

MIDL Information Pack

2. MIDL for users

Version 1.1

Melbourne Institute Data Lab (MIDL)
Melbourne Institute: Applied Economic &
Social Research, University of Melbourne



Melbourne
Institute
Data Lab
In collaboration with Cyconsol

Introduction

This guide provides an introduction to the Melbourne Institute Data Lab (MIDL, pronounced 'middle'), a secure data environment for housing, curating, and analysing sensitive data.

This guide is part 2 of 3 of the MIDL Information Pack.

Part 1 (identifiable by its blue cover) is intended for users, data custodians and the public who wish to understand MIDL. It provides information to explain the purpose and use of MIDL, its background and history, its service offering, and information about the governance of providing access to data housed in MIDL.

Part 2 (yellow cover) is for MIDL users, and provides information on gaining access to MIDL, the user journey and associated costs for using MIDL.

Part 3 (red cover) is for data custodians and provides information about hosting data assets in MIDL, as well as providing additional information on the services the Melbourne Institute: Applied Economic & Social Research (MI) offers for curating data and increasing data usability for approved users. Part 3 also provides a detailed description of MIDL's security measures and information on security controls. **Part 3 is a sensitive document and is not available for general audience access.**

Background

The Melbourne Institute Data Lab (MIDL, pronounced 'middle') is a secure, purpose-built data enclave that enables virtual access to micro-level data for curation, analysis, and visualisation. MIDL provides an environment that permits the development of collaborative data environments for use by authorised researchers and analysts, permitting rigorous and deep study of critical issues important to Australian society.

MIDL is a collaboration between the Melbourne Institute: Applied Economic & Social Research (MI), a research department within the Faculty of Business and Economics at the University of Melbourne with a mission to inform Australian economic and social policy; and Cyconsol, an Australian-based professional services provider specialising in the design and operation of cybersecurity and ICT capabilities working with Australian Government and industry.

For more information on MIDL, please see Part 1 of the MIDL Information Pack: "What is MIDL?".

MIDL for the user¹

A critical contribution of MIDL is the relative importance of working with administrative data through the lens of an empirical social scientist. Often "user ready" administrative data sets as provided by a custodian are not in a "research ready" state. MIDL staff and the users of the data will contribute to the development of these "research ready"² data sets by enhancing the materials provided by the custodians. These materials will include memoranda and analysis code that enable the near immediate use of the data for an approved project.

MIDL is one of several vehicles available for a user to house sensitive or proprietary data. MIDL permits the user to apply for:

- (1) A [standalone project](#) with a workspace that houses the data from the data custodian for analyses to answer their approved research question; and/or apply for,
- (2) a [Shared Data Environment \(SDE\)](#), a larger project designed and developed to house multiple data sets with the aim of promoting research in a particular theme.

¹ Please note that 'user' and 'researcher' are used interchangeably in this document. A detailed classification of a 'user' in MIDL can be found in Part 1 of the MIDL Information Pack.

² Data with appropriate preparation and transformation made ready for analysis. This includes additional documentation of variables, any treatments applied to clean the data and underlying rationale and example code to reproduce the transformation to raw data (as-collected data or as-provided data from the data custodian).

Benefits to the user

There are four benefits to users for accessing data through MIDL.

The primary benefit is access to multiple sensitive or proprietary data assets from primary or secondary data sources. MIDL is designed to achieve a very high standard for data security and a standard trusted within Government for managing their sensitive data. It is the hope of the MIDL team that these gold standards practices in data security and management enables trust in data custodians to provide access to approved users to their data assets through MIDL such that research may be undertaken to inform Australian economic and social policy.

A secondary benefit is the ability to access multiple data sets such that they can be combined³ to undertake richer data analyses for deeper insights into behaviours and/or trends that can better inform the approved research question by the users. All of the above mentioned activities are subject to data sharing agreements between MI and the various data custodians. These requirements will be provided to all approved users on MIDL through the project onboarding process.

