SEAD

Information for prospective users

AIHW ICT

Version 1.0

December 2025

Table of Contents

1. Introduction

The Secure Environment for Analysing Data (SEAD) is a scalable and secure cloud computing environment for Australian Institute of Health Welfare (AIHW) (the Institute) which enables staff and external users to analyse health and welfare data in the development of evidence-based policy development, impactful research, and improved public health and welfare outcomes.

SEAD is underpinned by Australian Bureau of Statistics (ABS) owned cloud infrastructure, with the AIHW responsible for the management and administration of the private AIHW SEADpod. AIHW ICT and project administration teams jointly facilitate user access to data via approved projects to empower users in securely accessing complex health and welfare datasets.

In parallel to the ABS's Safe Setting protections, AIHW administration of SEAD adheres to legislative, policy and risk requirements (see AIHW Data Governance).

This policy should be read in conjunction with:

- AIHW Corporate Information
- AIHW Privacy Policy
- AIHW Data Governance Framework
- The Five Safes Framework
- Data Sharing principles (ONDC)
- SEAD conditions of use
- NHDH Governance Protocols
- ABS Privacy Policy
- Australian Privacy Principles
- APS Code of Conduct

1.1. Purpose

This document provides prospective users, particularly those in policy and research, with a clear understanding of SEAD and its alignment with AIHW strategic goals of enabling trusted, innovative, and authoritative data use; supporting strategic partnerships; and promoting organisational excellence in data capability and infrastructure.

Further, provided information outlines SEAD compliance with underlying governance and legal frameworks, user conditions of use, and skill requirements, stakeholder roles and responsibilities, and additional resources relevant to prospective SEAD users.

1.2. Scope

This document aims to provide prospective users of SEAD with information and guidance on data governance processes, including SEAD access pathways and support resources, system information, and conditions of use to support the ethical, secure, and efficient use of data and resources. It also highlights the key enterprise stakeholders such as the National Health Data Hub (NHDH), related Data Custodians and AIHW ICT and their role in enabling access, and collaboration for users.

Note: This document is subject to modification in response to changes to the governance, system or information needs.

2. Governance

2.1. About Our Data

The Australian Institute of Health and Welfare (AIHW) provides trusted statistical data to government agencies and the community to promote discussion and inform decisions.

SEAD infrastructure and governance controls allow access to project data upon data custodian approval. User access to data in SEAD may be restricted due to:

- conditions set by the AIHW Ethics Committee
- agreements with the suppliers of data
- practicalities in providing access to older collections
- data quality considerations
- objectives outlined in project proposals and subsequent project approvals

Some projects in SEAD are AIHW linked data assets, for example, the NHDH. For further information on and guiding principles in working in data linkage projects see AIHW Linked Data Assets.

2.2. Data Governance

The AIHW's Data Governance Framework ensure that data use aligns with national legislation, and ethical standards, and strong data governance arrangements:

Key components include:

- AIHW compliance with the Privacy Act 1988 and the AIHW Act 1987
- clear data governance structures and roles
- AIHW systems and tools to support data governance
- AIHW data-related policies, procedures, and guidelines

Our data governance arrangements apply to all data we hold, including those:

- collected and/or enhanced by us
- collected on our behalf (for example under agreements)
- obtained from any third party(ies).

Specific governance arrangements apply for some AIHW linked data assets within SEAD – for example, see NHDH Governance Protocols for more information.

2.3. AIHW Standards

The AIHW has embedded The Five Safes Framework to ensure data is used safely, ethically, and for public benefit. *Users of data must demonstrate the knowledge, skills and accountability required to meet these standards.*

2.4. Confidentiality and access

The AIHW holds nationally significant health and welfare datasets that underpin evidence-based policy, related services, and strategic research. Access to data in SEAD requires user agreement to the *AIHW Confidentiality Undertaking* (Section 29 of the AIHW Act 1987) to ensure ethical use and safeguarding of public trust in data-driven initiatives.

3. System Governance Information

SEAD provides a secure environment for analysing complex datasets to support strategic research and policy development by offering:

- high level Data Security (ABS) protections which facilitate research
- governance controls allowing tailored project visibility and data delivery
- provides industry leading infrastructure and modern data science tools

Source: Features ABS SEAD

3.1. Multi-factor authentication requirement

To gain access to SEAD users must set up of multi-factor authentication (MFA) to personally verify their identity and gain access to the platform. MFA is a security measure that requires two or more proofs of identity to grant access. For further information see ABS System security information and Data Security (ABS).

