Version 1.0
13 February 2023

# GRDC RD&E DATA CAPTURE AND STORAGE – GUIDELINES FOR RESEARCH PARTNERS

## Table of contents


BACKGROUND ..... 3

SECTION A: OVERVIEW ..... 3

A-1. Identifying RD&E Data for collection and storage ..... 3
A-2. Uploading RD&E Data to an Approved Repository ..... 4
A-3. Registering Project Metadata with the GRDC Data Catalogue ..... 5
A-4. Checklist for GRDC Research Partners ..... 6

SECTION B ..... 7

B-1. GRDC RD&E Data Management Plan Guideline for Research Partners v1.0 ..... 7

B-1.1. Investment Details Worksheet ..... 7

B-1.1.1. GRDC Contract Code ..... 8
B-1.1.2. Contract title ..... 8
B-1.1.3. Prepared by ..... 8
B-1.1.4. Date Prepared ..... 8
B-1.1.5. Principal Investigator ..... 8
B-1.1.6. Responsible party ..... 8

B-1.2. Project Data Management ..... 8

B-1.2.1. RD&E Data organisation ..... 8
B-1.2.2. Ethical, confidentiality or privacy considerations ..... 9
B-1.2.3. Access and security ..... 9
B-1.2.4. Backup ..... 9
B-1.2.5. GRDC background data ..... 9

B-1.3. Data storage ..... 9

B-1.3.1. RD&E Dataset Name ..... 10
B-1.3.2. Storage location ..... 10
B-1.3.3. Milestone No. ..... 10
B-1.3.4. RD&E Data format ..... 10
B-1.3.5. Project Metadata standard/s ..... 11
B-1.3.6 Access conditions ..... 11

B-2. GRDC RD&E Data Management Plan Template v1.0 ..... 13



**P** PO Box 5367 Kingston, ACT 2604 Australia
**T** +61 2 6166 4500 **F** +61 2 6166 4599 **E** grdc@grdc.com.au

SECTION C ..... 14
C-1. GRDC List of Approved Repositories for Research Partners v1.0 ..... 14
C-1.2. Institutional Data Repositories ..... 14
C-1.3. Domain-Specific Data Repositories ..... 14
C-1.4. General Research Repositories ..... 14
C-2. GRDC Guideline for Identifying Approved Repositories for Research Partners v1.0 ..... 15
C-2.1. Table: Criteria for Approved Repositories ..... 15

SECTION D ..... 16
D-1. GRDC Metadata Collection Form v1.0 ..... 16
D-2. GRDC Guideline for Research Partners for Registration of Metadata with the GRDC Data Catalogue v1.0 ..... 21
D-2.1. Automated Harvesting of Project Metadata from an Approved Repository ..... 21
D-2.2. Manual registration of Project Metadata downloaded from an Approved Repository 22
D-2.3. Manual addition of Project Metadata using the GRDC Data Catalogue web-based form ..... 22
D-3. GRDC Guideline for Research Partners for Describing GRDC RD&E Datasets v1.0 ..... 22
D-3.1. Sharing/access information ..... 22
D-3.2. Data & file overview ..... 22
D-3.3. Methodological information ..... 23
D-3.4. Environmental/experimental conditions ..... 23
D-3.5. Data-specific information ..... 23

# BACKGROUND

The GRDC RD&E Data Capture and Storage – Guidelines for Research Partners relates to the following three actions necessary to collect and store RD&E Data:

1. Identifying RD&E Data for collection and storage.
2. Uploading RD&E Data to an Approved Repository.
3. Registering Project Metadata with the GRDC Data Catalogue.

# SECTION A: OVERVIEW

## A-1. Identifying RD&E Data for collection and storage

Valuable RD&E Datasets to be created in a GRDC Investment will be identified and documented in the Data Management Plan for the investment.

RD&E Datasets to be created in a GRDC Investment should be identified during contract negotiation or as early as possible in the investment. The relationship between the RD&E Datasets listed in the Data Management Plan and entries in the Project IP and Project Outputs (IPPO) Register will be clearly recorded.

**Relevant Guideline – Section B**

**B-1. GRDC RD&E Data Management Plan Guideline for Research Partners v1.0**

This guideline assists the GRDC Partner to complete the Data Management Plan (DMP) with a field-by-field guide to the required information.

