Assessment Information

[CoreTrustSeal Requirements 2017–2019](#)

Repository: CSIRO Data Access Portal
Website: https://data.csiro.au/dap/
Certification Date: 25 October 2018

This repository is owned by: Commonwealth Scientific and Industrial Research Organisation

CSIRO Data Access Portal

Notes Before Completing the Application

We have read and understood the notes concerning our application submission.

True

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:

CORE TRUSTWORTHY DATA REPOSITORIES REQUIREMENTS

Background & General Guidance

Glossary of Terms

BACKGROUND INFORMATION

Context

R0. Please provide context for your repository.

Repository Type. Select all relevant types from:

Institutional repository

Comments

The Commonwealth Scientific and Industrial Research Organisation (CSIRO) is Australia's national science agency and one of the largest and most diverse research agencies in the world.

This application is for CSIRO's Data Access Portal (DAP). The DAP is an institutional repository and the mission is to publish research data of significance to CSIRO and to ensure its ongoing access. The DAP provides access to research data, software and other digital assets published by CSIRO. A collection in the DAP includes metadata, files and a persistent unique identifier to describe and identify the data assets. Data assets from a wide range of disciplines is included with 17 of the 22 fields of research represented from the Australian and New Zealand Standard Research Classification (ANZSRC).
The repository is developed and maintained by CSIRO Information Management & Technology to facilitate sharing and reuse.

The DAP provides a number of levels of access:
• Public access to both metadata and data - metadata can be viewed and data can be directly downloaded from the repository.
• Public access to metadata only - metadata can be viewed. Further information and in some cases data can be accessed by enquiring.
• CSIRO staff access to both metadata and data - metadata can be viewed and data can be downloaded from the repository after authentication.
• CSIRO staff access to metadata only - metadata can be viewed. Further information and in some cases data can be accessed by enquiring.
• CSIRO project staff/affiliates access to metadata and data - specified and authenticated project members can view metadata and download data.

Comments:
Accept

Brief Description of the Repository's Designated Community.

The designated community of the DAP is broad and therefore the content covers a wide range of scientific research. Data users may include:

• general public

• industry specific groups such as agriculture

• policymakers

• students

• researchers

The research disciplines represented and based on the field of research codes from ANZSRC include in order of most number of collections:

• Physical Sciences

• Environmental Sciences

• Earth Sciences

• Agricultural and Veterinary Sciences

• Biological Sciences

• Information and Computing Sciences

• Engineering

• Chemical Sciences

• Technology

• Mathematical Sciences

• Economics

• Built Environment and Design

• Medical and Health Sciences

• Psychology and Cognitive Sciences

• Studies in Human Society

• History and Archaeology

• Language, Communication and Culture

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Level of Curation Performed. Select all relevant types from:

A. Content distributed as deposited, B. Basic curation – e.g. brief checking; addition of basic metadata or documentation, C. Enhanced curation – e.g. conversion to new formats; enhancement of documentation

Comments

The level of curation can be:

• Level A - content distributed as deposited

• Level B - basic curation

• As an optional service Level C - enhanced curation is offered on request

Outsource Partners. If applicable, please list them.

N/A

Other Relevant Information.

Research Data Australia is a web-based discovery service that helps the public find, access, and reuse data for research from over one hundred Australian research organisations, government agencies, and cultural institutions. To increase the discovery of publicly accessible collections from the DAP, Open Archives Initiative Protocol for Metadata Harvesting is used to harvest Registry Interchange Format - Collections and Services records to Research Data Australia. Research Data Australia does not store the data itself but provides descriptions of, and links to, the data. https://researchdata.ands.org.au/csiro/561690/

The DAP is listed in the Registry of Data Repositories re3Data.org under ID: r3d100010734. http://www.re3data.org/repository/r3d100010734

DAP records are included in the Clarivate Data Citation Index - Web of Science. http://wokinfo.com/cgi-bin/dci/search.cgi

The DAP has been recently added to the list of recommended repositories by PLOS.

As of November 2017, the DAP holds more than 877 TB of data across more than 3000 collections of which 2090 are available for public access. There were over 350 unique depositors. The number of discoverable data assets in the DAP is expected to continue to grow significantly.

Over the past year, December 2016 to November 2017, the average usage per month included more than: 4,000 visits, 16,000 total page views and 18,000 file downloads.

In a report by Anderson, Reeson and Box (2017) it was estimated the annualised total economic value generated by the DAP to its users is in the order of $67 million.

