility	Key inclusion criteria Min/max age Gender Can healthy volunteers participate? Key exclusion criteria
Study Design	All fields
Recruitment	Final sample size
Funding & Sponsors	All fields
Ethics & Summary	Brief study summary
Contacts	Contact person for scientific queries
Data Sharing Statement	Will individual participant data (IPD) for this trial be available (including data dictionaries)? What data in particular will be shared? When will data be available (start and end dates)? Available to whom? Available for what types of analyses? How or where can data be obtained? What supporting documents are/will be available?
Summary Results	

3.2. ANZCTR Data Sharing Statement

An important section on the ANZCTR is the Data Sharing Statement. This section indicates to researchers how the data may be used and includes supporting documents, such as a data dictionary, the study protocol, the summary of results from published papers which used the original dataset, and a link to the institution's Data Sharing Policy who owns the data.

The below outlines how to prepare the supporting documents. For a step-by-step guide on how to fill the Data Sharing Statement field, go to Appendix D.

3.2.1. Data Dictionary

A data dictionary is the compilation of definitions, names, and information about the data. The data dictionary will assist secondary data users to understand the data and in turn, will help them to analyse it efficiently. The data dictionary should define all variables and items clearly and concisely. This can be achieved by defining all acronyms, providing descriptions of all instruments used in the project, and explaining all variables listed in the dataset. To accurately collate your data dictionary, the tools/instruments used throughout your study should be recorded in the dictionary from the outset and throughout.

The data dictionary must be added onto the ANZCTR as either 1) a persistent download link, 2) listed as a DOI, or 3) PDF/Word/Excel/CSV document of the dictionary uploaded directly onto the ANZCTR directly.

Note: If the study data is stored on REDCap, a data dictionary can be seamlessly extracted as a CSV file. All detailed fields should be filled during the REDCap project setup to ensure the data dictionary is rich with useful information that will help potential Secondary Users understand the dataset. If your study data is in a .sav SPSS file, a data dictionary can also be extracted. The MHN Data Manager can assist with extraction when your meta-data is ready to be included on the HeSANDA federated catalogue and/or onboarded to SHeBa where required.

3.2.2. Study Protocol

A study protocol is a document that describes the project procedures. The protocol is vital for Secondary Users, as it will help them understand how the project was conducted, what sort of data was collected, and subsequently guide them on how they can analyse the data.

The study team members involved in the development, review and revision of study protocols must ensure that protocols are consistent with international ethical and scientific standards, including the National Statement on Ethical Conduct in Human Research 2007 (updated 2018), SPIRIT 2013, and in compliance with the guidelines for Good Clinical Practice (GCP). When producing a study protocol, researchers must ensure that the rights and well-being of participants take precedence over the expected benefits of knowledge. The study team members involved in the development, review and revision of study protocols must also follow their Institution's SOP and must ensure that the study design is methodologically appropriate and adequately documented.

Resources, such as the Australian Clinical Trials Alliance (ACTA) Consumer Involvement and Engagement Toolkit (2022), provide practical advice on how to engage consumers in clinical research protocol development and delivery.

The study protocol must be added onto the ANZCTR as 1) a DOI to the original published protocol (if it has been published), 2) a persistent download link, or 3) PDF/Word document of the protocol uploaded directly onto the ANZCTR.

3.2.3. Summary of Results

The primary results of the project must be updated on the ANZCTR when the primary paper has been published. The summary of results must be added onto the ANZCTR as either 1) a DOI to the original published paper, 2) a persistent download link to the article, 3) PDF of the paper uploaded directly onto the ANZCTR, or 4) typed as free text.

3.2.4. Data Sharing Policy

The Data Sharing Policy is an institutional policy on the sharing of data from research projects. This differs from the Data Sharing Agreement, as the agreement are the guidelines for a specific project, whilst the Data Sharing Policy refers to an institution's guidelines for data sharing. Secondary Users must adhere to both the agreement and policy.

This document is often public-facing and can be found on your institution's website or can be obtained by contacting an institution's library/legal team. The Data Sharing Policy must be added onto the ANZCTR as either 1) a DOI listed, 2) a persistent URL (where the policy can be read on your institution's website), or 3) PDF of the data sharing policy uploaded directly onto the ANZCTR.

3.3. Data Storage

Research data should be stored indefinitely in a secure digital repository. This repository should be password encrypted and only be accessible by the project Research Team. The MHN strongly recommends researchers use **REDCap**, a completely free, secure, and user-friendly digital data recording and repository website. To check if your institution is a REDCap partner, click here. If your institution is not a REDCap partner or you use a different repository, please inform the MHN to determine if your repository can connect with SHeBa. To become a REDCap Partner, click here to find out more.

Individual Patient Data (IPD) meta-data can be seamlessly extracted from REDCap, including the data dictionary. This will save time for the required HeSANDA federated catalogue meta-data (the data dictionary) and at the SHeBa onboarding process when your data is ready to be shared. REDCap can also connect with the SHeBa platform, allowing Secondary Users to access your data securely (see point 6. The SHeBa platform to understand SHeBa's capabilities). Previously recorded data stored in CSV format can also be imported onto REDCap, meaning a data dictionary can be created retrospectively. The SHeBa Data Manager can assist with importing your data onto REDCap.