**Relevant Template – Section B**

**B-2. GRDC RD&E Data Management Plan Template v1.0**

The DMP template is provided for the convenience of the GRDC Partner to help meet the requirements detailed in Schedule 3 of the latest version of the GRDC Research Contract.

## A-2. Uploading RD&E Data to an Approved Repository

RD&E Data created in GRDC Investments will be uploaded to an agreed Approved Repository.

The GRDC Partner will nominate and seek agreement from their GRDC Investment Manager on the Approved Repository where RD&E Data created in GRDC Investments will be stored. The agreed Approved Repository will be recorded on the Data Management Plan.

An Approved Repository list will be provided to GRDC Partners. There is an internal GRDC process for approving the use of a Repository that does not appear on this list.

The GRDC Partner is responsible for uploading RD&E Data created in a GRDC Investment to the Approved Repository.

The process for uploading RD&E Data to the Approved Repository will be specific to the Approved Repository. The GRDC Partner should consult the manager of the Approved Repository for further guidance on this if required.

**Relevant Guidelines – Section C**

**C-1. GRDC List of Approved Repositories for Research Partners v1.0**

This document provides a list of Approved Repositories that fulfil the criteria defined in the GRDC Guidelines for Identifying Approved Repositories. As additional Approved Repositories are identified, the list will be updated.

**C-2. GRDC Guideline for Identifying Approved Repositories for Research Partners v1.0**

This document provides a list of criteria by which Approved Repositories are defined.

## A-3. Registering Project Metadata with the GRDC Data Catalogue

For each RD&E data set that is uploaded to an Approved Repository, Project Metadata must be created and registered with the GRDC Data Catalogue.

Project Metadata must be created and submitted to the GRDC Data Catalogue to ensure that interested parties can:

- find RD&E Data that is relevant
- use RD&E Data according to agreed terms of access
- preserve and re-use RD&E Data.

The GRDC Partner is responsible for ensuring the quality of the Project Metadata and it will be reviewed by the GRDC Investment Manager.

Guidance on how to provide a suitable description of the RD&E data set using Project Metadata is provided in the GRDC Guideline for Research Partners for Describing GRDC RD&E Datasets v1.0.

Guidance on how to register Project Metadata with the GRDC Data Catalogue can be found in the GRDC Guideline for Research Partners for Registration of Metadata with the GRDC Data Catalogue v1.0.

**Relevant Guidelines – Section D**

**D-1. GRDC Metadata Collection Form v1.0**
The Metadata Collection Form has 21 fields, some of the fields are compulsory as indicated by an asterisk.

**D-2. GRDC Guideline for Research Partners for Registration of Metadata with the GRDC Data Catalogue v1.0**
This document details the three different methods of uploading metadata to the GRDC Data Catalogue. Harvesting metadata directly from Approved Repositories is the preferred method. However, when this is not possible, two other methods for capturing metadata are also detailed in this document.

**D-3. GRDC Guideline for Research Partners for Describing GRDC RD&E Datasets v1.0**
The metadata that is recorded in the GRDC Data Catalogue has fields that are general in nature. That is, they enable the full diversity of GRDC RD&E Data to be accommodated, but they do not specify the exact information that is required about given experiments, processes, or conditions around the gathering of the dataset. This guideline helps the GRDC Partner to include necessary clarifying information into the free text ‘description’ metadata field, that is particular to the dataset being stored.

## A-4. Checklist for GRDC Research Partners

- Identify RD&E Datasets to be created in each investment and record them in a Data Management Plan
- Upload RD&E Datasets to the agreed Approved Repository according to the Data Management Plan
- Ensure quality of the uploaded RD&E Datasets
- Ensure quality of the Project Metadata created
- Ensure that Project Metadata has been registered in the GRDC Data Catalogue.

# SECTION B

## B-1. GRDC RD&E Data Management Plan Guideline for Research Partners v1.0

For every GRDC Investment that generates RD&E Data, a GRDC Data Management Plan will be filled out that describes how RD&E Data will be managed within the investment and how newly created RD&E Datasets will be stored and catalogued for future use.

Under the GRDC Research Contract, it is the responsibility of the GDRC Partner to complete the Data Management Plan. GRDC recommends that the party responsible for RD&E Data management within the investment contact their organisation's Data Steward if they have one. Data Stewards can provide information on the resources available to the GRDC Partner, including shared workspaces and institutional data repositories.

The Data Management Plan will ideally be completed during the contract negotiation phase of the investment management cycle. It will be updated throughout the course of the contracted work to keep up to date with any changes.

