Version 2.0
1 July 2024

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

Table of contents

BACKGROUND **2**

SECTION A: OVERVIEW **2**

A-1. IDENTIFYING RD&E DATA FOR COLLECTION AND STORAGE 2
A-2. UPLOADING RD&E DATA TO AN APPROVED REPOSITORY 3
A-3. REGISTERING PROJECT METADATA WITH THE GRDC DATA CATALOGUE 5
A-4. CHECKLIST FOR GRDC PARTNERS 6

SECTION B: INVESTMENT RD&E DATA MANAGEMENT **7**

B-1. GRDC RD&E DATA MANAGEMENT PLAN GUIDELINE FOR RESEARCH PARTNERS V1.1 7
B-1.1. Help Notes Worksheet *7*
B-1.2. Investment Details Worksheet *8*
B-1.3. GRDC Contract Code *8*
B-1.4. Investment Data Management Worksheet *9*
B-1.5. Data Storage Worksheet *10*

SECTION C: APPROVED REPOSITORIES **14**

C-1. GRDC LISTS OF APPROVED REPOSITORIES FOR RESEARCH PARTNERS V1.1 14
C-1.1. Institutional Data Repositories *15*
C-1.2. Domain-Specific Data Repositories *16*
C-1.3. General Research Data Repositories *16*
C-2. GRDC GUIDELINE FOR IDENTIFYING APPROVED REPOSITORIES FOR RESEARCH PARTNERS V1.0 17
C-3. GRDC GUIDELINE FOR DATA ACCESS CONDITIONS FOR RESEARCH PARTNERS 18
C-4. GRDC GUIDELINE FOR PROCESSING R&DE DATA ACCESS REQUESTS FOR REUSE BY RESEARCH PARTNERS 19

SECTION D: WORKING WITH METADATA **20**

D-1. GRDC METADATA COLLECTION FORM V3.2.1 20
D-2. GRDC GUIDELINE FOR RESEARCH PARTNERS FOR REGISTRATION OF METADATA WITH THE GRDC DATA CATALOGUE V1.0 30
D-2.1. Automated Harvesting of Project Metadata from an Approved Repository *30*
D-2.2. Import of XML metadata file that is compatible with the Data Catalogue XML harvester *31*
D-2.3. Manual creation of Project Metadata using the GRDC Data Catalogue web-based form *31*
D-3. GRDC GUIDELINE FOR RESEARCH PARTNERS FOR DESCRIBING GRDC RD&E DATASETS V1.0 32
D-3.1. Sharing/access information *32*
D-3.2. Data & file overview *32*
D-3.3. Methodological information *32*
D-3.4. Environmental/experimental conditions *33*
D-3.5. Data-specific information *33*
D-4. GRDC GUIDELINE FOR RESEARCH PARTNERS FOR DATA UPLOAD AND METADATA CREATION IN ZENODO 33
D-4.1. Adding a new dataset and metadata creation to the Zenodo 'GRDC' Community *33*

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

BACKGROUND

This document is intended as a reference for GRDC's Partners in following GRDC's RD&E Data Management Process. These guidelines will help you meet your obligations to capture and securely store the valuable data that you and fellow investment participants generate in a GRDC RD&E Investment, making it findable for future reference and use. By meeting these obligations, you contribute to a growing body of valuable Australian grains-related data that has the power to transform the Australian grains industry.

This document 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

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

The DMP template is available on the GRDC website 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 is responsible for uploading RD&E Data created in a GRDC Investment to the Approved Repository.

In these Guidelines, we distinguish between three different types of Approved Repository:

- Institutional Data Repositories
- Domain-specific Data Repositories
- General Research Repositories

Institutional Data Repositories are provided by institutions for use by staff, Domain-specific Data Repositories relate to a given research domain or scope, and General Research Repositories accept research data of a very broad or unlimited scope. See Section C-1 for a list of Approved Repositories broken down by these types.

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 in the Data Management Plan.