References:

Sanderson T, Reeson A and Box P (2017) Understanding and unlocking the value of public research data: OzNome social architecture report. CSIRO, Canberra. Viewed 24 October 2018, <https://doi.org/10.4225/08/58a5e8d940904>.

Australian Bureau of Statistics (2008) 1297.0 - Australian and New Zealand Standard Research Classification. ABS, Canberra. Viewed 24 October 2018, <http://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/1297.0Main+Features12008>.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

ORGANIZATIONAL INFRASTRUCTURE

I. Mission/Scope

R1. The repository has an explicit mission to provide access to and preserve data in its domain.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Response:

The objective of the CSIRO Research Data Service is to ensure that the organisation captures, publishes and manages the right data to support innovation, collaboration and scientific integrity. The Research Data Service will:

• Ensure research data under the custodianship of CSIRO is securely stored, easily located and where appropriate, accessible to others for reuse.

• Provide the ability to publish data to support reproducibility and research integrity.

• Manage data in an increasingly collaborative research environment.

• Respond to government and funding body requirements to share the data from publicly funded research.

CSIRO Research Data Service is a collaboration between Information Management & Technology teams as well as research partners to deliver a holistic solution for the management of research data in CSIRO. The DAP is the point of capture, and a discovery portal, for CSIRO's research data assets.

Through data.gov.au, Research Data Australia and other subject discipline portals, CSIRO will make available data collections and/or descriptions of data collections across a broad range of disciplines to increase the availability of data for reuse.

CSIRO recognises the significant value in the data generated by its substantial investment in research. Durable research data is essential to justify, and defend when required, the outcomes of the research. Research data may also have value

for other researchers or the wider community.

The Research Data Management Procedure provides the basis for management of CSIRO's digital research data assets in terms of storage, retention, and accessibility for reference, use or reuse.

Research data assets are those selected for long term storage and management to enable validation of research findings and re-use of high value or unique data. This may include third-party data which may have originated within CSIRO or elsewhere.

The Research Data Management Procedure is consistent with the requirements of the Australian Code for Responsible Conduct of Research, the recommendations of the Australian Government 2.0 Taskforce and the research data management requirements of the CSIRO Code of Conduct (National Health and Medical Research Council (2007), Australian Government Information Management Office (2009)).

Evidence:

Mission Statement
https://confluence.csiro.au/display/daphelp/Mission+Statement

Research Data Management Procedure (Internal Document)

Research Data Service Functional Overview (Internal Document)

CSIRO Code of Conduct (Internal Document)

National Health and Medical Research Council (2007) Australian Code for the Responsible Conduct of Research. NHMRC, Canberra. Viewed 24 October 2018, <https://nhmrc.gov.au/about-us/publications/australian-code-responsible-conduct-research-2007>.

Australian Government Information Management Office (2009) Engage: Getting on with Government 2.0: Report of the Government 2.0 Taskforce. Viewed 24 October 2018, <https://www.finance.gov.au/archive/publications/gov20taskforcereport/>.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

II. Licenses

R2. The repository maintains all applicable licenses covering data access and use and monitors compliance.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Response:

Depositors have access to CSIRO specific and a subset of Creative Commons licences. Guidance on choosing a licence is available from the Data Deposit Checklist.

Licence deeds in the DAP Users Guide includes information about the CSIRO specific licences. Data users are made aware of and accept the terms of the licence when the data is downloaded. The DAP does not monitor compliance of licences by data users. It is not feasible as data users do not register. Monitoring for attribution or commercial use is beyond CSIRO's resources.

CSIRO:
• Does not record details of any person accessing the data and is not required to identify any person accessing the data.
• Is not required to take any steps to enforce any intellectual property or other rights in the data or the terms of any licence applicable to the data.

Evidence:

Data Deposit Checklist (Internal Document)

Licence Deeds
https://confluence.csiro.au/display/daphelp/Licence+Deeds

Office of the Australian Information Commissioner (n.d.) Legally binding guidelines and rules. Viewed 24 October 2018, <https://www.oaic.gov.au/agencies-and-organisations/legally-binding-guidelines-and-rules/>.

DAP Legal notice and disclaimer
https://data.csiro.au/dap/legal

III. Continuity of access

R3. The repository has a continuity plan to ensure ongoing access to and preservation of its holdings.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Response:

CSIRO is an Australian Government corporate entity. CSIRO's parent began as the Advisory Council of Science and Industry in 1916. CSIRO is constituted by and operates under the provisions of the Australian Government Science and Industry Research Act 1949. CSIRO's main funding is directly from the Australian Government and this funding is in excess of 1 billion per year. As this funding is based on a triennial funding program, if funding was to cease, CSIRO would have approximately 3 years notice to plan for the succession of the DAP's collections. CSIRO considers the possibility of it ceasing to operate or changing its scope as highly unlikely.