Download the Excel template for the Data Management Plan.

In creating a Data Management Plan for a GRDC RD&E Investment, please use the provided Data Management Plan template. It is necessary that the Data Management Plan be submitted as a Microsoft Excel spreadsheet. It is comprised of three worksheets with the following labels:

- Investment details
- Project data management
- Data storage.

In section B-2, guidance for completing each of the worksheets is provided.

A Data Management Plan template is provided as a Microsoft Excel file. It is required that GRDC Partners create the Data Management Plan using this template. Please do not enter the information in the Research Contract Schedule.

### B-1.1. Investment Details Worksheet

This worksheet is for capturing information about the investments for which the Data Management Plan is being created. It is included to help link the Data Management Plan back to investment details found in GRDC's Investment Management System (IMS) and other GRDC systems, and to provide details about parties responsible for RD&E Data management at the GRDC Partner organisation.

This section provides details on how to fill out each of the fields that appear on this worksheet.

#### B-1.1.1. GRDC Contract Code

This field contains the GRDC Contract Code as per Research Contract.

#### B-1.1.2. Contract title

This matches the Research Contract Title.

#### B-1.1.3. Prepared by

The party that has prepared this Data Management Plan, including their organisation, position, phone number.

#### B-1.1.4. Date Prepared

The date that the first complete version of this Data Management Plan was completed by the party named in the 'Prepared by' field.

#### B-1.1.5. Principal Investigator

The Principal Investigator for the GRDC investment for which this Data Management Plan has been prepared.

#### B-1.1.6. Responsible party

The person employed by the GRDC Partner that is responsible for ensuring the Data Management Plan is followed and updated where necessary. It is important to clarify who has responsibility for RD&E Data management for the investment.

### B-1.2. Project Data Management

This worksheet contains information on how RD&E Data will be managed by the GRDC Partner throughout the life of the investment and encourages them to consider this in advance.

A description of how to fill out each field in this worksheet is provided below.

#### B-1.2.1. RD&E Data organisation

It is important to think through in advance the practicalities of how RD&E Data will be managed within the investment.

Plans to ensure synchronisation of RD&E Data files and documents between different parties will avoid unintentionally creating multiple versions of the RD&E Dataset or document. This can be achieved by using a shared

working space. Many GRDC Partners offer shared working spaces to researchers for the purpose of collaborative research. To avoid RD&E Data loss, many shared working spaces also provide file versioning.

It is also helpful to think in advance how collaborators in the investment will name their RD&E Data files and documents and how the RD&E Data will be organised in folders.

#### B-1.2.2. Ethical, confidentiality or privacy considerations

Here the GRDC Partner will list any sensitivities associated with the RD&E Data being used in the investment. Any ethical and privacy requirements can be met through access control and data security within the investment. More information on sensitive data is found on the ARDC website.

#### B-1.2.3. Access and security

Personnel who will have access to the RD&E Data in the investment will be listed here. Measures taken to protect sensitive RD&E Data from unauthorised access, or to alter the RD&E Data to remove sensitive information, should be provided here.

#### B-1.2.4. Backup

RD&E Data loss can represent a significant cost to an investment in terms of time and resources. This field clarifies how RD&E Data and documents in the investment will be backed up to insure against loss. Many shared platforms provided by GRDC Partners will automatically provide RD&E Data backup on a regular basis. If such an environment is not being used, a strategy for performing regular backups will need to be defined.

#### B-1.2.5. GRDC background data

As GRDC improves its approach to Data Management, there will be historical RD&E Datasets that have not been registered in the GRDC Data Catalogue or stored in an Approved Repository. This is a chance to identify such RD&E Datasets and to arrange for them to be managed so they can be found through the GRDC Data Catalogue and used in future investments.

### B-1.3. Data storage

This worksheet is focused on archiving the RD&E Datasets that are created throughout the investment. Each row entry corresponds to an RD&E Dataset created within the investment. Using this worksheet, these RD&E Datasets can be identified in advance and a plan made for their cataloguing and storage in an Approved Repository.

### B-1.3.1. RD&E Dataset Name

This is a descriptive name of the RD&E Dataset. When deciding on a name, think of what ought to be included to properly characterise the RD&E Dataset and make it easily findable via a Google-style search query. The name should also match one of the Project Outputs listed in the Project IP and Project Outputs Register (IPPO Register). This will ensure that any limits on commercialisation and dissemination of this RD&E Dataset are well understood by referring to its entry in the IPPO Register.