The GRDC Partner will collaborate with the GRDC Investment Manager to establish the Data Access Conditions for the RD&E Data uploaded to an Approved Repository. These Access Conditions determine the level of accessibility for third parties and are agreed upon by the data owners at the time of upload by the GRDC Partner's Data Custodian. GRDC's policy position is to make RD&E Data as open as possible and as closed as necessary.

The process for uploading RD&E Data to the Approved Repository will be specific to the Approved Repository. Approved Repositories will have staff that are on hand to assist with inquiries. Specifically, many Institutional Data Repositories at universities have processes in place to assist with compliance with GRDC contracts, including automating registration of project metadata with the GRDC Data Catalogue (Section A-3).

GRDC Partners should contact staff at the chosen Approved Repository for information about the repositories and how to use them. Contact details can be found in Table 1, Table 2, and Table 3 in Section C-1.

Relevant Guidelines – Section C

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

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

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

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

C-3. GRDC Guideline for Data Access Conditions for Research Partners

This document provides detailed guidance on the three possible RD&E Data Access Conditions applicable to RD&E Data.

C-4. GRDC Guideline for processing R&DE Data Access Requests for reuse by Research Partners

This document outlines the workflow for handling RD&E Data requests, involving GRDC and Research Partners in decision-making, and adhering to GRDC's overarching principle of making data as open as possible and as closed as necessary.

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

For each RD&E Dataset 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; and
- 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.

Relevant Guidelines – Section D

D-1. GRDC Metadata Collection Form
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
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
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.

D-4. GRDC Guideline for Research Partners for data upload and metadata creation in Zenodo
This guideline provides detailed steps on how to enter data into the Zenodo Repository upload form to ensure a high-quality metadata record is harvested into the GRDC Data Catalogue.

A-4. Checklist for GRDC Partners

- Identify RD&E Datasets to be created in each investment and record them in a Data Management Plan.
- Update the Data Management Plan during the term of the investment.
- 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: INVESTMENT RD&E DATA MANAGEMENT

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

For every GRDC Investment that generates RD&E Data, a GRDC Data Management Plan must 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, the GDRC Partner is responsible for completing and updating the Data Management Plan. GRDC recommends contacting your organisation's Data Steward, Library, or other Research Data support service for assistance. They can provide information on available resources including shared workspaces and institutional data repositories.

The Data Management Plan will be completed by the GRDC Partner during the contract negotiation/execution phase of the investment management cycle. The GRDC Partner will update the Data Management Plan throughout the course of the contracted work to keep up to date with any changes.

GRDC has created a Data Management Plan template in Microsoft Excel spreadsheet for use in GRDC Investments. The GRDC Data Management Plan template is available for download from the GRDC website (here). The GRDC Data Management Plan is comprised of four worksheets with the following labels:

- Help notes
- Investment details
- Investment data management
- Data storage.

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

GRDC Partners are required to complete the GRDC Data Management Plan template and submit the document to GRDC as the Microsoft Excel file. Please do not enter the information in the Research Contract Schedule 3.

B-1.1. Help Notes Worksheet

Useful details for GRDC Partners include:

(i) File naming convention to distinguish the different versions of the Data Management Plan (DMP) submitted during the life of the investment.

The file name for DMP submitted at Contract execution will be:

CONTRACT_CODE-Data Management Plan-Execution version.xlsx

The file name for the DMP submitted as an Updated DMP will follow the convention: "CONTRACT CODE- Data Management Plan-MS NUMBER", e.g., WSU2303-001RTX-Data Management Plan-MS102.xlsx

(ii) Contractual requirements/activities associated with "Submitting an Updated DMP" Milestone. These include the following:

- Preparing data for long-term storage and re-use
- Depositing data in an Approved Repository as stated in the DMP
- Clarifying RD&E data Access Conditions (Open, Restricted, Conditional) of the data being deposited in an Approved repository
- Registering Project Metadata with the GRDC Data Catalogue (refer to Guideline for Registration of Metadata – see Section D2)
- Obtaining confirmation (URL/DOI) that the metadata/dataset has been deposited in an Approved repository and Project Metadata registered with the GRDC Data Catalogue and providing those details in column H of the Data storage sheet of the DMP template

(iii) GRDC Investment Manager will verify the URL/DOI as confirmation of successful deposition of data in the Approved Repository and metadata registration with the GRDC Data Catalogue. This is a requirement for approving an Updated DMP Milestone.