The Information Management & Technology Business Continuity Plan covers recovery from loss of building, equipment, technology, human resources, and third parties/suppliers.

Evidence:

Information Management & Technology Business Continuity Plan (Internal document)

DAP Business Continuity Plan (Internal document)

Data Preservation Principles
https://confluence.csiro.au/display/daphelp/Data+Preservation+Principles

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

IV. Confidentiality/Ethics

R4. The repository ensures, to the extent possible, that data are created, curated, accessed, and used in compliance with disciplinary and ethical norms.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

Responsibility is placed on CSIRO staff, who assume the role of depositors and approvers in the DAP, to ensure data published is free from legal and ethical issues. Depositors are required to consider the risks when publishing data using the Data Deposit Checklist. Approvers assess the risk before a data collection is published using the same Data Deposit Checklist.

Depositors and approvers are referred to experts within the organisation to seek further advice prior to publishing their collection. Experts within CSIRO include staff from: legal, information management and technology, ethics, business and commercial development, and communications.

The DAP will collect personal information where the data user is required to email the depositor to access a restricted collection or where a data user submits a question to the repository staff. However, CSIRO will not disclose this information without an individual's consent except where CSIRO may be required by law to disclose certain information.

Evidence:

Data Deposit Checklist (Internal document)

CSIRO Code of Conduct (Internal document)

Human Research Ethics in CSIRO (Internal document)

DAP Approver Training (Internal document)

Ethical Conduct in Human Research Procedure (Internal document)

Privacy Statement
https://data.csiro.au/dap/privacy

National Health and Medical Research Council (2007) Australian Code for the Responsible Conduct of Research. NHMRC, Canberra. Viewed 24 October 2018, <https://nhmrc.gov.au/about-us/publications/australian-code-responsible-conduct-research-2007>.

Office of the Australian Information Commissioner (n.d.) Legally binding guidelines and rules. Viewed 24 October 2018, <https://www.oaic.gov.au/agencies-and-organisations/legally-binding-guidelines-and-rules/>.

Reviewer Entry
Reviewer 1
Comments:
Accept

V. Organizational infrastructure

R5. The repository has adequate funding and sufficient numbers of qualified staff managed through a clear system of governance to effectively carry out the mission.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

CSIRO is an Australian Government corporate entity. CSIRO's parent began as the Advisory Council of Science and Industry in 1916. CSIRO is constituted by and operates under the provisions of the Australian Government Science and Industry Research Act 1949. CSIRO's main funding is directly from the Australian Government and this funding is in excess of 1 billion per year. As this funding is based on a triennial funding program, if funding was to cease, CSIRO would have approximately 3 years notice to plan for the succession of the DAP's collections. CSIRO considers the possibility of it ceasing to operate or changing its scope as highly unlikely.

The DAP is an institutional repository that has been available since 2012. It is a service within the Information Management & Technology department of CSIRO and is supported by qualified technical and information services staff. The Data Management Capability Enhancement Program ensures the DAP operates on best practice principles in data asset preservation and technical capability. The DAP is developed and managed by CSIRO and employs up to 11 staff in the roles of Project Manager, Business Analyst, Software Developer, Tester, Infrastructure Specialist and Database Administrator. In addition the DAP employs 5 Data Librarians who liaise, provide advice and training to the CSIRO staff who are depositing and approving collections. It is funded within the Information Management & Technology budget.

Evidence:

Science and Industry Research Act 1949 (Cth) (Austl.). Viewed 24 October 2018,
<https://www.legislation.gov.au/Details/C2012C00352>

CSIRO (2017) Annual Report 2016-17. CSIRO, Canberra. Viewed 24 October 2018,
<https://www.csiro.au/en/About/Our-impact/Reporting-our-impact/Annual-reports/16-17-annual-report>.

Data Preservation Principles
https://confluence.csiro.au/display/daphelp/Data+Preservation+Principles

Research Data Services (Internal document)

CSIRO (2017) Information Management and Technology Strategy 2020 overview. CSIRO, Canberra. (Internal document)

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

VI. Expert guidance

R6. The repository adopts mechanism(s) to secure ongoing expert guidance and feedback (either inhouse or external, including scientific guidance, if relevant).