### B-1.3.2. Storage location

The storage location identified here will be an Approved Repository that has been agreed upon by both the GRDC Investment Manager and the GRDC Partner. For minimum criteria in identifying an Approved Repository, please see the GRDC Guidelines for Identifying Approved Repositories for Research Partners.

### B-1.3.3. Milestone No.

RD&E Data storage and cataloguing within an investment is tied to Research Contract Milestones. Each row in the table corresponds to a RD&E Dataset that is expected to be generated through activities described in a contract Milestone. The 'Milestone No.' field records at which contract Milestone the listed RD&E Dataset will be created and helps to plan for when RD&E Data storage and cataloguing will take place throughout the investment.

In the figure below, RD&E Data is collected at contract Milestone 103 (left). Text is added to the contract Milestone to indicate a RD&E Data storage Milestone has been identified and recorded in the Data Management Plan. The Data Management Plan record (right) shows the Milestone at which we expect the RD&E Dataset to be created.

| | | | |
|---|---|---|---|
| | with varying levels of disease inoculum in the following five locations:<br>• Location 1<br>• Location 2<br>• Location 3<br>• Location 4<br>• Location 4 | | |
| 103 | Harvest all successful field trials<br>Collate yield information into single data file that includes information inoculum level.<br>Update IPPO register<br>Complete Data Management Milestone. | 30/09/2022 | $155,590.00 |

| Dataset Name | Milestone No. | Description | Data form |
|---|---|---|---|
| *Enter a short descriptive name of the dataset to be managed* | *Enter the corresponding contract milestone for this data management milestone.* | *Describe the dataset* | *See list be for some s data forma try to avoi proprietary standards* |
| Trial results 2022 | 103 | Results of field trials across all five sites as described in the relevant milestone | CSV |

### B-1.3.4. RD&E Data format

The format of the RD&E Dataset is an important factor in its reusability. This field encourages planning of the RD&E Data format when writing the Data Management Plan. To maximise opportunities for future re-use, the chosen

RD&E Data format should be non-proprietary, widely adopted, and machine-readable.

The following table provides a list of suggested RD&E Data formats for several different Data types.

| Dataset type | Suggested format |
|---|---|
| Document | HTML, TXT, XML, RTF, PDF, MS-Word |
| General numerical data | CSV, NetCDF, MS-Excel |
| Spatial data | Esri shapefile or geodatabase, GeoTIFF |
| Database | SQLite, MS-Access |
| Image | EPS, SVG (for vector images), PNG, TIFF (for images with sharp edges), JPEG (for photographs) |
| Audio | WAV or formats with lossless compression where possible |

### B-1.3.5. Project Metadata standard/s

The Project Metadata that is managed by the GRDC Data Catalogue is like the type of data that you will find in a library catalogue. It can be used to refer to any type of RD&E Dataset and provides high-level information for the purposes of RD&E Dataset findability. When harvesting Project Metadata from an Approved Repository, the GRDC Data Catalogue can import Project Metadata of the RIF-CS or DataCite types. It is therefore important to ensure that the Approved Repository chosen to store the RD&E Data can export Project Metadata in one of these formats.

The GRDC Investment Manager or a Data Steward at the host institute will be able to help identify whether the Project Metadata in an Approved Repository can be harvested by the GRDC Data Catalogue.

When creating Project Metadata, it must have sufficient information to enable reuse of the RD&E Dataset it describes. The Project Metadata will allow for a free text field to be filled in with information about the RD&E Dataset. For guidance on information that ought to be included in this field, see the GRDC Guideline for Research Partners for Describing RD&E Datasets.

### B-1.3.6 Access conditions

This defines to what extent the RD&E Data will be accessible and downloadable by third parties. The access conditions will be consistent with the Research Contract or as negotiated between GRDC and the GRDC Partner as provided by the Research Contract.

The following options can be written in this field.

- **Open**: Project Metadata are findable and RD&E Data are accessible with any Creative Commons Licence.
- **Closed**: Project Metadata are findable and RD&E Data are not available for use by third parties.
- **Restricted**: Project Metadata are findable and data are only accessible by third parties with approval from the Data Custodian and the relevant GRDC Investment Manager. This is the default for GRDC RD&E Datasets.
- **Embargoed**: Project Metadata are findable and RD&E Data are closed until the end of a specified embargo period to allow for publication of research or commercialisation activities. After the embargo period, the RD&E Data will adopt the default 'Restricted' status unless otherwise indicated.