4. Data Holdings in SEAD

4.1. The NHDH

The National Health Data Hub (NHDH) is a major national de-identified linked data system that draws together core government administrative health, welfare, disability and aged care datasets. The NHDH facilitates person-based and longitudinal studies to support contemporary medical research, inform health, welfare, disability and aged care services planning and policy development by government and non-government organisations, and monitoring of service delivery.

To contact the NHDH email NHDH@aihw.gov.au. For NHDH specific information please visit NHDH website and NHDH Governance Protocols.

4.2. Other Data in SEAD

Australian Teacher Workforce Data (ATWD) brings together data from higher education providers, teacher regulatory authorities, and from teachers themselves on topics important to the national teacher workforce. The ATWD offers crucial insights into the pipeline of new teachers as they move through initial teacher education then enter the workforce and helps us understand the workforce experiences of our current teachers.

For information on access ATWD data contact atwd@aihw.gov.au. For further information on access and guiding principles for data linkage projects see AIHW linked data assets.

5. Roles and Responsibilities

5.1. Data Custodians

The Data Custodian is an AIHW staff member with delegation to exercise overall responsibility for a specific data holding (or Data Asset e.g. the NHDH), in accordance with legislation, policies, guidelines and any specific conditions for use applicable to that data collection. They have responsibility and oversight of the data lifecycle to ensure compliance with data governance policies.

5.2. Project Administration Aims

AIHW Project Administration is data asset specific (e.g. the NHDH, and ATWD each have a project administration team), and they are responsible for:

- Providing a trusted platform for users to access de-identified, person-centric data across health, welfare, disability, and aged care sectors
- Facilitating user access to evidence-based that informs government programs, policy decisions, and public health initiatives
- Uphold strong governance principles including privacy, ethical use, and community trust through rigorous data governance protocols
- Offer clear processes for project applications, renewals, amendments, and invoicing, ensuring efficient user support and administration

5.3. ICT Administrator Support Aims

AIHW ICT is responsible for:

- Providing the system administration of the AIHW instance of SEAD
- Providing clear guidance on ICT processes and functions to approved users
- Facilitating user access to projects, data and approved software tools
- Managing and resolving user in-scope queries related to access and system functionality

- Delivering timely and in-scope support for SEAD users and approved project-related ICT needs efficiently through a ticketing system, aiming to respond within two days.

Please note: AIHW support does not extend to analytical coding or data interpretation.

5.4. SEAD User Responsibilities

A user is defined as any person approved to access data in SEAD by the corresponding Data Custodian. Users include AIHW staff, or contractors, or any individual from a government or non-government organisation.

Prior to accessing SEAD, users must:

- **Review** User Competency Requirements and SEAD conditions of use
- **Review** Software in SEAD information for project suitability
- **Review** ICT support available
- **Be approved** for project access by Project Administration team
- **Attend** Mandatory data onboarding training and ICT information session

5.5. Project leaders

The project leader is the individual responsible for the management, administration and supervision of a research project in SEAD. Project leaders are responsible in ensuring that users in their project team individually meet SEAD user responsibilities (as outlined in 5.4 SEAD User Responsibilities.

5.6. Mandatory data training and ICT information session

As part of the SEAD onboarding processes, users are required to complete mandatory training delivered by the Project Administration team. In parallel, AIHW ICT deliver an orientation session on system access, governance protocols, compliance requirements, including conditions of use, and secure workspace navigation. These sessions equip users with the foundational knowledge needed to operate within the SEAD to uphold shared responsibility within data governance standards.

5.7. User Competency Requirements

As part of the Safe People assessment of the 5 Safes Framework, users must demonstrate experience and expertise in data analysis and programming to perform their role effectively.

In addition to User Competency Requirements, the AIHW provides Appendix 1: User Competency Matrix – which outlines skills needed for data analysis in SEAD. These resources are designed to help users self-assess and determine minimum skills needed for their study.