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository
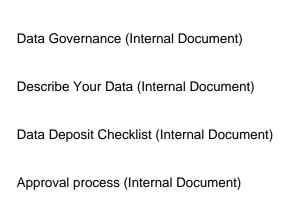
Response:

A range of complex mechanisms for expert guidance and feedback is available in CSIRO for the continued development of the DAP. In-house expertise comprises multi-disciplinary teams including: information management, information technology and scientific staff from a broad range of disciplines. CSIRO staff collaborate with many organisations to ensure the repository continues to meet the needs of the scientific data community. For example, the CSIRO is a partner with the Australian National Data Service who play an important role in the international research data community. CSIRO is active in a number of data sharing initiatives such as National Collaborative Research Infrastructure Strategy and Australian Square Kilometre Array Pathfinder.

DAP staff are annually allocated per capita allowance to develop and maintain their professional knowledge. Staff can use this allowance to attend committee meetings, conferences and undertake training. Repository staff attend the annual Australian Collaborative Conference on Computational and Data Intensive Science. In 2017 and 2018 staff attended the international Open Repositories conference. Staff are members of a range of committees, for example, Standards Australia IT-019 Information and Communications Technology Services - Learning, Education, Training and Research.

All staff who deposit data into the DAP are encouraged to contact the Research Data Service for advice and are referred, if necessary, to expert advisers.

Approvers of DAP collections are CSIRO science leaders who attend compulsory training so that they are aware of their roles and responsibilities. An approver has the option to request one or more advisers to review the collection(s) they are approving for publication.

All data records include a point of contact. This contact information can be used by data users for providing feedback and communicating with the data depositor.

Evidence:

Data Governance (Internal Document)

Describe Your Data (Internal Document)

Data Deposit Checklist (Internal Document)

Approval process (Internal Document)

DAP Approver Training (Internal document)

Data Access Portal User Guide

https://confluence.csiro.au/display/daphelp/Data+Access+Portal+Users+Guide

Research Data Management

http://libguides.csiro.au/ResearchDataManagement

DIGITAL OBJECT MANAGEMENT

VII. Data integrity and authenticity

R7. The repository guarantees the integrity and authenticity of the data.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Response:

The DAP is based on the recommended practice of the CCSDS 650.0–M–2 as published June 2012 and is also known as the Magenta Book.

Deposited data is stored in what we call a collection so that all data within a single deposit is kept together in a single unit. This collection has a cryptographically strong checksum generated at the time of deposit. This can be used at any time to

validate the data as it was at time of deposition in the system. The middleware used to manipulate the collections has been specifically coded to prevent changes to a collection once it has been created.

For ingest activity there are minimum requirements for metadata. Once a collection is published changes to metadata are recorded and the history of a collection is accessible when logged into the DAP. Version control applies to all collections. All PIDS and DOIs are persistent.

Data users access the metadata and data as it was deposited.

Dissemination Information Packages have a checksum and are validated when the data user downloads the file. Checksums are created for each file in the Archival Information Package and checked when copied.

Identity of depositors and approvers is by CSIRO credentials and is traceable.

Evidence:

Consultative Committee for Space Data Systems (2012) Reference Model for an Open Archival Information System (OIAS): Recommended practice, CCSDS 650.0-M-2 Magenta Book, Issue 2, June 2012. CCSDS, Washington. Viewed 24 October 2018, <https://public.ccsds.org/Publications/MagentaBooks.aspx>.

Approval process (Internal Document)

Delegations information (Internal Document)

Data Governance (Internal Document)

DOI Service (Cite My Data)
http://www.ands.org.au/online-services/doi-service

Data Access Portal User Guide
https://confluence.csiro.au/display/daphelp/Data+Access+Portal+Users+Guide

Research Data Management
http://libguides.csiro.au/ResearchDataManagement

National Archives of Australia (2015) Digital Continuity 2020 Policy. NAA, Canberra. Viewed 24 October 2018, <http://www.naa.gov.au/information-management/digital-transition-and-digital-continuity/digital-continuity-2020/index.aspx>.

Reviewer Entry
Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

VIII. Appraisal

R8. The repository accepts data and metadata based on defined criteria to ensure relevance and understandability for data users.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

The Data Collection Development Principles states the DAP will include collections originating from research conducted at CSIRO. The DAP will include external data collections under the following conditions:

• Data identified in a collaboration agreement with an external individual or organisation.

• Data identified in a contract with an external individual or organisation.