B-1.2. Investment Details Worksheet

This worksheet is for capturing information about the investment 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 the personnel 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-1.3. GRDC Contract Code

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

B-1.3.1. Contract title

This matches the Research Contract Title.

B-1.3.2. Prepared by

The party that has prepared this Data Management Plan, including their organisation,

position, phone number.

B-1.3.3. 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-1.3.4. Principal Investigator

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

B-1.3.5. 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-1.4. Investment Data Management Worksheet

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-1.4.1. 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-1.4.2. Ethical, confidentiality or privacy considerations

Here the GRDC Partner will identify any sensitivities associated with the Background or Third Party RD&E data being used in the GRDC 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-1.4.3. 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-1.4.4. 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-1.4.5. Background Data

As GRDC improves its approach to RD&E 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. Background Data that is used in the GRDC investment should also be listed in the IPPO Register.

B-1.5. Data Storage Worksheet

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-1.5.1. 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 dissemination or commercialisation and of this RD&E Dataset are well understood by referring to its entry in the IPPO Register.

B-1.5.2. Dataset description

This is a short description of the RD&E Dataset that details its contents. The description should inform other parties about the content of the RD&E Data and allow users to decide whether it is of interest to them.

B-1.5.3. Storage location

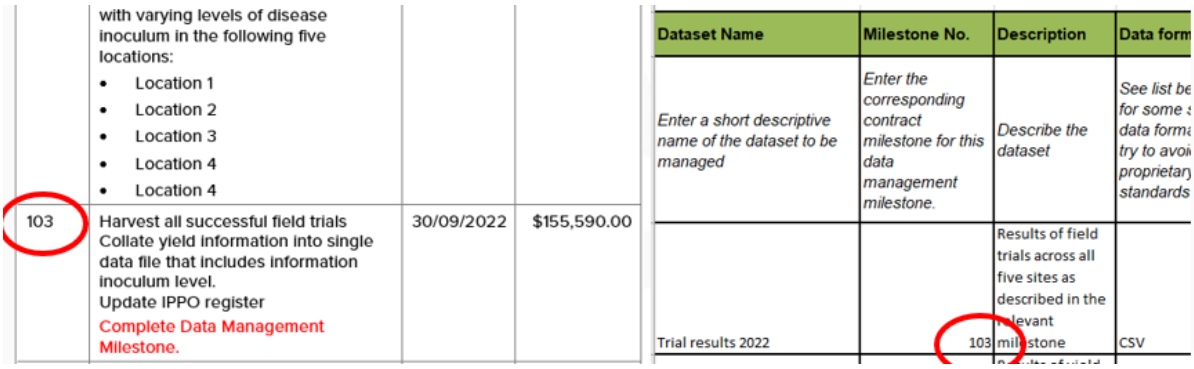
The storage location identified here will be an Approved Repository that has been agreed upon by both the GRDC Investment Manager and the GRDC Partner. A list of Approved Repositories is provided in the **GRDC List of Approved Repositories for Research Partners v1.1**

B-1.5.4. Milestone No.

RD&E Data storage and cataloguing activities within an investment are tied to Research Contract Milestones. Each row in the table corresponds to a valuable RD&E Dataset that is expected to be generated through activities described in a contract Milestone. The associated Milestone number is entered here.

The 'Milestone No.' field records the contract Milestone when the listed RD&E Dataset will be deposited into the selected Approved Repository and helps to plan for when RD&E Data storage and cataloguing will take place throughout the investment. At the Milestone time, Project metadata will also be generated and registered with the GRDC Data Catalogue.