- Proficiency in using very large datasets, data transformation, efficient data coding and processing, resource optimisation
- Experience with understanding file formats (e.g. parquet, SAS formats) and compatibility, memory and computational constraints, scalability and resources limits
- Data analysis techniques, such as statistical modeling, and data visualisation

- Data quality and validation: ability to design and implement basic data quality checks (e.g. completeness, consistency, plausibility, reconciliation against known totals).
- Familiarity with techniques for comparing versions of datasets over time (e.g. drift checks, change detection).
- Experience with programming languages such as SAS, Python, R, or SQL, and familiarity with analytical libraries and packages
- Ability to setup and maintain efficient reproducible workflows, apply secure coding practices, and use version control in collaborative environments
- For users intending to use *Databricks* within SEAD, they need to demonstrate experience in:
    - **Managing and working in notebooks and/or repo**s. Understanding clusters and cluster policies (job clusters vs all-purpose clusters, autoscaling, appropriate driver/worker sizing).
    - **Working with the Lakehouse environment** (e.g. Delta tables, catalog / schema / table concepts, use of meta store/Unity Catalog where applicable).
    - **Cost and resource stewardship**: understand that consumption-based costs apply to their project and the cost implications of cluster choices, long-running interactive clusters, and inefficient queries. Ability to design workloads that are efficient and shut down or clean up resources when no longer required

For further information on Databricks see 9.1.2 Databricks and Databricks Access Request Form (Appendix 2).

For further assistance in assessing User Competency Requirements see Appendix 1: User Competency Matrix.

Users will be requested by Project Administration to demonstrate their competency level and experience within a project proposal.

5.8. SEAD Conditions of Use

To access data in SEAD, users **must** agree to comply with the following requirements and obligations outlined in this document, including:

- Keeping passwords safe and do not share your login details.
- Using SEAD in secure locations i.e. office environments, and secure other locations.
- Do not “screen share” your desktop over online meetings, with any unauthorised team members.
- Do not capture any onscreen information, including electronically (e.g. screenshots, meeting recordings) and handwritten notes.

- Always locking your workstation when not in use.
- Making no attempt to identify individuals or organisations in the data.
- Do not attempt to avoid, override, or bypass the system or procedures.
- Utilising output and input clearance procedures as advised by the AIHW in all instances.
- Making no attempt to match or link information with any other data.
- Notifying AIHW Service Desk promptly of any suspected breach of security relating to SEAD or related data.
- Do not discuss content or analysis of a SEAD project if the individual(s) are not listed as analysts or discussants within an approved project proposal form.

A breach of these conditions may result in sanctions which may include, but are not limited to, the revocation of access to your project in SEAD permanently or for a set period.

6. How to access data

6.1. Project Administration contacts

AIHW ICT and Project Administration teams collaborate to provide secure, streamlined access to data. In the first instance, prospective users should contact the Project Administration team as per the information in the table below.

When to contact Project Administration team
For further information on access and guiding principles for data linkage projects see AIHW linked data assets. For NHDH specific information please visit NHDH website and NHDH Governance Protocols.
Please contact the Project Administration team regarding: • Application to access a data asset • Invoicing and pricing (inc. SAS, Databricks) • Project extension (renewal) • Project amendments For **NHDH projects** – NHDH@aihw.gov.au inbox. For the **linking of additional datasets to the NHDH** for analysis purposes (referred to as *NHDH+n projects*) – linkage@aihw.gov.au inbox.
Other contacts: to contact the ATWD project team contact atwd@aihw.gov.au inbox.

7. Contacting ICT

7.1. ICT Support Scope

In addition to AIHW ICT aims, AIHW ICT functions include:

- enabling user access to data, projects, software and tools
- facilitation of ingress/egress as approved by data custodian or Project Administration team

Out of scope
Assistance with complex code analysis, or coding efficiency in Databricks or other software fall outside the scope of AIHW support.

7.2. ICT contacts

ICT can be contacted via secureenvironments@aihw.gov.au for the following project functions.

- User access to platform, projects, and connection/access to data
- Password and access queries
- Ingress/Egress requests

8. Virtual Machines

Users once approved will access SEAD via a Virtual Machine (VM). The VM is provisioned to an individual user via their individual login username and password. Virtual Machines cannot be shared among project users. Large size VMs are allocated to users as standard practice. For further information see ABS information on Virtual Machines.

Standard virtual machines

Name	Windows server	CPU	RAM
Small	Small Windows 10 DSVM	CPU Cores 2	8GB
Medium	Standard Windows 10 DSVM	CPU Cores 2	16GB
Large	Standard Windows 10 DSVM	CPU Cores 2-8	64GB

9. Software in SEAD

Prospective users should assess available tools and software closely, noting that while similar tools may exist in other environments, software versions are often inconsistent.