# B-2. GRDC RD&E Data Management Plan Template v1.0

The RD&E Data Management Plan Template has three worksheets, with each field in each worksheet described here.

Worksheet 1. Investment details

| | A | B |
|---|---|---|
| 1 | | |
| 2 | **GRDC contract code** | |
| 3 | **Contract title** | |
| 4 | **Prepared by** | |
| 5 | **Date prepared** | |
| 6 | **Principal Investigator** | |
| 7 | **Responsible party** | |
| 8 | | |
| 9 | | |

**Investment details** | Project data management | Data storage

Worksheet 2. Project data management details

| | A | B |
|---|---|---|
| 1 | | |
| 2 | **Data organisation** | |
| 3 | **Ethical, confidentiality or privacy considerations** | |
| 4 | **Access and security** | |
| 5 | **Backup** | |
| 6 | **GRDC Background Data** | |
| 7 | | |
| 8 | | |

Investment details | **Project data management** | Data storage

Worksheet 3. Data storage details

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| 1 | | | | | | |
| 2 | **Dataset Name** | **Storage location** | **Milestone No.** | **Data format** | **Metadata standard/s** | **Access permissions** |
| 3 | *Enter a short descriptive name of the dataset to be managed. This should match the corresponding record in the GRDC IPPO Register* | *Identify the Approved Repository where the data will be stored.* | *Enter the associated contract milestone for the deposition of the dataset in the storage location.* | *Enter the data format for the dataset. See list below this table for suggested data formats. Proprietary data standards should be avoided.* | *Describe the metadata standards that have been used to describe the dataset.* | *Open/closed/embargoed/restricted. Please provide as much detail as possible.* |
| 4 | | | | | | |
| 5 | | | | | | |
| 6 | | | | | | |
| 7 | | | | | | |
| 8 | | | | | | |
| 9 | **Suggested Data formats** | | | | | |
| 10 | | | | | | |
| 11 | Documents - HTML, TXT, XML, RTF, PDF, MS-Word | | | | | |
| 12 | General observational and measurement data - CSV, NetCDF, MS-Excel | | | | | |
| 13 | Spatial data - Esri shapefile or geodatabase, GeoTIFF | | | | | |
| 14 | Databases - SQLite, MS-Access | | | | | |
| 15 | Images - EPS, SVG (for vector images), PNG, TIFF (for images with sharp edges), JPEG (for photographs) | | | | | |
| 16 | Audio - WAV or formats with lossless compression where possible | | | | | |
| 17 | | | | | | |

Investment details | Project data management | **Data storage**

# SECTION C

## C-1. GRDC List of Approved Repositories for Research Partners v1.0

GRDC has reviewed and approved a number of repositories for RD&E Data in accordance with the GRDC Guideline for Identifying Approved Repositories for Research Partners v1.0. The GRDC Approved Repository list is shown here. It is broken down between Institutional Data Repositories, Domain-Specific Data Repositories, and General Research Data Repositories. Additional repositories may be assessed and approved by GRDC as needed.

### C-1.2. Institutional Data Repositories

| Name | URL |
|---|---|
| University of Sydney eScholarship Repository | https://ses.library.usyd.edu.au/ |
| UQ eSpace | https://espace.library.uq.edu.au/ |
| CSIRO Data Access Portal | https://data.csiro.au/ |
| UWA Profiles and Research Repository | https://research-repository.uwa.edu.au/ |
| University of Adelaide Figshare | https://adelaide.figshare.com/ |
| Curtin Research Data Collection | https://researchdata.edu.au |
| University of Melbourne Figshare | https://dataservices.research.unimelb.edu.au/services/43/ |

### C-1.3. Domain-Specific Data Repositories

| Name | Data Domain | URL |
|---|---|---|
| Atlas of Living Australia | Biodiversity | https://www.ala.org.au/ |
| Bioplatforms Australia | ‘Omics | https://data.bioplatforms.com/ |
| Australian Data Archive | Social sciences | https://ada.edu.au/ |
| GenBank | DNA Sequencing | https://www.ncbi.nlm.nih.gov/genbank/<br>view FAIRsharing entry |
| NCBI Sequence Read Archive (SRA) | DNA Sequencing | https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?<br>view FAIRsharing entry |