In the figure below, RD&E Data is collected at contract Milestone 103 (extract from Schedule 1 Milestones section on the 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 (extract shown on the 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: • Location 1 • Location 2 • Location 3 • Location 4 • Location 4		
103	Harvest all successful field trials Collate yield information into single data file that includes information inoculum level. Update IPPO register 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 proprietar standards*
Trial results 2022	103	Results of field trials across all five sites as described in the relevant milestone	CSV

B-1.5.5. 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-1.5.6. Access Conditions

The Project Metadata describing RD&E Data will be publicly available via the GRDC Data Catalogue and will clarify which of the following Access Conditions apply to a given Dataset, see GRDC Guideline for Data Access Conditions for Research Partners.

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 (https://creativecommons.org/about/cclicenses/). GRDC advocates for Open Access Conditions for GRDC RD&E Data.
- Conditional: RD&E Data are publicly available subject to certain conditions being met. A common example of Conditional Access is an embargo placed on access to allow for publication of research or commercialisation activities. If an embargo is applied, the embargo period should be specified and a date for expiry of the embargo should be included.
- Restricted: Metadata are findable and data are only accessible by third parties with approval from the Data Custodian and the relevant GRDC Investment Manager.

B-1.5.7. IPPO Register "Project Outputs" Number

This field would be blank at the start of the investment (for a DMP submitted at contract execution) and will be populated over time as the Project Outputs are generated and the IPPO Register is updated.

This is the number that appears in the “No” column of the IPPO Register Project Outputs table to help clarify any limitations on dissemination or commercialisation of the GRDC RD&E Data due to and Background Material or Third Party Material embodied in the dataset.

B-1.5.8. Data uploaded to Approved Repository?

Once the associated data management Milestone has been carried out and data appears in an Approved Repository, a URL from the Approved Repository for the Project Metadata should be available. This provides the GRDC Investment Manager with an opportunity to check that the URL is working, and if need be, check data in the Approved Repository.

SECTION C: APPROVED REPOSITORIES

Approved Repositories are an essential component of GRDC's approach to RD&E Data Management and are used to publish RD&E Data and related metadata generated within GRDC investments.

GRDC have developed a set of criteria by which Approved Repositories are defined (See Table 4). Based on these, we've compiled lists of Approved Repositories, repositories that can be used to store valuable RD&E Datasets generated in GRDC Investments. These are presented below.

Procedures to publish metadata and data in accordance with the GRDC's requirements will be specific to the Approved Repository. The GRDC Partner should consult the relevant contact as listed in Table 1,Table 2, and Table 3 in Section C-1. for support and instruction for uploading data to the Approved Repository.

C-1. GRDC Lists of Approved Repositories for Research Partners v1.1

GRDC has approved several repositories as Approved Repositories for GRDC RD&E Data in accordance with the criteria found in Table 4. Lists of these repositories are presented in this Section. These lists are for Institutional Data Repositories, Domain-Specific Data Repositories, and General Research Data Repositories, respectively.

If there is a data repository that is not included in these lists and you think it will be suitable for long-term storage of data from a GRDC investment, please let us know by emailing data.catalogue@grdc.com.au.

C-1.1. Institutional Data Repositories

This section contains a list of Approved Repositories that are owned by research institutions.