The third benefit is the ability for users to be involved in a collaborative research environment that we call a 'Shared Data Environment' or SDE. An SDE provides approved users with access to multiple data sets specifically for the purpose of undertaking research under a specific theme. Some or all data sets permitted for use within a SDE are available for approved users of that SDE based on their project requirements or any other requirements under specific data sharing agreements. Alongside the "as provided" versions of data sets provided by the data custodians, users of the SDE also have access to a "research ready" version of these data as prepared by MI staff. Users in the SDE also have access to other material relating to these data as shared by MI staff or other users inside the SDE. This is made possible through the use of the MIDL's information management system (the MIDL Wiki) and contributions made to this wiki by users inside the SDE. These contributions are intended for users to share information relating to their use and analyses of their data including any derived variables which will better inform and speed up research undertaken towards the SDE's core research theme. These contributions may take the form of additional documentation, memoranda or example code for analyses of data stored in the SDE.

Finally, as MIDL was designed by researchers *for* researchers and other users for conducting research, MIDL uses a cost-recovery based costing model. It is the intent that MIDL provides a cost-effective solution for researchers requiring a secure platform for the analyses of sensitive data assets.

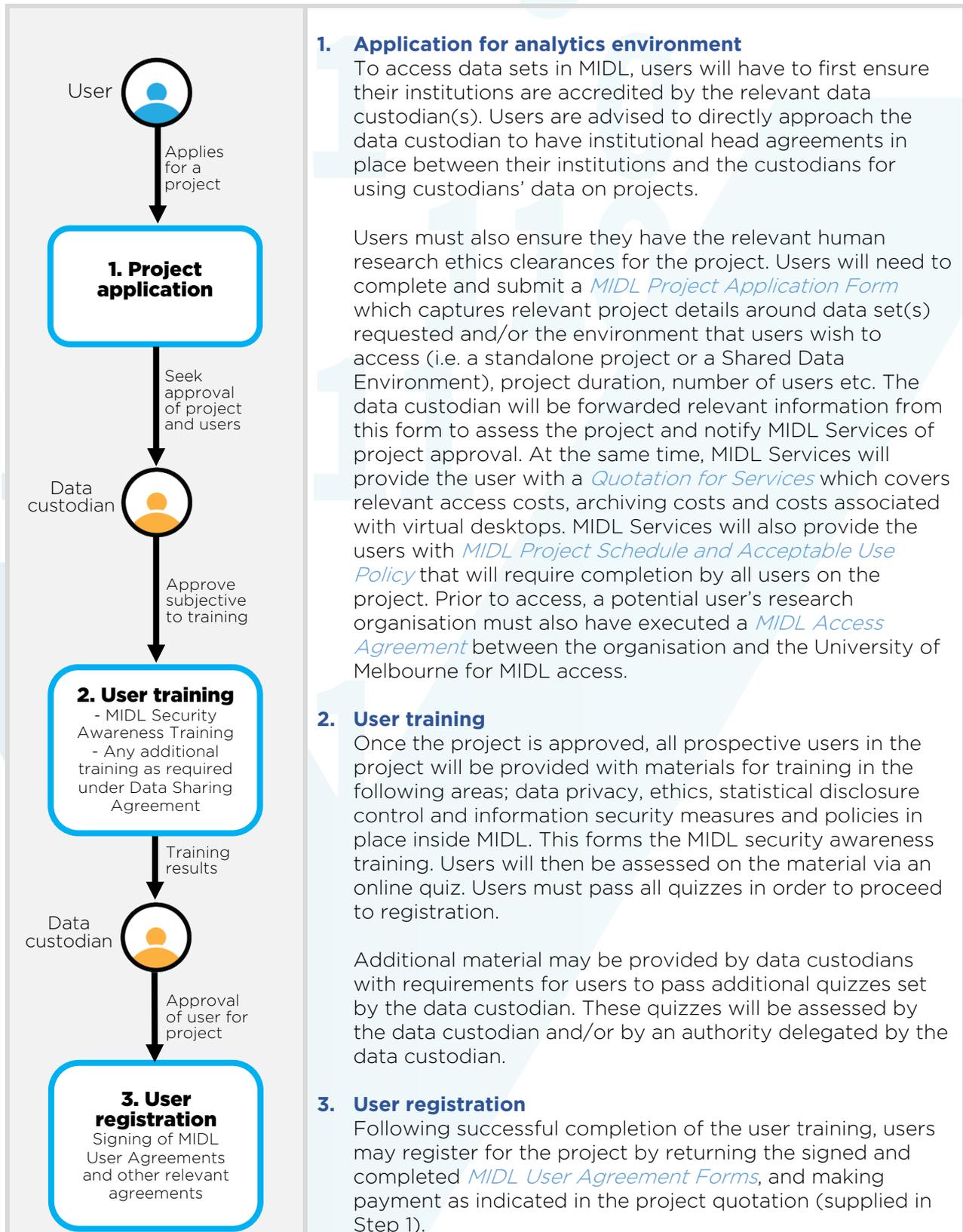
For more information on SDE's and the differences between standalone projects and a SDE, please see Part 1 of the MIDL Information Pack: "What is MIDL?".