3.4. Data collection and recording

Data must be collected in accordance with the NHMRC National Statement on Ethical Conduct in Human Research in relation to the collection, use and management of data and information. Clinical trials must conform to Good Clinical Practice (GCP) to ensure that data collected from research with human participants is high quality, rigorously obtained and, most importantly, that the conduct of the project continues to protect the rights, safety, and wellbeing of human research participants.

- all personnel involved in the conduct of clinical trials must follow their Institution's SOP for obtaining and organising data generated in a research project, for the purpose of monitoring, reviewing, analysing, and reporting project progress, acceptability, and outcomes.
- site Investigators or their delegated staff should complete the Case Report Form (CRF) as specified by the protocol when relevant data become available. The data that is entered into the CRF should, as a best practice rule, be recorded first in the source data, unless there is a clearly documented rationale for why the data is to be entered directly into the CRF. Source data is defined as the place where the information is first recorded before being transcribed to the CRF.
- the Data Custodian keeps track of data entry or transfer progress and pending/missing CRFs, and informs the Project Manager of progress and any issues as per the Data Management Plan
- data that cannot be linked to a participant's personal or identifiable information should be stored indefinitely. Please ensure to check your state and institutional governance regarding the indefinite storage of research data.
- include provisions for deidentification. Any identifiable data should be hidden or removed from Secondary Users (SHeBa will allow for hiding or deidentification of specific variables from the dataset during the

onboarding process. For requests through the HeSANDA federated catalogue, the Data Custodian must remove identifiable information from the dataset before sharing it with a Secondary User).

- be labelled in detail to allow Secondary Users who have no prior knowledge of the study to identify and understand the data (see section on Meta-data for instructions and minimum requirements).
- include a clear and concise data dictionary that matches the recorded dataset.
- include a study protocol that outlines the procedure of the project in detail.

3.5. Preparing pre-collected datasets for reuse

For pre-collected datasets, you must ensure it is clean and labelled appropriately, and follows the requirements as listed above to allow for ease of secondary use (see point 4. Meta-data for instructions on minimum requirements). The project must have sought the appropriate level of consent from research participants before sharing unidentifiable data with Secondary Users (see point 2. Ethical considerations regarding data sharing for more information). If your project did not seek the appropriate level of consent, your project meta-data can still be listed on the HeSANDA federated catalogue. By listing only study meta-data, secondary researchers can communicate with Data Custodians, establish collaborations, and potentially investigate if re-consent of participants or a waiver of consent is possible to share the project IPD.

3.6. Data Sharing Agreements

Data Custodians must ensure a Data Sharing Agreement is in place before Secondary Users access the data. The responsibility and governance of data sharing between institutions and researchers is the sole responsibility of the Data Custodian, research team, and their institution/s. The MHN, HeSANDA, ARDC, SHeBa, Barwon Health, and Deakin University do not take responsibility for the governance of data used in data sharing.

A template agreement can be found as Appendix C in this SOP. This template outlines various provisions of data sharing and reuse third-party researchers can follow. The template will serve as a guide only. You must ensure you follow institutional and local data sharing policies and seek approval from your institution before utilising templated policies/guidelines from the MHN. Custodians may use the template provided and make edits where appropriate to suit their usage requirements, limitations, and institutional/local governance on data sharing.

The Data Sharing Agreement should:

- ask for a list of names of the secondary research team requesting/viewing the data.
- any limitations of the use of the data, including how long the data will be available for secondary use, institutional or project-based limits, privacy, and intellectual property rights as dictated by the governance and policies of the institution.
- request a summary of the intended use (topic summary, statistical analysis plan, publication plan)
- authorship guidelines (see below)

3.6.1. Authorship

Not only does data sharing benefit the health of Australia's population through extending scientific discoveries and creating new knowledge, but it also benefits researchers by enabling them to think about their data in new ways, develop collaborations with other researchers and institutions, and potentially increasing the number of publications, citations and other research outputs credited to them and their institution. The Data Sharing Agreement should outline authorship requirements. Authorship should be a collaborative discussion between parties. The MHN recommends that Data Custodians discuss authorship with their original research team first to outline:

- which of their research team members should be credited on secondary papers.

– the order of authorship from the original research team members.

The MHN recommends Data Custodians then discuss authorship with the secondary data users to outline:

– the overall authorship order of the secondary paper.

– how the original study should be credited in the published article (consider institution credit, original research team, funding bodies, and sponsors)

3.7. Indigenous Data Sovereignty

Data Sovereignty refers to the rights of Aboriginal and Torres Strait Islander peoples to own and protect the use of data that relates to them. Any research that produces conclusions about Aboriginal and Torres Strait Islander people, communities, or organisations must recognise Data Sovereignty. Consult with your institution and local community to understand more about Data Sovereignty.

Note: consent for data sharing at the individual level, i.e., an Aboriginal or Torres Strait Islander person provided data for a research study that is not directly related to Aboriginal or Torres Strait Islander people, is not bound by Data Sovereignty.

4. Meta-data

The research studies participating in the HeSANDA initiative that are available for data sharing will be listed on both the HeSANDA federated catalogue and on SHeBa. The SHeBa platform is jointly owned by Barwon Health and Deakin University (see point 6. The SHeBa platform for more information). Researchers can not only search and request data for secondary analyses through the HeSANDA federated catalogue, but they can also access a secure virtual environment to visualise and analyse the data. Both the HeSANDA federated catalogue and SHeBa need to collect important information about the study to make it findable. This information is called meta-data.