Table 1. List of Institutional Data Repositories

Name	URL	Support contact
University of Sydney eScholarship Repository	https://ses.library.usyd.edu.au/	ses.admin@sydney.edu.au
University of Queensland UQ-RDM/eSpace	RDM: https://rdm.uq.edu.au/ (UQ internal) https://espace.library.uq.edu.au/	data@library.uq.edu.au
CSIRO Data Access Portal	https://data.csiro.au/	researchdatasupport@csiro.au
UWA Profiles and Research Repository	https://research-repository.uwa.edu.au/	staffsupport-lib@uwa.edu.au
University of Adelaide Figshare	https://adelaide.figshare.com/	library@adelaide.edu.au
espace-Curtin University	https://espace.curtin.edu.au/	researchdata@curtin.edu.au
University of Melbourne Figshare	https://dataservices.research.unimelb.edu.au/services/43/	digital-stewardship@unimelb.edu.au
Queensland Department of Agriculture and Fisheries eResearch Archive	https://era.daf.qld.gov.au/	RIS.Enquiries@daf.qld.gov.au
Australia National University ANU Data Commons	https://datacommons.anu.edu.au/DataCommons	repository.admin@anu.edu.au
UniSQ RISE Research Repository	https://www.unisq.edu.au/library/research-support/research-repository	research.repository@unisq.edu.au
UniSA Research Data Storage (University of South Australia)	https://i.unisa.edu.au/askresearch/data-management/data-storage/	
Murdoch Research Portal	https://researchrepository.murdoch.edu.au/cgi/oai2	repository@murdoch.edu.au
LaTrobe University (Opal)	https://opal.latrobe.edu.au	libraryresearchsupport@latrobe.edu.au
Charles Sturt University Research Outputs (CRO)	https://researchoutput.csu.edu.au/	cro@csu.edu.au

C-1.2. Domain-Specific Data Repositories

This section contains a list of domain-specific Approved Repositories. Such repositories exist to collect data of the same type and provide opportunities for data aggregation on a large scale. It may be more valuable to store data in a domain-specific repository than in an institutional repository due to such opportunities.

Table 2. List of Domain-Specific Data Repositories

Name	Data Domain	URL	Support Contact
Atlas of Living Australia	Biodiversity	www.ala.org.au/	support@ala.org.au
Bioplatforms Australia	'Omics	https://data.bioplatforms.com/	help@bioplatforms.com
Australian Data Archive	Social sciences	https://ada.edu.au/	ada@ada.edu.au
NCBI GenBank	DNA Sequencing	www.ncbi.nlm.nih.gov/genbank/	https://support.nlm.nih.gov/support/create-case/
NCBI Sequence Read Archive (SRA)	DNA Sequencing	www.ncbi.nlm.nih.gov/sra	https://support.nlm.nih.gov/support/create-case/
Harvard Dataverse Repository	General	https://support.dataverse.harvard.edu/home	support@dataverse.harvard.edu

C-1.3. General Research Data Repositories

A general research repository can be used to store data that spans disciplines and institutions. The general research Approved Repositories listed in the table below can be used if a suitable institutional or domain-specific data repository cannot be found. There may still be a limitation on the scope of these repositories. For example, the 'Data WA' repository is a general repository insofar as it can store data across institutions and disciplines, but it only accepts data from Western Australian state departments.

Table 3. List of General Research Data Repositories

Name	URL	Support Contact
Zenodo	https://www.zenodo.org	https://zenodo.org/support
Data WA	https://data.wa.gov.au	info@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 Table 4.

Table 4. Criteria for Identifying an Approved Repository

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: • provides a publicly facing web-based search tool for data discovery; • assigns a persistent identifier to each RD&E Dataset and provides it as Uniform Resource Locator (URL) as part of the associated Project Metadata; • 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.

C-3. GRDC Guideline for Data Access Conditions for Research Partners

Access Conditions define the extent to which RD&E Data will be accessible by third parties. This guideline provides detailed guidance on the three possible RD&E Data Access Conditions applicable to RD&E Data.

The Access Condition for an RD&E Dataset is determined by agreement of the owners of the RD&E Data and is assigned to the RD&E Data when the RD&E Data is uploaded to an Approved Repository by the GRDC Partner's Data Custodian.

GRDC's policy position is for GRDC RD&E Data to be as open as possible and as closed as necessary.

C-3.1.1. Open Access

Open Access means the RD&E Data are publicly accessible by a third party without the requirement to seek prior permission from the owners of the RD&E Data.

Applying Creative Commons (CC) licensing is the recommended way to achieve flexible publication and re-use of RD&E Data without further permission being required. Most of GRDC's research partners require the use of CC licences to facilitate in designating RD&E Data as Open Access.