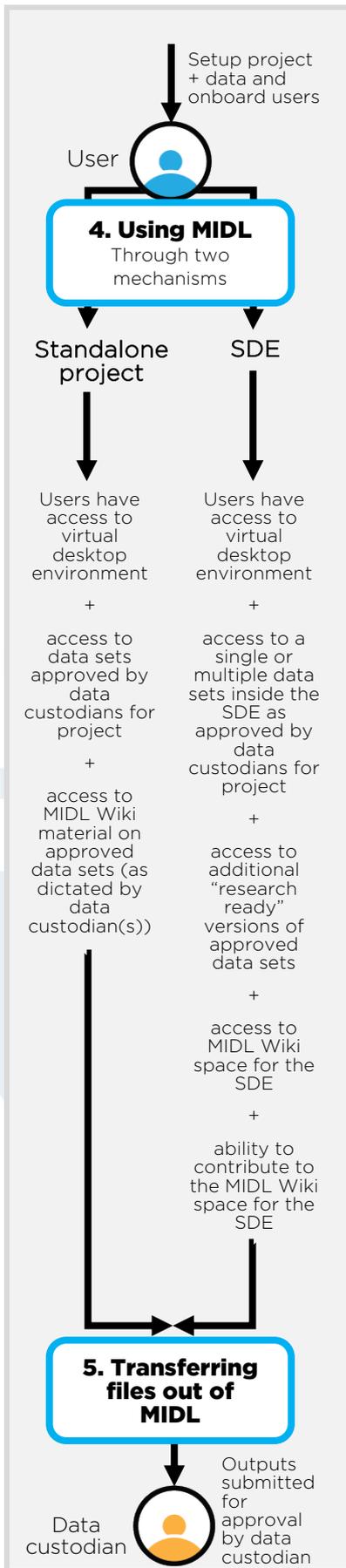
³ Linking can be done at an individual or aggregated level (dependent on data sharing agreement requirements)

The MIDL user journey

The following steps depict an end-to-end user journey, including the steps required to gain access to MIDL, using MIDL and accessing data and the output clearance and project termination.

Key documentation and forms that users require at various stages of this journey are *emphasised* for clarity.





MIDL Services will process the payment and assess the User Agreement Forms. Once complete, and if the data custodian training has been successfully completed, user accounts will be created, and the research environment set up. MIDL Services will provide registered users with the [MIDL Gaining Access Guide](#) that details the process to setup their computer(s) for accessing MIDL.

4. Using MIDL

Users access MIDL by visiting <https://sso.midl.unimelb.edu.au/>. From here users can enter their username, password and access token provided through their Okta Verify app on their smartphones to access their Citrix Launchpad web page. This page provides users with a list of project workspaces that they have been approved to access.