Your project's meta-data will be harvested from the ANZCTR for the purposes of the HeSANDA federated catalogue, which include the minimum required ANZCTR fields (see Table 1), and associated documents (see point 3. Data Management). The ANZCTR fields will then be minted into a DOI (see point 7. DOI minting for further information).

Your project's meta-data will also be 'on-boarded' to SHeBa with the assistance of the SHeBa Data Manager. Onboarding refers to the harvesting of your project's meta-data which includes information about the variables in your dataset. If the IPD is stored on a digital repository (like REDCap), SHeBa can request access to the repository to seamlessly harvest only the meta-data to create a listing about the trial. Meta-data harvesting for the HeSANDA federated catalogue will occur *before* the harvesting of meta-data for the SHeBa platform.

Important note: *The HeSANDA federated catalogue or SHeBa do not collect or store IPD. Instead, it harvests the information about the IPD (meta-data), this is information such as the variable names, their format and description. IPD will only be temporarily accessible for secondary analysis within SHeBa's secure online virtual environment upon approval by the Data Custodian of the data access request submitted by the secondary data user (see point 6. The SHeBa platform for more information on SHeBa's capabilities). Only the data users' members of the corresponding project workspace will have access to the data through this secure environment. The HeSANDA federated catalogue only stores meta-data accessible from the ANZCTR and associated documents (protocol, dictionary, summary of results) and does not store IPD.*

5. Secondary use requests

Secondary Users will be able to search for datasets on either the HeSANDA federated catalogue or SHeBa. When a Secondary User finds a dataset they would like to use for secondary analysis, they need to submit an

Expression of Interest (EOI) to access the data. This request will then be forwarded to the Data Custodian for review.

As the dataset will be listed both on SHeBa and the HeSANDA federated catalogue, the expression of interest pathways will differ. Below explains the request pathway through the HeSANDA federated catalogue and SHeBa platform, and what is required of the Data Custodian.

5.1. Requests through the HeSANDA federated catalogue

FIND: Secondary Users search the Federation Service data catalogue to identify datasets of relevance for their planned Research Project. The metadata made available in the catalogue by Data Custodians will facilitate this by adhering to the requirements specified in this SOP.

ENQUIRE: Secondary Users should review the metadata associated with records in the Data Catalogue to assess the feasibility of each dataset for use in their Planned Project. To improve this assessment, Secondary Users should contact the Data Custodian directly (i.e., outside of the Federation Service). If feasibility cannot be established at this stage, then the process may be discontinued.

REQUEST: If the Secondary User can determine the reasonable feasibility and availability of items in the Data Catalogue for their Planned Project, then a formal Data Request may be submitted to the Data Custodian via the Federation Service. The Federation Service will collect a standard set of information about the Planned Project from the Secondary User and forward it to the Data Custodian. The Data Custodian then reviews the request in line with their data sharing policy and procedures and records the outcome of the Request stage in the Federation Service. The outcomes of the Request stage may be either that the request is not approved, or that the Data Custodian provides in-principal approval for the request. If the Data Custodian requires that modification or the inclusion of additional information to the Data Request prior to completion of the Request stage, this may be requested via the optional Negotiate stage.

NEGOTIATE (optional): If the Data Provider requires modifications or further information about the Planned Project to be included in the Data Request, this may be requested via the Federation Service. The Secondary User may update their Data request accordingly and re-submit it for review. Alternatively, the Secondary User may discontinue the request at this point.

AGREE: If the Data Custodian gives in-principal approval for the Data Request, they will then confirm the terms of the Data Sharing Agreement with the Secondary User. If the parties are unable to reach an agreement and the process is discontinued, then this will be recorded in the Federation Service. Other than recording the outcome of this stage, this process and any associated records are beyond the purview of the Federation Service.

ACCESS: If the parties agree to the terms of a data sharing agreement and the Secondary User is provided access to the requested Data Catalogue items [Endpoint 6], then this will be recorded in the Federation Service. The provision of data access and handling of IPD are beyond the purview of the Federation Service and are to be carried out by the Secondary User and Data Custodian in accordance with the terms agreed to in the Access stage.

5.2. Requests through SHeBa

Once a request has been sent by a Secondary User, the Data Sharing Agreement is discussed (see point 3.5. Data Sharing Agreements for more information). When the Data Sharing Agreement has been ratified by all parties, the requesting researcher must then apply for ethical approval from their local Human Research Ethics Committee (HREC) to conduct their study (see point 2. Ethical considerations). Once ethically approved, the Data Custodian will accept the request, and the SHeBa Data Manager will open a project workspace in SHeBa for the Secondary User so analyses can commence, under the provisions of the Data Sharing Agreement, within SHeBa's

secure data analysis environment (see point 6. The SHeBa platform for more information on SHeBa's capabilities).

Once the data analysis (and therefore the project) is completed, the Principal Investigator of the project submits a request to export the summary data. This includes the results of their analyses and any other artefact created during the project, such as scripts. No IPD can be exported. Once the request is submitted, the SHeBa Data Manager will inspect the files before the Secondary User can access them. This allows an extra layer of security for Data Custodians, as the Data Manager will ensure that no IPD or any identifying information is inadvertently exported. Once this is verified, a download link will be provided to the Principal Investigator, and the project will be archived.