There are six different CC license types, ranging from least permissive to most permissive. For more information on the CC licences visit the website https://creativecommons.org/share-your-work/cclicenses/

The CC-BY licence is recommended for GRDC RD&E data that has been assigned Open Access. This licence enables re-users to distribute, remix, adapt and build upon the RD&E Data in any medium or format, so long as attribution is given to the creator. The CC-BY licence allows for research and commercial use of the RD&E Data.

The following attribution statement should be used: "Copyright Grains Research and Development Corporation (GRDC) and [name of co-owner], [YEAR]".

C-3.1.2. Conditional Access

Conditional Access means the RD&E Data are publicly available subject to certain conditions being met, such as:

(i) Expiry of an Embargo period. In this scenario, the RD&E Data will be available but only after a specified date. Embargoes are often applied to RD&E Data to allow researchers to publish the results of their research in scientific publications. The default embargo period for GRDC RD&E Data is 12 months after the end of the contract in which the RD&E Data was generated. GRDC may consider a longer

embargo period upon request from the GRDC Partner.

(ii) A fee applies to access the data.

(iii) The data is only accessible at a specific physical location.

(iv) The data is only available to certain users (e.g. Australian users only, non-commercial use only).

Once the condition for access has been met, the RD&E Data should be made available under an appropriate Creative Commons licence.

C-3.1.3. Restricted Access

Restricted Access means that the data is not publicly accessible and a third party must request access to the RD&E Data from the owners. If the owners of the RD&E Data agree to share the data, then the third party will be required to enter into a licence to access the RD&E Data.

Restricted Access may be required where the RD&E Data:

(i) contains personal information that needs to be de-identified before the data can be shared with third parties;

(ii) contains confidential or commercially sensitive information such as data that underpins intellectual property applications or will be packaged with other project outputs for commercialisation;

(iii) contains industry-sensitive information that may impact trade and market access such as chemical residue testing, biosecurity incursions, volumes of grain in storage, purchase and sales prices; or

(iv) is subject to contractual arrangements that prevent the sharing of the RD&E Data.

C-3.1.4. Project Metadata accessibility to third parties

The Project Metadata for RD&E Data will be registered and accessible to third parties on the GRDC Data Catalogue irrespective of the Access Condition that has been assigned to the actual RD&E data.\

C-4. GRDC Guideline for processing R&DE Data Access Requests for reuse by Research Partners

The GRDC Guideline for processing R&DE Data Access Requests for reuse by Research Partners outlines the workflow for handling RD&E Data requests, involving GRDC and Research Partners in decision-making, and adhering to GRDC's overarching principle of making data as open as possible and as closed as necessary.

SECTION D: WORKING WITH METADATA

The GRDC Metadata Collection Form v3.2 aims to standardise and streamline the collection of metadata for RD&E Data. This schema is an implementation of the RIF-CS metadata standard. All records in the GRDC Data Catalogue follow this standard. Below each question (metadata field), there is helper text for the information that is expected. Mandatory fields are marked with an asterisk.

D-1. GRDC Metadata Collection Form v3.2.1

1. **What is the name of the collection or dataset?** *
   Aim to be descriptive but concise, so that the name is unique and meaningful to users of the GRDC Data Catalogue. Note: the name of the collection or dataset should be different to the GRDC project title and should not include the GRDC contract code.

2. **What is this metadata record for?** ***Select one of the following.*** *
   The GRDC Data Catalogue will default to Collection.
   Collection: A collection of objects, grouped according to a shared criteria, which are stored and managed as a collective group. E.g., an entire book.
   Dataset: Structured data that is an input to, or output of research. E.g., a chapter of a book.

   ☐ Collection ☐ Dataset

3. **Which organisation is contributing this metadata record?** *
   Specify the name and Research Organisation Registry (ROR) of the research organisation that is contributing this metadata record in full, e.g., New South Wales Department of Primary Industries (https://ror.org/050khh066).
   Click here to search the Research Organisation Registry (ROR).

Name:	
ROR:	

Provide the name, ROR and role of any other organisations that have contributed to this collection or dataset. ***Repeat this block as needed.***

Roles: Owner (legally owns the collection or dataset), Manager (maintains the collection or dataset and makes it accessible), Collector (collected, generated, created or aggregated the collection or dataset), Enricher (provided additional value to the collection or dataset) and Associated (has an unspecified relationship with the collection or dataset).