All software in SEAD is maintained, licenced and administered by the ABS (as *Platform Owner*) whom provide system information and a list of included software versions available in SEAD – see Available Features.

9.1. Add-on Software (by approval, charges apply)

9.1.1. SAS 9.4 (EG 8.2)

SAS 9.4 (EG 8.2) is available upon request (charges apply) for use in SEAD. SAS is charged at a per user cost. Please contact the Project Administration team for full cost details. Refer to Project Administration contacts.

9.1.2. Databricks

Databricks is available to approved users. The process to utilise Databricks in SEAD involves submission of a Databricks Access Request Form (Appendix 2) to provide details of prior experience and analysis justification for approval.

Refer to Project Administration contacts for further information on process, charges, and eligibility. If using Databricks with NHDH data see NHDH access, grant submissions & costs.

See also SEAD Conditions of Use and ABS Databricks information.

Databricks requires **specialised knowledge and is not a standard inclusion for projects.**
It is **recommended** that advanced level analysts request access to Databricks according to Appendix 1: User Competency Matrix, and additionally have prior experience in **managing notebooks clusters, machine learning, and collaborative workspaces.**
Databricks costs are consumption based, increasing with processing power, data size, and code efficiency – making potential cost estimates difficult.
Invoicing process: invoicing of Databricks is consumption based and are issued on a biannual basis: a first invoice is sent for payment of upfront access, and a second invoice is sent post-financial year for usage (based on consumption). Refer to Project Administration contacts to obtain further information on process, charges, and eligibility.

10. Incident management

Users are asked to notify AIHW Service Desk promptly of any suspected breach of security relating to use of the SEAD platform or data. The AIHW will ensure the effective management of and response to Information Security incidents to maintain secure operations, see AIHW Vulnerability and disclosure policy.

10.1. NHDH Response Plan

All NHDH users and discussants must comply with the **NHDH Response Plan**. Project Leaders are required to report a breach (suspected or actual) to the **AIHW NHDH Data Custodian immediately** via NHDH@aihw.gov.au. See NHDH resources information.

10.2. Review and Update of Policy

This document is subject to modifications in response to changes in technology services and support needs. This policy is reviewed annually.

Appendix 1: User Competency Matrix

As part of the Safe People assessment (see 5 Safes Framework) users must demonstrate experience and expertise in data analysis to ensure that data is handled safely and accurately, and in compliance with governance and legal frameworks, data integrity and security.

In addition to User Competency Requirements, the AIHW provides this Matrix to outline skills needed for data analysis in SEAD. These resources are designed to help users self-assess and determine minimum skills needed for their study. Skills required within a project will also depend on many variables (e.g. analysis question, data set size and variable use, formats etc).

RECOMMENDATIONS

To analyse data in **SEAD**: users must have at least a **Beginner to Intermediate Data Analysis Competency.**

For **Databricks,** an **Advanced level Data Analysis Competency is recommended**.

COMPETENCY AREA	BEGINNER	INTERMEDIATE	ADVANCED
Large Dataset Handling	Understands principles of data integrity, can clean and preprocess datasets (up to x # of rows using basic tools (e.g., pandas, SQL).	Can transform and process large datasets using workflows	Optimises resource usage, manages complex transformations at scale. **Databricks:** understands and has experience with Lakehouse environment (e.g., Delta tables, catalog / schema / table concepts, use of meta store/Unity Catalog where applicable).
File Format & Compatibility	Works confidently with multiple common formats (CSV, Excel, JSON) and understands encoding issues and conversions.	Works with formats like Parquet, SAS, JSON	Manages memory constraints, scalability, and system limitations
Data Analysis & Visualization	Creates descriptive statistics and visualisations using libraries (e.g., here?) and interprets basic trends.	Applies statistical methods and creates visualisations	Implements advanced analytics and develops complex dashboards
Programming Skills	Writes structured scripts in Python or R, uses functions, and applies error handling for data tasks.	Uses multiple libraries and frameworks for data processing	Develops modular, reusable code and integrates advanced techniques (e.g., ML) **Databricks:** understands clusters and cluster policies (job clusters vs all-purpose clusters, autoscaling, appropriate driver/worker sizing).
Reproducible & Secure Workflows	Implements basic version control, documents workflows, and always applies simple security practices (e.g., Password security or can find these resources easily).	Uses version control and follows coding practices	Builds collaborative workflows and ensures security compliance. **Databricks:** cost and resource stewardship – understanding consumption-based costs project costs and implications of cluster choices, long-running interactive clusters, and inefficient queries. Ability to design efficient workloads, shut down or clean up resources no longer required

Appendix 2: Databricks Access Request Form

Please complete the following sections to request access to Databricks and demonstrate your experience with the platform.