Datasets for analyses are initially provided for a limited period of 12 months. Should the Secondary Data Users need to extend their access to their project beyond 12 months, for example, to finalise the analysis, they will need to submit an extension request that will need to be approved by the corresponding Data Custodian. If no extension request is submitted/approved, the project will expire, and its members will lose access to the project workspace and its contents. A request to re-initiate an expired project can be submitted by the principal investigator in the 12 months after the project expires. This request will need to be approved by the Data Custodian, as it will allow access to the dataset.

In the case that the Secondary Data Users need to access their archived project after they have exported their summary data, for example for extra data analysis advised by journal reviewers, they will need to submit an access request that will need to be approved by the corresponding Data Custodian.

Note: it is the responsibility of the Data Custodian and Secondary User to adhere to local policy and governance associated with data sharing. This includes adhering to the Data Sharing Agreement, and proof of ethics approval before accepting a request.

6. The SHeBa platform

SHeBa is a federated data catalogue and distributed data network, enabling the secure aggregation and analysis of data and biosamples for secondary analyses. The platform is co-produced by Deakin University and Barwon Health to enable research across the continuum of clinical care and public health.

- SHeBa allows only unidentifiable data access to Secondary Users, with potentially identifying data securely tokenized prior to export from the Data Custodian's data repository. Tokenization means that the actual value of the variable, e.g., driver's license number, will be substituted by a random series of numbers and letters that will be associated to that unique driver's license number, allowing linkage with other datasets in SHeBa for which the driver's license has also been tokenized.
- SHeBa permits Data Custodians to completely allow/remove access to specific variables within their database during the initial on-boarding process, as well as by participant/patient level of consent at any stage. Note that when participant consent is revoked, their data will not be imported for subsequent data access requests, but they will remain in the projects that had access to the dataset prior to consent being revoked.
- Data is only viewable and analysable through SHeBa's secure virtual workstation within a project workspace accessible only by the users' members of the project workspace. IPD cannot be exported, copied, or pasted into external documents. Only a non-identifying summary of results from data analysis can be exported.
- SHeBa is not a data repository, and therefore it does not store IPD, but rather it stores meta-data (i.e., information) about the research project and about the corresponding IPD in its Data Dictionary (see point 4. Meta-data for more information). Upon approval of a Data Access request by the corresponding Data Custodian, the IPD is copied to the SHeBa virtual environment within the corresponding secure project

workspace. The IPD is then accessible for analysis only through secure virtual machines within the workspace, and only by the members of the workspace, for a limited period.

7. DOI minting

Digital Object Identifier (DOI) is a way to give a unique and permanent identification number to digital objects, such as datasets or research papers. It is a digital fingerprint that allows others to find and access information that remains permanently available on the internet. Minting DOIs is the act of applying a digital object its own individual code that will always identify it. In clinical research, DOIs are used to identify and share data from clinical trials. It is important for certain information from research projects to be DOI minted for the purposes of data sharing in the HeSANDA initiative. These DOI will be publicly accessible and will not include any IPD or identifying participant information.

The meta-data provided to the MHN will be minted with its own DOI. The minting process will be managed manually through Deakin. A MHN team member will collect the required meta-data from the ANZCTR and the Data Custodian, and it will then be submitted to Deakin Libraries – a partner within the DataCite consortium. The DOI is customizable, where it can point to any landing page (i.e., origin of dataset, the MHN website, SHeBa landing page, institutional landing page containing the dataset, etc.). This DOI will consist of the meta-data collected from the Data Custodian as listed on the ANZCTR and information integrated into DataCite. The DOI will be made available for both SHeBa and the HeSANDA federated catalogue.

A second DOI will be minted, specifically for SHeBa, that consists of the meta-data harvested from the dataset's repository. This includes information about the variables in the dataset and does not include any IPD. This DOI will be made available for SHeBa users.

8. Contact and support

For more information and support on establishing your project in line with data sharing principles, please contact the Mental Health Node.

General Enquiries
Email: mentalhealth_n@deakin.edu.au

Collaborations and Mental Health Node enquiries
Position: Project Lead
Name: Professor Alison Yung
Email: alison.yung@deakin.edu.au

Project setup, SOP, ethics, and research enquiries
Position: Mental Health Node Project Manager (Researcher Liaison)
Name: Marko Milicevic
Email: m.milicevic@deakin.edu.au

SHeBa enquiries
Position: SHeBa Data Manager
Name: Dr Lourdes Llorente Escrina
Email: lourdes.llorente@barwonhealth.org.au

Technical enquiries
Position: Mental Health Node Project Manager (Technology Liaison)
Name: Nemanja Zivanov
Email: n.zivanov@deakin.edu.au

Researcher resources
mentalhealthnode.org (coming soon)
sheba.com.au (coming soon)
researchdata.edu.au/health
magnetctn.com (includes useful resources for clinical trialists)
informedpicf.com.au
ctiq.com.au