Name:	
ROR:	
Role: *Select from the drop-down menu*	☐ Owner ☐ Manager ☐ Collector ☐ Enricher ☐ Associated

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

GRDC contract code:	
GRDC project title:	

Provide the contract code and project title of any additional GRDC projects and specify how they are related to the collection or dataset. *Repeat this block as needed.*

Relationships: Output (the collection or dataset is a product of the GRDC project) and Associated (the collection or dataset has an unspecified relationship with the GRDC project).

GRDC contract code:	
GRDC contract title:	
Relationship: *Select from the dropdown menu*	☐ Output ☐ Associated

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

While this field is mandatory, it may not be fillable during initial metadata gathering. Follow your institutional processes to mint a persistent identifier and add to the record when available.

6. **What is the Handle for this collection or dataset?**
   While this field is mandatory, it may not be fillable during initial metadata gathering. Follow your institutional processes to generate a Handle and add to the record when available.
   Click here to learn more about the Handle.Net Registry, run by Corporation for National Research Initiatives (CNRI).

7. **What is the URL for the metadata or a landing page which allows users to know how to access the collection or dataset?** *
   If the collection or dataset has a metadata record in a repository, include the URL here.

8. **What is the URL that triggers a direct data download?**
   If the collection or dataset is publicly downloadable, include the URL here.

9. **When was the data collected or the observations made? *Use the YYYY format.* ***
   Specify the year data collection started and the year it ended.

Start date:	
End date:	

10. **When was the collection or dataset published? *Use the YYYY format.* ***
    If the metadata record for the collection or dataset has been registered with your institutional repository, include the year here. E.g., 2023.

11. **What is the spatial coverage of the collection or dataset?** *
    The GRDC Data Catalogue will have a map view showing a highlighted region bounded by the minimum and maximum longitude and latitude coordinates, if these are defined. This bounding box also empowers searching through the catalogue by regions, e.g., if a user is interested in trials only in the Southern Region.
    Note: this bounding box also allows for privacy. If the locations of paddocks used to generate the data are confidential, then provide a larger bounding box to add some uncertainty. This allows for a safe obfuscation of the precise location, while still letting the dataset be accurately sorted in general region searches and categories.
    Optionally (and/or alternatively), describe the spatial coverage in text. This can be appropriate,

e.g., if there is too wide a range of locations for the collection or dataset, or there is additional information about the spatial extent that cannot be conveyed through the GPS coordinates alone.

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

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

The brief description should be a concise overview of the collection or dataset. The description should inform users about the content of the data and allow users to decide whether it is of interest to them.

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

A full description of the collection or dataset is not the same as a project summary or manuscript summary. The full description should provide details about the collection or dataset that would enable someone to use the data, without contacting the original researchers for further information. Suggested content includes:

- Data and file overview: file list with a brief description of files and the relationship between the files

- Methodological information: description of the methodology used for the collection/generation of the data, methods for processing the data and instrument/software specific information

- Environmental/experimental conditions

- Data specific information: the number of variables, variable list and definitions of codes or symbols

See D-3. Describing GRDC RD&E Datasets for more details on descriptive metadata to include in this section.

14. Provide the names, identifiers (at least one of the three) and roles of those involved in producing the collection or dataset? *Repeat this block as needed.* *

Roles: Principal Investigator, Collector (collected, generated, created or aggregated the collection or dataset), Enricher (provided additional value to the collection or dataset) and Associated (has an unspecified relationship with the collection or dataset).

Click here to search for an ORCID ID
Click here to search for a Scopus Author ID
Click here to search for a ResearcherID

Given name:	
Family name:	
ORCID ID:	
Scopus ID:	
Researcher ID:	
Role:	
	☐ Principal Investigator ☐Collector ☐ Enricher ☐ Associated

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

Where possible, use an email address for the lead organisation, rather than an individual researcher.