• Data that is used in CSIRO's research.

• Data that is the output of CSIRO's research.

• Is aligned with the functions of CSIRO as outlined in section 9 of the Science and Industry Research Act 1949.

• Is aligned with the Minister's Statement of Expectations and CSIROs Statement of Intent.

Self-deposit access to the DAP is available to CSIRO staff. Depositors must acknowledge they have worked through the off system Data Deposit Checklist or Software Licence Selection Process. The checklists guide depositors through key considerations: intellectual property rights, data ownership, ethics, quality, format for reuse, sensitive or personal information and licences. If a depositor requires guidance the checklist refers them to an expert including: legal, business development and contracts, ethics, communications or repository staff.

External data deposits are mediated by the DAP Data Librarians in liaison with CSIRO depositors and approvers who assess the collection according to the checklists.

The data deposit process includes selected mandatory fields and optional extra fields and metadata schema. Mandatory fields must be completed for the collection record to progress to the approval stage.

The data deposit process for externally accessible collections is reviewed by a CSIRO science leader who has completed mandatory training in the role of approver. The approver checks the collection for quality and accuracy. This review includes: metadata, associated outputs such as code, protocols, models and readme files being essential for data users to understand the data. Content that needs amending is flagged by the approver and the depositor is notified and resubmits the necessary information.

The approver can seek guidance from expert advisers including: legal, business development and contracts, ethics, communications or repository staff.

The repository publishes a list of preferred formats and provides links to best practice guides.

It is not mandated but recommended that depositors use a preferred format. On request, a file format can be updated and a new version of the data collection with a new DOI will be generated.

Evidence:

Data Collection Development Principles
https://confluence.csiro.au/display/daphelp/Data+Collection+Development+Principles

Data Preservation Principles
https://confluence.csiro.au/display/daphelp/Data+Preservation+Principles

Data Governance (Internal Document)

Describe Your Data (Internal Document)

Data Deposit Checklist (Internal Document)

Approval process (Internal Document)

DAP Approver Training (Internal document)

Delegations information (Internal Document)

File management recommendations for the DAP (Internal Document)

Research Data Management

http://libguides.csiro.au/ResearchDataManagement

IX. Documented storage procedures

R9. The repository applies documented processes and procedures in managing archival storage of the data.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Response:

Security around the storage of the collections is restricted to infrastructure system administrators and DAP system administrators.

The underlying Historical Storage Management (HSM) and collection management middleware (LCM - Logical Collection Manager) utilised within the DAP meets all requirements as set out in recommended practice CCSDS 650.0–M–2 Magenta Book.

At the collection level, there is geographic diversity of data. However at the application and business logic level there is currently one location where this exists. Backups are performed on critical information in the system and documentation exists to permit the creation of replacement servers which would allow restoration of service within 2-3 days.

Collections in each geographic location are generated from the same source copy and part of the process applied by LCM ensures that the data stored within the collection is a true copy of the source data. If LCM cannot validate this at time of creation an error occurs which alerts an administrator to resolve the problem. Part of this process involves the creation of a cryptographically strong checksum of the stored data as a whole which is validated by LCM on collection mount.

Bad storage media is detected from log entries from the hierarchical storage management (HSM) storage backend. When errors are detected, the media in question is manually flagged as suspect and all data is attempted to be migrated from it to a replacement tape, where all files cannot be recovered from the media, the recovery process requires the validation and copying of the collection from a secondary storage location.

Evidence:

RDS Functional Model (Internal document)

RDS Technical Starter Pack (Internal document)

RDS Storage Architecture (Internal document)

Consultative Committee for Space Data Systems (2012) Reference Model for an Open Archival Information System (OIAS): Recommended practice, CCSDS 650.0-M-2 Magenta Book, Issue 2, June 2012. CCSDS, Washington. Viewed 24 October 2018, <https://public.ccsds.org/Publications/MagentaBooks.aspx>.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

X. Preservation plan

R10. The repository assumes responsibility for long-term preservation and manages this function in a planned and documented way.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

Data management practices in CSIRO comply with recognised standards and legislation where they exist. This includes, for example the use of open standard formats for long term storage of data, standards based metadata schema to describe data, and the use of interoperability protocols such as Open Archives Initiative Protocol for Metadata Harvesting to facilitate data sharing and record keeping requirements. CSIRO's data management practices are in accordance with the Australian Code for Responsible Conduct of Research and align with national initiatives such as the Australian National Data Service.