Appendix A: List of terms and definitions

Term	Definition
SHeBa	**S**ecure **He**alth Data and **B**iob**a**nk Platform. SHeBa is a state-of-the-art platform for discovering, combining, accessing, and analysing data and biosamples collected by health services, research groups and other custodians in a secure remotely accessible platform. The SHeBa platform is jointly owned by Barwon Health and Deakin University.
Meta-data	Information about the variables collected in the trial, but not about their content (value), i.e., not individual patient data. For example, the name of trial instruments, headers, items, or 'PICO' data (patient/problem, intervention, comparison/control, outcomes), study protocol, data dictionary, and summary of results.
MHN	The **M**ental **H**ealth **N**ode: one of nine 'nodes' supported nationally by the Australian Research Data Commons and is facilitating the sharing of mental health related research data. The Mental Health Node is based at Deakin University, is led by Professor Alison Yung, and works closely with the SHeBa platform and Barwon Health in Geelong, Victoria.
DOI	**D**igital **O**bject **I**dentifier: unique alphanumeric string assigned by a registration agency (the International DOI Foundation) to identify digital content and provide a persistent link to its location on the internet, e.g., publishers assign a DOI when an article is published and made available electronically.
Data Custodian	The Data Custodian is an individual who is responsible for the management of a project's research data. The Data Custodian is typically the project Principal/Chief Investigator or their delegate.
Secondary Researcher	The Secondary Researcher is an individual or group of researchers that are requesting access to a Data Custodians research data for secondary analyses.
HeSANDA	**He**alth **S**tudies **A**ustralian **N**ational **D**ata **A**sset.
ARDC	**A**ustralian **R**esearch **D**ata **C**ommons.
IPD	**I**ndividual **P**atient **D**ata. This refers to the (unidentifiable) data that was collected from participants in a research study.

Appendix B: Data sharing preparation checklist

The below is a checklist you can use to ensure you are meeting the requirements for data sharing.

Item	Check
Registered the trial on the ANZCTR?	
Filled the minimum required meta-data fields on the ANZCTR?	
Appropriately filled the Data Sharing Statement on the ANZCTR, indicating IPD is sharable?	
Sought extended/unspecified consent in the PICF including opt-out option?	
Approval to share unidentifiable health data in the ethics application?	
Assigned a Data Custodian who can respond to data access requests?	
Prepared a copy of the study protocol that can be public facing on the ANZCTR?	
Prepared a template/draft Data Sharing Agreement for Secondary Users?	
Collated a data dictionary, or are able to export one from REDCap/SPSS (or other data repository)?	
Stored the data appropriately on REDCap (preferred) or another secure repository?	
Updated the ANZCTR with summary of results, study protocol, institutional data sharing policy, and data dictionary at the conclusion of the project?	
Responded to ANZCTR reminder emails to keep listing up to date?	

Appendix C: Data Sharing Agreement template

Data Custodians must ensure a Data Sharing Agreement is in place before Secondary Users access the data. The responsibility and governance of data sharing between institutions and researchers is the sole responsibility of the Data Custodian, research team, and their institution/s. The Mental Health Node, HeSANDA, ARDC, SHeBa, Barwon Health, or Deakin University do not take any responsibility for the governance of data used in data sharing.

This template Data Sharing Agreement outlines various provisions of data sharing third party researchers can follow. The template will serve as a guide only. You must ensure you follow institutional and local data sharing policies and seek approval from your institution before utilizing templated policies/guidelines from the Mental Health Node. Custodians may use the template provided and make edits where appropriate to suit their usage requirements, limitations, and institutional/local governance on data sharing.

Data Sharing Agreement template

This Data Sharing Agreement is for the sharing of The Data from The Project titled ***[insert project title name]*** as held by ***[institution name]***. The Project was first approved by ***[ethics committee, approval number]*** and is registered on the ***[clinical trials registry name, registry number]***.

Investigators from The Project:

Position: Data Custodian/CIA *(add new individuals where necessary)*
Name:
Institution:
Email address:
Description of project role:

Position: CIB
Name:
Institution:
Email address:
Description of project role:

Position: CIC
Name:
Institution:
Email address:
Description of project role:

Secondary Data User/s (researchers requesting and accessing The Data):

Position: Principal Investigator *(add new individuals where necessary)*
Name:
Institution:
Email address:
Description of project role:

Position:
Name:
Institution:
Email address:
Description of project role:

Position:
Name:
Institution:
Email address:
Description of project role:

To be completed by the Secondary User/s

Proposed project title.

Provide the title of the proposed project that uses The Data.

Data requested.

Describe in detail what data is being requested.

Intended use.

Describe in detail what you intend to do with The Data (e.g., meta-analyses, aggregate data with other studies).

Data analyses plan

Describe in detail what statistical methods you plan to conduct on The Data.

Dissemination plan.

Describe in detail how you intend to disseminate your findings (e.g., journals, forums, conferences).

Data governance.

[information regarding the owner institution Data Sharing Policy, link to policy]

Authorship of published results.

The Secondary User/s must include the following list of researchers as contributors on any published work using The Data.

[list of author names, ORCID ID]

The Secondary User/s must acknowledge The Data was provided to them by ***[institution providing the data]***, from the original published article ***[doi link, name]*** and the Data Custodian/PI ***[name]*** within the manuscript body related to descriptions of data, in any section within the journal that allows the author to list acknowledgements, and acknowledged at conferences or forums on presentation slides and or verbally. The Secondary User/s is/are solely responsible for any costs associated with publication or dissemination of results.