From here users can select which remote virtual desktop to view through a special linking application from their local computing environment. More details on specific steps to access the MIDL are covered in the [MIDL Gaining Access Guide](#)

MIDL virtual desktops run powerful, hardened versions of Microsoft Windows 10 Education which can have a range of statistical and analytical packages pre-loaded for data analysis. This includes Stata, SPSS, SAS and R Studio. Other software applications include Acrobat Reader, Microsoft Office Suite and popular text editors for programming.

NOTE: The initial list at launch is not indicative of long-term software packages and will be updated on a regular basis. Additional software packages can be requested using the MIDL Service Desk from inside MIDL.

Inside MIDL, a user cannot access their email or the internet⁴, print documents or copy/paste text or images in or out of the environment. There will also be no ability to copy data to any removable media such as USB memory sticks or removable hard drives. These functions are disabled to minimise the risk of unauthorised misuse of the data or the data being taken out of the environment against the control of data custodians. To take results and/or output from analysis outside the MIDL environment, users must use the output clearance and vetting mechanisms of the data curation management system.

5. Transferring files in and out of MIDL

Users can use the data curation management system (Citrix ShareFile) to send files into their MIDL project workspace or take files (or results) outside the environment.

⁴ Some websites may be accessible based on data sharing agreements e.g. access to ABS Census data products to enrich analysis.

↓
Outputs approved to
be taken outside of
MIDL

6. Getting help

MIDL
Service
Desk

Access
issues
Project
onboarding
queries
DSA
queries
Technical
issues
Request
additional
software

MIDL
Wiki

Access to
material re:
data from
MI staff
and other
approved
users

All inbound and outbound files will be reviewed a member in the MIDL Services, MIDL Data & Analytics, the data custodian or relevant authority to ensure the files are compliant with ethics approval and data custodian requirements. Where data custodian(s) and/or data sharing agreements allow it, there is the option to have free file ingress and egress permissions at a project or user level.

6. Getting help

Queries relating to account issues, requesting new software, data queries, technical issues and reporting incidents can be raised using the MIDL Service Desk application inside the MIDL environment. Tickets will be assessed by a member in the MIDL Services team and allocated accordingly to either MIDL Services, MIDL Data & Analytics or Cyconsol based on issue classification. All access queries can be emailed to servicedesk@midl.unimelb.edu.au. **NOTE:** No queries relating to any data set, wiki page and/or data memo can be raised using email as this will be a breach of your relevant data sharing agreements. These data requests must only be made using the MIDL Service Desk from inside the MIDL environment.

Users may access the MIDL Wiki pages from within the MIDL environment. This is a web-based online resource where users can access approved pages of resources that contain additional material relating to a particular data set or data sets. When allocated to a SDE, users also have the capability to contribute to the MIDL Wiki page with sample files and data documentation generated as part of project activities.

Support services

A detailed description of support services available for researchers is outlined in Part 1 of the MIDL Information Pack.

Contact information

The MIDL team is available to answer specific questions you may have about the MIDL platform. Please contact servicedesk@midl.unimelb.edu.au if you would like to discuss any further details.

Document History

Revision history

A version number X refers to a version of this document that discusses the feature set and/or processes for a specific release of MIDL. E.g. a document version number of 1.12 is the 12th version of this document for the MIDL platform at launch.

Version number	Author	Date	Details of changes
1.0	Rajeev Samarage	18/04/21	Final draft of version 1
1.1	Rajeev Samarage	09/06/22	Annual review. Updated sections on legal agreements.

Approvals and endorsements

Name	Position	Date approved
Prof. A. Abigail Payne	Director, Melbourne Institute	18/04/21

sso.midl.unimelb.edu.au/

© 2021, Melbourne Institute: Applied Economic & Social Research, University of Melbourne



Melbourne
Institute
Data Lab
In collaboration with Cyconsol