### C-1.4. General Research Repositories

| Name | URL |
|---|---|
| Zenodo | https://www.zenodo.org |
| Data WA | https://data.wa.gov.au |

# C-2. GRDC Guideline for Identifying Approved Repositories for Research Partners v1.0

This guideline lists the criteria GRDC has used for certifying Approved Repositories for the purpose of storing RD&E Data that are generated within GRDC investments. To be a GRDC Approved Repository, the repository must meet the requirements listed in C-2.1:

## C-2.1. Table: Criteria for Approved Repositories

| Requirement | Explanatory Notes |
|---|---|
| Fit-for-purpose | The purpose of the Approved Repository is to facilitate long-term storage for RD&E Data and to provide access to it. |
| Discovery and Identification | The Approved Repository:<br>• provides a publicly facing web-based search tool for data discovery;<br>• assigns a persistent identifier to each RD&E Dataset and provides it as Uniform Resource Locator (URL) as part of the associated Project Metadata;<br>• provides application programming interface (API) endpoints for metadata harvesting that return metadata according to the RIF-CS metadata standard |
| Security | The technical infrastructure of the Approved Repository provides for protection of the facility and its RD&E Data, products, services, and users |
| Technical infrastructure | The Approved Repository must operate on a reliable and stable core infrastructure that can support the expected usage load and storage requirements. |
| Access management, licensing, and rights | RD&E Data is findable to a third-party user but access by third parties can be blocked or restricted (gated). Access to RD&E Data can be granted by the Data Custodian. Project Metadata fields associated with the dataset include information on licensing and rights for datasets use and the repository monitors compliance. |
| OAI-PMH Implementation | The Approved Repository implements the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). An OAI-PMH implementation is a feature of an Approved Repository that allows the GRDC Data Catalogue to automatically harvest Project Metadata from the Approved Repository. For details on Project Metadata harvesting, please refer to the *GRDC Guideline for Research Partners for Registration of Metadata with the GRDC Data Catalogue v1.0*. |
| Continuity of service | The Approved Repository has a plan to ensure ongoing access to and preservation of its RD&E Data and Project Metadata, including regular backups. Disaster recovery and succession plans are in place |
| Guidance and assistance | Guidance on the use of the Approved Repository must be available in the form of both documentation for end-users, and human assistance. |

# SECTION D

## D-1. GRDC Metadata Collection Form v1.0

**1. What is the name of the collection or dataset? ***

**2. What is this metadata record for? *Select one of the following.* ***

☐ Collection ☐ Dataset

**3. Which organisation is contributing this metadata record? ***

**4. Which GRDC project does this collection or dataset relate to? ***

| | |
|---|---|
| GRDC contract code: | |
| GRDC project title: | |

**5. What is the DOI for this collection or dataset?**

**6. What is the Handle for this collection or dataset?**

**7. What is the URI for the metadata or a landing page which allows users to know how to access the collection or dataset? ***

**8. What is the URI that triggers a direct data download?**

**9. When was the data collected or the observations made?** ***Use the YYYY format.*** *****

| | |
|---|---|
| Start date: | |
| End date: | |

**10. When was the collection or dataset published?** ***Use the YYYY format.*** *****

| |
|---|
| |

**11. What is the spatial coverage of the collection or dataset? ***

| | |
|---|---|
| Spatial coverage: | |
| Spatial coverage type: *Select one of the following.* | ☐ DCMI Point ☐ GML KML coordinates<br>☐ GPX ☐ ISO 3166-1<br>☐ ISO 3166-2 ☐ ISO 3166-3<br>☐ DCMI Box ☐ KML coordinates<br>☐ Other |

**12. Provide a brief description of the collection or dataset. ***

| |
|---|
| |

**13. Provide a full description of the collection or dataset.**

| |
|---|
| |

**14. For the citation, provide the names of those involved in producing the collection or dataset?** ***Repeat this table as needed.***

| | |
|---|---|
| Given name: | |
| Family name: | |

**15. What is the contact email address for the collection or dataset?**

| |
|---|
| |

**16. What are the subject terms for the collection or dataset?** ***Repeat this table as needed.***

| | |
|---|---|
| Subject term: | |
| Subject term type: *Select one of the following.* | ☐ ANZSRC: Field of Research codes<br>☐ ANZSRC: Type of Activity<br>☐ ANZSRC: Socio-economic Objective codes<br>☐ Australian Pictorial Thesaurus<br>☐ Global Change Master Directory Keywords<br>☐ International Standard for language codes<br>☐ Library of Congress Subject Headings<br>☐ Powerhouse Museum Object Name Thesaurus<br>☐ Schools Online Thesaurus<br>☐ Thesaurus of Psychological Index Terms<br>☐ Local |