The Secondary User/s will have the right to analyse The Data as per the statistical analyses plan provided above and publish the results of any analyses in scientific journals. The Secondary User/s will not be provided any data with identifiable information about human participants. Participants will only be linked by a study ID number which does not include any identifiable information. Any potentially identifiable information will be removed from The Data before it is shared.

The Data Custodian or affiliated institution reserve the right to audit or review use of The Data to ensure compliance with The Data Sharing Agreement.

Limitations of use

The Data Sharing Agreement commences on ***[DD/MM/YYYY]*** and concludes on ***[DD/MM/YYYY]***. The Secondary Data User/s will have access to the Individual Patient Data from the proposed project between these dates only.

If an extension is required beyond the dates agreed, the Principal Investigator must contact the Data Custodian to amend The Data Sharing Agreement.

The Secondary User/s must retain a copy of secondary analyses results for ***XX*** years at the conclusion of The Data Sharing Agreement. The Secondary User/s is/are responsible for responding to journal review requests. The Data Custodian will assist with review requests where appropriate and within reason if the review request concerns the original dataset.

The Data may only be used for the current proposed project, and only for the intended purposes outlined in The Data Sharing Agreement. Any additional analyses and or publications that fall outside the scope of The Data Sharing Agreement which include the results of any analyses must first be discussed and reviewed by the Data Custodian.

The Data Custodian or any Investigators from The Project are not responsible for the accuracy of any secondary analyses nor any consequences resulting from the use of The Data. Secondary User/s must conduct any Secondary Analyses with integrity and follow NHMRC guidelines with respect to research with human participants.

The Data shall be supplied as a ***[file type]*** through the ***[Secure Health Data and Biobank platform]*** only. Access to SHeBa requires a subscription, and all costs are the responsibility of the Secondary User (see SHeBa website for more information).

Reporting requirements

The Secondary User must submit a project report ***XX days*** from the conclusion of the Data Sharing Agreement. The report should include a summary of results from the secondary analyses, any publications as a result of the analyses, and a summary of where the results have been presented.

The Data Custodian

As the Data Custodian providing The Data, I hereby:

- Agree to provide The Secondary User/s with The Data in line with the conditions stipulated in The Data Sharing Agreement.
- Confirm that Informed Consent and Ethical Approval permits the sharing of unidentifiable human research data to Secondary Users. Any participants who did not provide Informed Consent will not have their data shared.
- Agree to collaborate with The Principal Investigator and/or Secondary User/s on the proposed project, within reason.
- Confirm The Research Agreement follows the Data Sharing Policies as stipulated by ***[institution providing the data]***

Data Custodian Name (print): _____

Data Custodian Signature: _____

Date: _____

The Principal Investigator

As the Principal Investigator, I hereby:

- Agree to follow the conditions and restrictions of use of The Data for the secondary analyses stipulated in The Data Sharing Agreement.
- Agree to conduct the secondary analyses with integrity, respect The Data Custodian and Investigators from The Project, and protect the privacy and wellbeing of the research participants whose data is being shared.
- Agree to conduct the secondary analyses in accordance with The Australian Code for the Responsible Conduct of Research, 2018.

Principal Investigator Name (print): _____

Principal Investigator Signature: _____

Date: _____

Appendix D: ANZCTR Data Sharing Statement instructions

The ANZCTR holds important information about your trial. This information is called meta-data and will be used to populate your trials listing on the HeSANDA catalogue. An important section on the ANZCTR is the Data Sharing Statement. This section indicates to researchers how the data may be used and includes supporting documents which help researchers understand the trial and data. This document includes instructions on how to appropriately fill the Data Sharing Statement. You or a delegate will require a login to the ANZCTR and credentials to update the listing.

For any questions and support, please email:

Marko Milicevic
Mental Health Node Project Manager (Researcher Liaison)
m.milicevic@deakin.edu.au

Supporting Documents

There are a number of important documents that must accompany the meta-data listed in your trial's ANZCTR. Supporting documents must be made **publicly accessible** unless there is a legal or ethical requirement to restrict access. The supporting documents should already exist as part of your trial.

1. **The study protocol.**

A study protocol is a document which describes the project procedures. The protocol is vital for Secondary Users, as it will help them understand how the project was conducted and what sort of data was collected.

2. **The study data dictionary.**

The data dictionary is a document which explains the variables in your dataset. It will assist secondary users to understand how your data is labelled and what it means. The dictionary should list and define all the instruments used in your trial, define acronyms, and explain data formats. This can be a PDF file, excel spread sheet, or Word document. If a data dictionary was not collated manually from the outset of the study, a data dictionary can be exported from an SPSS file or REDCap. The Mental Health Node can help you export the data dictionary if required. If you are unable to generate a data dictionary for any reason, please notify Marko Milicevic m.milicevic@deakin.edu.au.

3. **The summary of results.**

The summary of results are the primary results found from the project. This will help Secondary Users understand what was found from the original analyses and inform them how the data could be used in different ways. If you are still collecting data or awaiting publication of primary results, the summary of results can be updated when they become available.