Evidence:

Data Preservation Principles
https://confluence.csiro.au/display/daphelp/Data+Preservation+Principles

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

XI. Data quality

R11. The repository has appropriate expertise to address technical data and metadata quality and ensures that sufficient information is available for end users to make quality-related evaluations.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

The repository employs five qualified data librarians who have expertise in metadata and data curation. The data librarians provide advice and training to depositors and approvers. Data librarians refer and liaise with experts within the organisation for domain advice about issues of metadata and data quality. Repository staff and domain specific researchers within CSIRO are involved in international and national committees, for example Committee on Data of the International Council for Science, Research Data Alliance, Standards Australia and Open Geospatial Consortium, to ensure the repository is up to date with quality standards on data archiving and data curation.

The DAP has an off system Data Deposit Checklist that requires depositors and approvers to assess the quality and accuracy of the data and the suitability of the format and size for long term sharing, saving and reuse. Depositors are urged to enter meaningful and relevant metadata in all fields. Approvers complete mandatory training and are asked to consider both metadata and data quality as part of the approval process. Approvers can send the collection back to the depositor to address a quality issue requiring the depositor to fix the issue before approval to publish is granted. Depositors and approvers are encouraged to seek advice if they require assistance in determining data and metadata quality.

Within the DAP interface are a range of features to ensure the quality of the metadata. Vocabularies for specific fields decrease input errors: collection type, fields of research codes, contributor and organisation names. The free text fields contain a dictionary to alert the depositor and approver to typing errors. Tooltip help for each field provides guidance about the information required and format. If a field format is added incorrectly, e.g. date, the depositor will be alerted of the invalid entry when saving the collection. The deposit system includes a number of mandatory metadata fields. If the mandatory fields are not complete the data deposit cannot continue to approval for publishing.

Logs are created for all activities within the DAP application. A system administrator is alerted if publishing a collection cannot be completed or if there is an error completing a data user's request.

To assist the data user's understanding of the data, the depositor can add links to related websites, code repositories,

publications (journal articles, reports) and related DAP collections. In the descriptive metadata the depositor is encouraged to add file(s) to further explain the data, such as a readme file or data dictionary.

Currently there is no formal process for obtaining feedback from the designated community, however all data records include a point of contact. This contact information is used by data users to communicate with the data depositor or repository staff.

Evidence:

Data Governance (Internal Document)

Data Access Portal FAQs (Internal Document)

Approval process (Internal Document)

DAP Approver Training (Internal Document)

Describe your data (Internal Document)

Data Preservation Principles
https://confluence.csiro.au/display/daphelp/Data+Preservation+Principles

Research Data Management
http://libguides.csiro.au/ResearchDataManagement

Reviewer Entry
Reviewer 1

Comments:
Accept
Reviewer 2

Comments:
Accept

XII. Workflows

R12. Archiving takes place according to defined workflows from ingest to dissemination.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Response:

The RDS Functional Model is based on the recommended practice of the CCSDS 650.0–M–2 Magenta Book. It contains the workflows for the DAP for deposit, archival storage, data and metadata management, administration, preservation planning and access.

Depositors and data users have access to information in the DAP Users Guide.

Data is secure and users traceable for authenticated access for depositing and accessing restricted collections.

Appraisal and selection of data is undertaken by depositors and approvers.

Data is accepted as deposited. Depositors have access to experts for advice on best practice. Documentation on best practice is available to depositors and approvers in the DAP Users Guide.

Evidence:

Data Access Portal Users Guide
https://confluence.csiro.au/display/daphelp/Data+Access+Portal+Users+Guide

Data Preservation Principles
https://confluence.csiro.au/display/daphelp/Data+Preservation+Principles

Data Governance (Internal Document)

Data Access Portal FAQs (Internal Document)

Approval process (Internal Document)

Describe your data (Internal Document)

RDS Functional Model (Internal Document)

Consultative Committee for Space Data Systems (2012) Reference Model for an Open Archival Information System (OIAS): Recommended practice, CCSDS 650.0-M-2 Magenta Book, Issue 2, June 2012. CCSDS, Washington. Viewed 24 October 2018, <https://public.ccsds.org/Publications/MagentaBooks.aspx>.

