

AIMS Data Repository Preservation Plan

Introduction

The preservation plan outlines the AIMS Data Repository's applying to the long-term retention and preservation of research data held within the Institute's repository. The FAIR data principles guiding the repository's approach provide a framework for implementing the preservation strategy. Preservation decisions are made with the repository's mission in mind.

Purpose

The repository was developed to carry out the functions to collect and disseminate information from AIMS' research projects. The AIMS Data Repository has been developed by the AIMS Data Systems Engineering team (formerly Aims Data Centre). As data is a core part of AIMS' business, the repository and its services maintain critical functions for archiving and sharing data within AIMS as well as the broader scientific community. Data are considered valuable assets and records of AIMS and, as such, shall be managed in a way that preserves and enhances the value.

AIMS Data Repository Mission Statement

The mission of the AIMS Data Repository is to capture, preserve and provide access to AIMS research data to support the institute's publicly funded research and data assets. The repository is committed to securely storing and publishing research data to enable reproducibility, promote research integrity and enhance access through innovative data tools linked to informative metadata.

<https://www.aims.gov.au/data/about-aims-data-repository>

Scope

The preservation plan applies to the AIMS Data Repository, services for data management and long-term data archiving infrastructure. The plan applies to digital processes applied to acquiring data resources at AIMS and the digital materials held by the repository.

Purposes of the Repository preservation plan are to:

- to ensure the authenticity and integrity of the data held within the repository collections
- Enable long term sustainable accessibility to research data generated by the institute
- provide reliable, long-term infrastructure to ensure the digital resources are managed throughout their lifecycle, including migrations where needed
- keep data available for reuse, and available replication and verification purposes



Alignment with National Policies

AIMS has a Records Authority produced through the National Archive of Australia (NAA), which outlines the responsibilities towards retention of research data and associated resources. This Records Authority was drafted in direct consultation with AIMS' Information Services team and NAA and is based on the National Archives Authority 37 (2022/00304018). The Records Authority (2011/00292508) is supported through the AIMS Records Management Procedure which notes the AIMS Data Repository is the designated long term storage facility for data, and all data deposits must be accompanied by rich metadata.

<https://www.naa.gov.au/sites/default/files/2019-12/agency-ra-2011-00292508.pdf> AIMS Records Authority (June 2011)

Further governance is described in the AIMS Responsible Conduct of Research Framework, which is based on the Australian Code for the Responsible Conduct of Research. The code sets out the expectations and responsibilities of both researchers and institutions. Specifically, the framework outlines the obligations for the management of research data:

- *AIMS is committed to implementing best practice in its record management practices and systems. Researchers are expected to:*
- *Retain research data and primary materials, manage storage of research data and primary materials, and maintain confidentiality of research data and primary materials.*

Responsibilities

Roles and Responsibilities are outlined in the AIMS Data Access Policy.

Repository: The Repository's objective is to identify and preserve scientific data generated by the Institute, make these resources publicly available and promote their use, value and impact. The repository assumes responsibility for the safe and secure storage and long-term preservation of research data.

The role of the AIMS Data Repository is documented (<https://www.aims.gov.au/data/about-aims-data-repository>) to:

- collaborate with AIMS' Information Communications and Technology (ICT) team to provide services for the safe and secure storage and long-term preservation of research data
- enable creation of metadata records and apply persistent identifiers, including Digital Object Identifier (DOI) where appropriate, to all data records
- assume responsibility for management and accessibility of datasets and to promote their use, value and impact
- make every effort to work with data custodians to ensure data standards are maintained, and the FAIR data principles are supported
- carry out preservation actions to mitigate data loss over time and maintain long term preservation standards.



Research/ Data contributors: Responsibilities of submitters are outlined in the AIMS Data Access Policy, section in the section Data Management Actions (<https://www.aims.gov.au/data/data-policy>) and include the below data management actions

The following fundamental data management actions apply:

1. Projects that collect or generate research data must have a Data Management Plan The project leader(s) is/are responsible for creating the Data Management Plan as part of the Project Concept Process and addressing any concerns raised by ICT or Data Systems Engineering teams including the possibility of having to budget for data storage and the work needed to prepare research data for distribution.
2. Datasets must have at least one metadata record lodged in the AIMS Data Repository. Occasionally, multiple metadata records are used to provide structure to the component parts of a dataset, particularly if usage constraints are different for different parts of the dataset. Metadata allows AIMS to keep a record of its data assets
3. The data custodian(s) is/are responsible for generating the metadata record. Metadata records must be finalized during the execution stage of the project and published with curated archived data prior to the closure stage
4. Datasets must be lodged in an approved repository to ensure data continuity and preservation and comply with;
 - The data custodian(s) is/are responsible for lodging the data in the AIMS Data Repository or other approved repository
 - Data formats are important to maintain accessibility over the long term. It is preferred that data files submitted to the repository are in non-proprietary, open-source software formats, and well documented.
 - It is the responsibility of the data custodian(s) to ensure data integrity and accuracy.

Formats

The AIMS Data Repository prefers data submissions to be received in open formats and open-source software, to enable long term preservation of data assets and prevent data loss for obsolete software. A list of recommended file formats for long term preservation are provided through the Factsheet 'Data Archiving Formats'.

It is important to consider the longevity of the data format and whether it is proprietary software for data preservation. Open access formats are preferred for archival file longevity and to prevent data loss (.csv, .pdf, .txt, .tiff, .shp, ascii files or netCDF). Excel files are also commonly used but will require format updates from time to time by the repository as new versions of the program are released.

For more detailed information on Formats, please see the Repository User Guide (Submissions Assistance) or Fact Sheet "Data Formats".

Persistent identifiers:

The Repository employs persistent identifiers to enable long term preservation strategies.



UUID: The metadata and on-premise data store are linked through the use of the universally unique identifier (UUID). The UUID is embedded into the metadata record identifier across all copies of the metadata, as well as the data store name, and forms part of the metadata link.

Digital Object Identifiers (DOI) are manually minted by the repository for high value datasets, where additional curatorial actions may apply. The repository maintains the DOI register where the resource must be permanently hosted and persistently available. The repository assumes the commitment to ensure the resource's continuity. The DOI will be used as the resource identifier on metadata records (including citation) in preference to the UUID. The URL with the UUID will be used at the landing page for the DOI.

Back Up and Archive:

The AIMS Data Repository uses stable software technology to ensure integrity and authenticity during the process of ingestion, archival storage and data access.

This includes on premise server archive and Oracle databases which are subject to ICT Back Up procedures.

On-premise data are replicated to cloud infrastructure to facilitate public data access via Amazon Web Services (AWS). Most data are stored on Amazon S3 storage, or stored in Amazon Relational Database Service (RDS).

These data are subject to

<https://docs.aws.amazon.com/AmazonS3/latest/userguide/DataDurability.html>

Backup and restoration processes are in place for cloud deployed components where both infrastructure and code are restorable through automated build and deployment processes.