4. **A copy of your institution's Data Sharing Policy.**

This will be included with any trial included on the HeSANDA federated catalogue. Each institution should have their own policy on data sharing, which is publicly available on their website, or can be found on your institute's intranet. If you are unable to locate it, contact your library/legal team.

There are three options for making supporting documents available on the ANZCTR:

Option 1 [DOI] - we strongly recommend creating a DOI for each supporting document. Your institution may be able to create this for you. Alternatively, you may consider creating a DOI by uploading your supporting documents to public/generalist repositories such as Figshare, Dryad, Zenodo, or OSF. These generalist repositories can create a DOI for you. The DOI can then be used in the 'link' option in response to this ANZCTR question. You should speak to your institution first for advice on what options they allow and are best suited for your trial. The Library or Research Office can support with repository choice and DOI considerations.

Note: If a DOI already exists, for example the protocol or summary results have been published in a journal, a DOI link of the article is sufficient.

Option 2 [Web link] - If it is not possible to mint a DOI for your supporting documents, then provide a link to a web page where the document can be downloaded.

Option 3 [Attachment] - If neither of the previous options are available, upload your supporting documents directly onto the ANZCTR. For study protocols please use the prefix "Study Protocol" in the filename of your attachment. For data dictionaries please use the prefix "Data Dictionary" in the filename of your attachment.

Filling the Data Sharing Statement

Below is a step-by-step guide on how to fill the Data Sharing Statement, including the supporting documents. You must have login credentials to edit the trial. Once the changes have been made, submit them for review by the ANZCTR and email **Marko Milicevic** at m.milicevic@deakin.edu.au. The changes will be reflected on the HeSANDA catalogue automatically once approved.

ANZCTR question	ANZCTR description	Guidance for HeSANDA trials
Will individual participant data (IPD) for this trial be available (including data dictionaries)?	* Indicate whether there is a plan to make individual participant data (IPD) publicly available for this trial. IPD refers to raw line-by-line data collected from each participant. • Yes; IPD and related data dictionaries are/will be available • No; IPD will not be available	Nominated Trials must select: "Yes; IPD and related data dictionaries are/will be available" If you select "No", you are indicating that you are not willing/able to share your data under ***any*** circumstances - if this is the case, then your trial is ***not*** eligible to be part of HeSANDA. You can answer "Yes" as long as you meet one of the following criteria: 1. **You don't have ethics approval (or a waiver of consent) to share data, but you would be willing to seek ethics approval in some circumstances:** If you didn't originally obtain ethics approval to share data, it might not be possible to obtain approval at a later stage to share data for 'unspecified future use' (i.e. the broadest form of approval for data sharing). However, your ethics committee may be willing to provide approval for you to share your data on a case-by-case basis. If you are willing to review requests from others to access your data for a specific project and (potentially) seek ethics approval to share your data with that project, then you can select 'Yes'. You should clarify what kinds of projects and groups you would be willing to do this for in the relevant sections below. 2. **You have ethics approval to share data:** This does not need to be approval to share with anyone who requests the data. It just needs to be approval to share data under some circumstances with people outside of the groups running the trial (e.g., a secondary researcher). While it is best practice to obtain consent from trial participants to

		share data, ethics committees may provide a waiver of consent under some circumstances - this constitutes ethics approval to share data. 3. **There are no other restrictions preventing you from sharing your data.**
What data will be shared? * (Mandatory when 'Yes' is selected for IPD question in Step 11)	Please describe what data will be shared e.g., all of the individual participant data collected during the trial, after de-identification; individual participant data underlying published results only, etc.	Nominated Trials should indicate at least one of the following: • All participant data collected • Participant data supporting the publication results • Participant data relating to primary outcomes • Participant safety data *Sample response* *"Participant data supporting the publication results* *Participant data relating to primary outcomes"*
When will data be available (start and end dates)? * (Mandatory when 'Yes' is selected for IPD question in Step 11)	Please outline the timeframe of data availability, i.e., beginning and end dates for when the data is expected to be available, e.g. Immediately following publication, no end date; Beginning 3 months and ending 5 years following main results publication; no end date determined etc.	Nominated Trials should describe when the data will be made available. Options/examples include: (Optional) Pick one from the following: • Data are available straight after publication • Data are available ['x' months] after publication Pick one from the following: • Data are available for a finite time • Data are available for an indefinite time Add the following information (approximate dates are ok): • Start date: [insert date] • End date: [insert date] / [if "data are available for an indefinite time" then insert "Unknown"]

		If dates are subject to change (e.g., if funding ends/is extended) then your trial registration must be updated (as per the general requirements for updating ANZCTR records) *Sample response* *Example 1:* *"Data are available 5 months after publication* *Data are available for an indefinite time* *Start date: July 2023 (approx.)* *End date: Unknown"* *Example 2:* *"Data are available for a finite time* *Start date: 1 July 2023* *End date: 1 July 2028"*
Available to whom? * (Mandatory when 'Yes' is selected for IPD question in Step 11)	Please specify who can/will be able to access the data, e.g., anyone who wishes to access it, only researchers who provide a methodologically sound proposal, case-by-case basis at the discretion of Primary Sponsor, etc.	Nominated Trials should specify one or more of the following: Data are [potentially] available to (select all that apply): • Researchers from not-for-profit organisations • Commercial organisations • Other Based in (select one): • Australia only • Any location • Other Further information: • [insert further details about the above responses if required] • [insert further information usage permissions if required]