XIII. Data discovery and identification

R13. The repository enables users to discover the data and refer to them in a persistent way through proper citation.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Response:

The DAP can be searched in the following ways:

• SEARCH text searches of one word or more

• SEARCH BY LOCATION drawing an Area or selecting a map Point

• BROWSE DATA on fields of research

• DOMAIN SEARCH on particular research area

• WEB SERVICE INTERFACE to search using a remote application

Metadata for depositors:

Compulsory

• Core Scientific Metadata Model CSMD (Standard)

Optional

• Virtual Observatory Resource (Standard)

• Darwin Core (Standard)

• ANZLIC (Standard)

• Marine Community Profile (Standard)

• Sensor (In house)

• Software (In house)

Metadata available as feeds:

• CSMD

• Registry Interchange Format - Collections and Services (RIF-CS)

• Dublin Core

• Darwin Core

• Virtual Observatory Resource

The metadata is ingested into DMS Fedora Repository via DAP for review and approval/publish. DMS Dataset Metadata Collector reads in approved/published metadata and generates corresponding RIF-CS metadata and adds (or updates) it to the DMS Fedora Repository. RIF-CS metadata gets harvested to Research Data Australia (RDA) from the DMS Fedora Repository via the Open Archives Initiative Protocol for Metadata Harvesting protocol.

Publicly accessible DAP records are syndicated to the national Research Data Australia repository. DAP depositors can choose to have records tagged for harvesting to national repositories: Atlas of Living Australia, TERN Data Discovery Portal: Australian Coastal Ecosystems Facility and TERN Data Discovery Portal: Soil and Landscape Grid of Australia or the regional repository Pacific-Australia Climate Change Science Adaptation Planning program.

DAP is registered with re3data.org, DataCite, PLOS preferred repository list
https://journals.plos.org/plosone/s/data-availability#loc-unstructured-and-or-large-data

DAP records are included in the Clarivate Data Citation Index - Web of Science
http://wokinfo.com/cgi-bin/dci/search.cgi

The system requires depositors to build the attribution statement for the data collection.

DOIs are minted for publicly accessible data collections to provide internationally unique identifiers. DOIs are also

generated for data collections where the metadata is publicly accessible, but the data is not.

CSIRO endorses the Force11 Manifesto and the FAIR Data Principles.

Evidence:

Data Access Portal
https://data.csiro.au/

Data Access Portal Users Guide
https://confluence.csiro.au/display/daphelp/Data+Access+Portal+Users+Guide

Find Data
https://confluence.csiro.au/display/daphelp/Find+Data

Create Your Citation (Internal document)

Reviewer Entry
Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

XIV. Data reuse

R14. The repository enables reuse of the data over time, ensuring that appropriate metadata are available to support the understanding and use of the data.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry
Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

Metadata in use:

For depositors

Compulsory

• Core Scientific Metadata Model CSMD (Standard)

Optional

• Virtual Observatory Resource (Standard)

• Darwin Core (Standard)

• ANZLIC (Standard)

• Marine Community Profile (Standard)

• Sensor (In house)

• Software (In house)

Metadata available as feeds:

• CSMD

• RIF-CS

• Dublin Core

• Darwin Core

• Virtual Observatory Resource

The basic required fields are:

• Collection title

• Collection description

• Keywords

• Field of research

• Licence

• Metadata schema specific required fields

The repository does publish a list of preferred formats and provides links to best practice guides.

Measures to account for the possible evolution of formats is in development. Staff watch international repository trends to inform development.

There are ongoing discussions with the Designated Community to determine best practice for formats.

There is an ongoing process in DAP evolution on an as needed basis.

The DAP ensures the data can be understood for reuse as collections include appropriate metadata: description, lineage and readme files.

Evidence:

Data Governance (Internal Document)

File management recommendations for the DAP (Internal Document)

Data Deposit Checklist (Internal Document)

Contact for help
https://data.csiro.au/dap/contact

Access or download data
https://confluence.csiro.au/display/daphelp/Access+or+Download+Data

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

TECHNOLOGY

XV. Technical infrastructure

R15. The repository functions on well-supported operating systems and other core infrastructural software and is using hardware and software technologies appropriate to the services it provides to its Designated Community.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

The repository functions on well supported operating systems and software described in the RDS Technical Starter Pack. Information Management & Technology teams review the infrastructure to ensure it continues to meets the needs of the repository. The technical starter pack includes documentation to maintain the system.

Software Development Patterns

The repository code follows several well documented and validated patterns of software development which has delivered robust and scalable performance.

Some of these patterns are:

Data Capture and Storage Standards:
• Dependency injection/ or IoC (inversion of control) – Is the main principle behind decoupling process of Spring.
• Factory – Spring uses factory pattern to create objects of beans using Application Context reference.
• Proxy – Used heavily in Aspect Oriented Programming, and remoting.