**17. What is the rights statement for the collection or dataset?**

| |
|---|
| |

**18. What licence applies to the collection or dataset?** *

| | | |
|---|---|---|
| Licence description: | | |
| Licence type: *Select one of the following.* | ☐ CC-BY<br>☐ CC-BY-ND<br>☐ CC-BY-NC-SA<br>☐ CC0<br>☐ AusGoal Restrictive | ☐ CC-BY-SA<br>☐ CC-BY-NC<br>☐ CC-BY-NC-ND<br>☐ GPL<br>☐ No Licence |

**19. What are the access rights for the collection or dataset?** *

| | | |
|---|---|---|
| Access rights: | | |
| Access rights type: *Select one of the following.* | ☐ Open<br>☐ Restricted | ☐ Conditional |

**20. Provide related information for the collection or dataset. *Repeat this table as needed.***

| | |
|---|---|
| Title: | |
| Identifier: | |
| Identifier type: | ☐ Australian Research Council (ARC) identifier<br>☐ Archival Resource Key (ARK)<br>☐ National Library of Australia party identifier<br>☐ Digital Object Identifier (DOI)<br>☐ International Article Number<br>☐ Crossref Funder Registry<br>☐ Global Research Identifier Database<br>☐ Handle System Identifier<br>☐ International Standard Book Number<br>☐ ORCID Identifier<br>☐ PubMed ID<br>☐ Persistent Uniform Resource Locator (PURL)<br>☐ Research Activity Identifier<br>☐ Thomson Reuters ResearcherID<br>☐ Scopus Author ID<br>☐ Uniform Resource Identifier (URI)<br>☐ Uniform Resource Name (URN)<br>☐ Other |
| Related information type: | ☐ Activity ☐ Collection<br>☐ Website ☐ Metadata<br>☐ Party ☐ Provenance<br>☐ Reuse information ☐ Service<br>☐ Data quality information |
| Relationship: | *Collection-Collection relation:*<br>☐ Describes ☐ Has part<br>☐ Has association with ☐ Is described by<br>☐ Is part of ☐ Is derived from<br>☐ Has derived collection ☐ Has version<br>☐ Has association with ☐ Is version of<br><br>*Collection-Party relation:*<br>☐ Is managed by ☐ Has collector<br>☐ Has principal investigator ☐ Is enriched by<br>☐ Has association with ☐ Is owned by<br><br>*Collection-Activity relation:*<br>☐ Has association with ☐ Is output of<br><br>*Collection-Service relation:*<br>☐ Is available through ☐ Supports<br>☐ Is operated on by ☐ Is presented by<br>☐ Has value added by ☐ Is produced by<br>☐ Has association with |

**21. Provide related publications for the collection or dataset.** ***Repeat this table as needed.***

| | |
|---|---|
| Title: | |
| Identifier: | |
| Identifier type: | ☐ Australian Business Number<br>☐ Australian Research Council (ARC) identifier<br>☐ Crossref Funder Registry<br>☐ Digital Object Identifier (DOI)<br>☐ Global Research Identifier Database<br>☐ International Geo Sample Number<br>☐ 'info' URI scheme<br>☐ MIME type<br>☐ ORCID Identifier<br>☐ Research Activity Identifier<br>☐ Thomson Reuters ResearcherID<br>☐ Scopus Author ID<br>☐ Universal Product Code |
| Relationship: | ☐ Is cited by ☐ Is referenced by<br>☐ Is documented by ☐ Is supplemented by<br>☐ Is supplement to ☐ Is reviewed by<br>☐ Is supported by |

## D-2. GRDC Guideline for Research Partners for Registration of Metadata with the GRDC Data Catalogue v1.0

There are several ways to register Project Metadata describing RD&E Datasets with the GRDC Data Catalogue. In order of precedence, Project Metadata can be added to the GRDC Data Catalogue in the following ways:

1. Automated harvesting of Project Metadata from an Approved Repository
2. Manual registration of Project Metadata downloaded from an Approved Repository
3. Manual creation of a Project Metadata record using the GRDC Data Catalogue web-based form.