		• [insert a link to your institutions data sharing policy] *<u>Sample response</u>* *"Data are potentially available to:* *- Researchers from not-for-profit organisations* *- Commercial organisations* *- Other* *Based in:* *- Any location* *Further information:* *All data requests will be considered by the primary sponsor on a case-by-case basis. Requests must include a methodologically sound proposal. Specific conditions of use may apply and will be specified in a data sharing agreement (or similar) that the requester must agree to before access is granted. For further information, see our data sharing policy (http://www.university.edu/datasharing policy)."*
Available for what types of analyses? * (Mandatory when 'Yes' is selected for IPD question in Step 11)	Please clarify if there is a specific type of analysis for which the data are/will be available, e.g. any purpose, only to achieve the aims in the approved proposal, for IPD me	Nominated Trials should indicate at least one of the following: • Any type of analysis • IPD meta-analysis or systematic review • Assessed on a case-by-case basis • Any approved protocol • Audit/verification of results *<u>Sample response</u>* *"IPD meta-analysis or systematic review* *Assessed on a case-by-case basis"*
How or where can data be obtained? * (Mandatory when 'Yes'	Please specify how/where data are/will be shared e.g., unrestricted access via web address (provide link), access subject to approvals by Principal	Nominated Trials should include the following:

is selected for IPD question in Step 11)	Investigator (provide email or other contact details), etc.	"Access can be requested via the Health Data Australia catalogue (https://researchdata.edu.au/health). Search for the ACTRN number in the catalogue to find datasets associated with this trial." ARDC also recommends including a link to the data sharing policy for the trial (where available). For e.g. "For further information, see our data sharing policy: {insert URL}" Please note that you can use your own processes handling data requests in addition to the HeSANDA platform. If so, you should provide those details in addition to the suggested wording above. *Sample response* *"Access can be requested via the Health Data Australia catalogue (https://researchdata.edu.au/health). Search for the ACTRN number in the catalogue to find datasets associated with this trial.* *For further information, see our data sharing policy (http://www.university.edu/datasharing policy)."*
What supporting documents are/will be available? *	Select all types of supporting information that will be shared. Choose the appropriate type(s) from the list. • No other documents available† • **Study protocol (mandatory)** • Statistical analysis plan • Informed consent form • Clinical study report • Ethical approval • Analytic code • **Data Dictionary (mandatory)** † Note that if this option is selected, other options will not be available for selection.	**Nominated Trials are required to make their Study Protocol and Data Dictionary publicly accessible unless there is a legal or ethical requirement to restrict access.** For study protocols, Nominated Trials should tick the "Study Protocol" option. The ANZCTR accepts full protocols for this section or published protocols. If the protocol has been published in a journal, please list the DOI link to the article. For data dictionaries, Nominated Trials should tick the "Other (please specify)" option then specify that you will share your data dictionary.

	†† If 'Other' is selected, please note that it is mandatory to specify the other type of document that is/will be available.	
How or where can supporting documents be obtained? * (Mandatory for each of the documents selected in the previous field)	Indicate how the corresponding document can be obtained, e.g., citation, link, email, other, attachment. Note that it is mandatory to complete at least one of these. Attachment: It is the responsibility of the registrant to ensure that any uploaded documents comply with copyright regulations. Please note that any files attached will be publicly available via the trial ANZCTR registration record. Attached files cannot exceed the maximum size of 35MB per file. Maximum number of attachments allowed: 20	**Supporting documents must be made publicly accessible unless there is a legal or ethical requirement to restrict access**. Nominated Trials should avoid using the Citation, Email, or Other options. ARDC recommends the following: • **Option 1 [DOI]** - ARDC strongly recommends creating a DOI for each supporting document. Your institution may be able to create this for you. Alternatively, you may consider creating a DOI by uploading your supporting documents to public/generalist repositories such as Figshare, Dryad, Zenodo, or OSF. These generalist repositories can create a DOI for you. The DOI can then be used in the 'link' option in response to this ANZCTR question. You should speak to your institution first for advice on what options they allow and are best suited for your trial. The Library or Research Office can support with repository choice and DOI considerations. • **Option 2 [Web link]** - If it is not possible to mint a DOI for your supporting documents, then provide a link to a web page where the document can be downloaded. (NB. The URLs for your supporting documents should also be listed as 'Related Items' (DataCite element 20) in your dataset DOI) • **Option 3 [Attachment]** - If neither of the previous options are available, upload your supporting documents to ANZCTR. For study protocols please use the prefix "Study Protocol" in the filename of your attachment. For data dictionaries please use the prefix "Data Dictionary" in the filename of your attachment.

<u>Summary results:</u> **Have study results been published in a peer-reviewed journal? (Step 12)**	**Yes** No† † If your results are awaiting publication, select no and update this field when the results are available.	List the DOI of any published papers which used the dataset. <u>*Sample response*</u> **Journal publication details** **Publication date and citation/details** Yung AR, Milicevic M, Berk M. Fair funding for mental health research. Australian & New Zealand Journal of Psychiatry. 2023;0(0). https://doi.org/10.1177/00048674231177226

Note: ANZCTR Data Sharing Statement guidelines created by Dr Kristan Kang, ARDC.