For more information
https://premaseem.wordpress.com/2013/02/09/spring-design-patterns-used-in-java-spring-framework/

Metadata for depositors:

Compulsory
• Core Scientific Metadata Model CSMD (Standard)

Optional
• Virtual Observatory Resource (Standard)
• Darwin Core (Standard)
• ANZLIC (Standard)

- Marine Community Profile (Standard)
- Sensor (In house)
- Software (In house)

Metadata available as feeds:
- CSMD
- RIF-CS
- Dublin Core
- Darwin Core
- Virtual Observatory Resource

The metadata is used to generate the Registry Interchange Format - Collections and Services (RIF-CS) schema which is a well-known schema for metadata exchange.

The repository uses a number of open-source libraries that are acknowledged on the DAP's website.

We maintain a record of the software configuration in the code and the build is automated using Gradle. Builds occur on our CI machine to ensure independence from a developer's local machine configuration. A version of the code for the DAP was released as open source in 2016.

Release history of the DAP is available on the website.

Provision for real-time to near real-time data streams have been discussed but have not been developed.

Data management practices in CSIRO comply with recognised standards and legislation where they exist. This includes, for example the use of open standard formats for long term storage of data, standards based metadata schema to describe data, and the use of interoperability protocols such as Open Archives Initiative Protocol for Metadata Harvesting to facilitate data sharing and record keeping requirements. CSIRO's data management practices are in accordance with the Australian Code for Responsible Conduct of Research and align with national initiatives such as the Australian National Data Service.

Evidence:

RDS Technical Starter Pack (Internal Document)

Acknowledgements: Open source and 3rd party components in the Data Access Portal
https://data.csiro.au/dap/acknowledgements

Code for CSIRO Data Access Portal

https://sourceforge.net/projects/csiro-dap/

Release history

https://confluence.csiro.au/display/daphelp/Release+History

National Health and Medical Research Council (2007) Australian Code for the Responsible Conduct of Research.
NHMRC, Canberra. Viewed 24 October 2018,
<https://nhmrc.gov.au/about-us/publications/australian-code-responsible-conduct-research-2007>.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

XVI. Security

R16. The technical infrastructure of the repository provides for protection of the facility and its data, products, services, and users.

Compliance Level:

3 – The repository is in the implementation phase

Reviewer Entry

Reviewer 1

Comments:
3 – The repository is in the implementation phase

Reviewer 2

Comments:
3 – The repository is in the implementation phase

Response:

All systems in the DAP are covered by the CSIRO IMT Business Continuity Plan.

Validation of the Disaster Recovery Plan for the DAP is planned for 2018.

Security evaluations are done on the boxes to the pre-productions systems, prior to the release to production. A report on the environment comes back and we work towards resolving any issues that are highlighted in the security report.

Security around the storage of the collections is restricted to infrastructure system administrators and DAP system administrators. All administrators have a minimum level of security clearance (managed by Commonwealth Government).

Evidence:

CSIRO Information Security Procedure (Internal Document)

IMT Business Continuity Plan (Internal Document)

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

APPLICANT FEEDBACK

Comments/feedback

These requirements are not seen as final, and we value your input to improve the core certification procedure. To this end, please leave any comments you wish to make on both the quality of the Catalogue and its relevance to your organization, as well as any other related thoughts.

Response:

There have been immediate benefits in undertaking this self-assessment for certification. A benefit is documentation has been organised and we have been able to develop quick applications for journal publishers and funding agencies to recognise the DAP on their repository recommended list.

As an institutional repository we question the requirement to make internal documents available to the public. Some of the

new organisational requirements in the certification require evidence that is sensitive in nature or is based on procedures of our parent organisation that are restricted to internal use.

There is a lack of guidance on what minimum requirements are needed for certification.

The procedure and timeframes for the processing of an application is unclear.

In early December 2017, we contacted the Core Trust Seal via the website and still had no response several weeks later.

There is an element of repetitiveness in the requirements. For example we found that we were discussing similar issues and pointing to the same evidence throughout.

It was useful to be able to see other repositories certification applications on the Data Seal of Approval website.

Currently, for the Data Seal of Approval, when a repository adds the seal to their repository page when you click on the seal it leads the user to their certification application. See the example for PARADESIC http://www.paradisec.org.au/. A suggestion is to link to a summary of the certification requirements for the Core Seal Trust.

Reviewer Entry
Reviewer 1

Comments:
Accept
Reviewer 2

Comments:
Accept