Purpose of Request

Describe why you are requesting access to Databricks and how it supports your approved project.

Experience with Databricks

Outline your experience with Databricks, including tools and features you have used (e.g., Apache Spark, SQL, Python, R).

Expected Benefits

Explain how access to Databricks will benefit your project and improve your data analytics capabilities.

Conclusion

Summarise your request and affirm your ability to use the platform responsibly and effectively.

User Acknowledgement:

I acknowledge that access to Databricks requires an advanced level of experience and capability in data analytics and cloud-based platforms. I understand that usage and consumption costs apply and that no refunds will be provided if I am unable or choose not to use the platform after access is granted.

Signature: _____

Date: _____

Appendix 3: SEAD ABS resources

ABS information

Users	System	Tools and Software
SEAD General information	Navigating the SEAD portal	Available features (Software)
SEAD User guide		SEAD Databricks
SEAD FAQ		

SEAD Information for users

File formats determine how data are created, stored, and read. Each file format has a unique extension. The terms “file format” and “file type” are often used interchangeably. Certain formats are designed to compress data for fast processing. Files in SEAD are available in SAS and Parquet format.

SAS

The SAS *sas7bdat* form is the file extension for data products (individual data files) within SEAD. SAS7BDAT files is a binary database storage file. Users wishing to use, read and manipulate these files using SAS software will require SAS enabled on their VM which incurs an annual cost.

PARQUET

Apache Parquet format uses a columnar storage format that addresses big data processing challenges. Unlike traditional row-based storage, it organises data into columns. This structure allows you to read only the necessary columns, making data queries faster and reducing resource consumption.

CHOOSING A FORMAT

Selection of a data format is about ensuring the format matches your downstream intended use for the data. Use code optimisation formatting should match the storage method and data usage to effectively manage engineering time and other resources.

Source: Explaining the Row vs. Columnar Big Data File Formats

Glossary of Terms – SEAD

Term	Definition
Project data	The local objects (e.g. datasets) project users have created and saved in their Project or Output drives. This may include code or local copies of accessible products. Project and Output folders within each workspace are backed up each night and retained for 14 days. SEAD Administrators are responsible for clearing and outputting any user analysis from the system.
User	Any one person who has been approved to access data in SEAD by the corresponding Data Custodian. Includes staff and contractors from non-government organisations, or individuals.
Project Leader	The project leader is the individual responsible for the management, administration and supervision of the NHDH research project. The project leader is typically the lead researcher of the project.
Data Custodian	A Data Custodian is an AIHW staff member with delegation to exercise overall responsibility for a specific data collection, in accordance with legislation, policies, guidelines and any specific conditions for use applicable to that data collection.
The NHDH	AIHW National Health Data Hub (NHDH) repository which contains hospital services data, Medicare Benefits Schedule, (MBS) and Pharmaceutical Benefits Scheme (PBS) data, National Death Index (NDI) data, Australian Immunisation Register (AIR) data, National Aged Care Data Clearinghouse data, National Disability Insurance Scheme data and Australian and New Zealand Intensive Care Society (ANZICS) data.
ATWD	The ATWD brings together data from higher education providers, teacher regulatory authorities and teachers themselves on topics important to the national teacher workforce.
SEAD partner	“SEAD partners” are government agencies who adopt a SEADpod and inherit exclusive administration of that self-contained environment through self-service features.

SEAD Project	In the SEAD environment, workspaces are based around and linked to a Projects. A Project represents a shared space for approved users to work in, access data and store all their Project files self-contained from other Projects.
SEAD Project Shared Library	All users can see all files in the Shared Library. The ABS will continue to maintain/upload support information, such as Statistical language documentation, ANZSIC classification, code and packages. Files cannot be saved to this drive by anyone other than ABS System Administrators.
Virtual Machine	As part of Security features of SEAD, users are allocated one Virtual Machine for each Project they are approved to access. Virtual Machines are automatically destroyed and rebuilt every 30 days for security and maintenance purposes. Users are reminded about a rebuild three days ahead of their rebuild and again 24 hours prior. Rebuilding can take up to 45 minutes to complete. Virtual Machines (through use of a login) cannot be shared among approved analysts in SEAD.