This order of precedence is designed to prioritise the creation of a single, unduplicated Project Metadata record for each RD&E Dataset. Following GRDC's distributed repository model, the original authoritative Project Metadata record will reside with the RD&E Dataset at the Approved Repository. Therefore, highest precedence has been attributed to the method of automated harvesting of Project Metadata directly from the Approved Repository.

### D-2.1. Automated Harvesting of Project Metadata from an Approved Repository

Project Metadata harvesting can be set up using the GRDC Data Catalogue platform to follow a defined schedule for upload. For this to work, the Approved Repository must

- have an application programming interface (API) implemented that is Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) compliant
- be able to provide Project Metadata in either of the RIF-CS or DataCite formats.

When metadata is harvested, the metadata that has been created at the site of the Approved Repository is copied automatically to the GRDC Data Catalogue. GRDC requires that the original record is completed and structured so that the resultant record harvested by the GRDC Data Catalogue has all mandatory fields completed.

This means that when the data is uploaded to the Approved Repository, care must be taken to create metadata that meets this criterion. GRDC also requests that where possible, as much metadata be included as possible, i.e., above and beyond those that are flagged as mandatory. This is good for data findability and will ensure that the dataset enjoys maximum visibility and potential for reuse in valuable future applications.

If in doubt about how to create a suitable metadata record, please contact your data steward or librarian for assistance.

### D-2.2. Manual registration of Project Metadata downloaded from an Approved Repository

Where automated harvesting of Project Metadata from an Approved Repository is not possible, the GRDC Data Catalogue is able to import Project Metadata from files that contain XML information structured according to either the RIF-CS or DataCite Project Metadata formats.

This is a manual process that involves downloading a Project Metadata file from the Approved Repository and importing it into the GRDC Data Catalogue.

Therefore, Project Metadata records in the GRDC Data Catalogue that are created in this way are not automatically synchronised with the record in the Approved Repository if any modifications are made to the latter.

### D-2.3. Manual addition of Project Metadata using the GRDC Data Catalogue web-based form

Where the options detailed in D-2.1 and D-2.2 are not possible, Project Metadata can also be created using the online form that is part of the GRDC Data Catalogue.

## D-3. GRDC Guideline for Research Partners for Describing GRDC RD&E Datasets v1.0

This Guideline provides a list of suggested information to include in the description of Project Metadata to ensure that the RD&E Dataset is adequately described for future use. This has been adapted from Cornell University's suggested descriptive metadata.

### D-3.1. Sharing/access information

- Licenses/restrictions placed on the RD&E Data. These should agree with the GRDC Project IP and Project Outputs (IPPO) Register entry that corresponds to the RD&E Dataset.
- Links to publications that cite or use the RD&E Data.
- Links to other publicly accessible locations of the RD&E Data.
- Links/relationships to ancillary datasets.
- Was data derived from another source? If yes, list source(s).
- Recommended citation for this RD&E dataset.

### D-3.2. Data & file overview

- File List: list all files (or folders, as appropriate for dataset organization) contained in the RD&E Dataset, with a brief description.
- Relationship between files, if important.

- Additional related RD&E Data collected that was not included in the current RD&E Dataset.
- Are there multiple versions of the RD&E Dataset? If yes, name of file(s) that was updated.
- Why was the file updated?
- When was the file updated?

### D-3.3. Methodological information

- Description of methods used for collection/generation of RD&E Data. Include links or references to publications or other documentation containing experimental design or protocols used in RD&E Data collection.
- Methods for processing the RD&E Data. Describe how the submitted RD&E Datasets were generated from the raw or collected RD&E Data.
- Instrument- or software-specific information needed to interpret the RD&E Data. Include full name and version of software, and any necessary packages or libraries needed to run scripts.
- Standards and calibration information, if appropriate.

### D-3.4. Environmental/experimental conditions

- Describe any quality-assurance procedures performed on the RD&E Data.
- People involved with sample collection, processing, analysis and/or submission.
- Description of experimental or environmental conditions that may affect the interpretation of the data.
- Degree to which the listed experimental or environment conditions are expected to affect interpretation of the data, quantified where possible. Where appropriate, specifics about which data are expected to be affected by a given condition should be provided.

### D-3.5. Data-specific information

- Number of variables
- Number of cases/rows
- Variable List: list variable name(s), description(s), unit(s) and value labels as appropriate for each.
- Definitions for codes or symbols used to record missing information in the RD&E Data.
- Specialised formats or other abbreviations used.

ENDS