16. What are the keywords for the collection or dataset? *Repeat this block as needed.*

The AGROVOC Multilingual Thesaurus provides a list of agricultural keywords that can be used.

Click here to search the AGROVOC Multilingual Thesaurus.

Keyword:	

Provide Fields of Research (FoR) codes for this collection or dataset. *Repeat this block as needed.*

Provide at least one FoR code. FoR allows R&D activity to be categorised according to common knowledge domains and/or methodologies used in the R&D.

Click here to access the 2020 ANZSRC FoR classification.

FoR code: *Select from the dropdown menu or add another code*	☐ 300105 Genetically modified field crops and pasture ☐ 300109 Non-genetically modified uses of biotechnology ☐300202 Agricultural land management ☐ 300204 Agricultural management of nutrients ☐ 300205 Agricultural production systems simulation ☐ 300206 Agricultural spatial analysis and modelling ☐ 300207 Agricultural systems analysis and modelling ☐ 300208 Farm management, rural management and agribusiness ☐ 300209 Germplasm management ☐ 300299 Agriculture, land and farm management not elsewhere classified ☐ 300403 Agronomy ☐ 300404 Crop and pasture biochemistry and physiology ☐ 300405 Crop and pasture biomass and bioproducts ☐ 300406 Crop and pasture improvement (incl. selection and breeding) ☐ 300407 Crop and pasture nutrition ☐ 300408 Crop and pasture post harvest technologies (incl. transportation and storage) ☐ 300409 Crop and pasture protection (incl. pests, diseases and weeds) ☐ 300410 Crop and pasture waste water use ☐ 300412 Organic and low chemical input crop production ☐ 300413 Pollination biology and systems ☐ 300499 Crop and pasture production not elsewhere classified ☐ 310802 Plant biochemistry

	☐ 310803 Plant cell and molecular biology ☐ 310804 Plant developmental and reproductive biology ☐ 310805 Plant pathology ☐ 310806 Plant physiology ☐ 310899 Plant biology not elsewhere classified ☐ 410101 Carbon sequestration science ☐ 410302 Biological control ☐ 410601 Land capability and soil productivity ☐ 410602 Pedology and pedometrics ☐ 410603 Soil biology ☐ 410604 Soil chemistry and soil carbon sequestration (excl. carbon sequestration science) ☐ 410605 Soil physics ☐ 419999 Other environmental sciences not elsewhere classified ☐ 450301 Aboriginal and Torres Strait Islander agriculture and forestry ☐ 450399 Aboriginal and Torres Strait Islander environmental knowledges and management not elsewhere classified ☐ 451599 Pacific Peoples environmental knowledges not elsewhere classified ☐ 451902 Global Indigenous studies environmental knowledges and management

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

Describe the access rights and access constraints, including who may access the collection or dataset and when access may occur (including any embargos). Refer to the GRDC contract and the IP and Project Outputs (IPPO) register.

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

A data licence is a legal arrangement between the creator of the data and the end-user specifying what users can do with the data. Specify the licence used for the collection or dataset from the dropdown menu, including an optional textual description if desired.

Licence description:	
Licence type: *Select one of the following.*	☐ CC-BY ☐ CC-BY-SA ☐ CC-BY-ND ☐ CC-BY-NC ☐ CC-BY-NC-SA ☐ CC-BY-NC-ND ☐ CC0 ☐ Other

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

Specify the access rights for the collection or dataset, including an optional access rights statement. Q. 17 can be used to elaborate on the conditions if 'conditional' access is selected.

Access rights:	
Access rights type: *Select one of the following.*	☐ Open ☐ Conditional ☐ Restricted ☐ Embargoed

20. Provide the names and identifiers of any other collections or datasets and specify how they are related to the collection or dataset. Repeat this table as needed.

Provide a DOI or a Handle as the identifier for the related collection or dataset.

Relationships: Has part (contains the related collection), Is part of (is contained within the related collection), Has derived collection (the related collection is derived from the collection or dataset e.g., through analysis), Is derived from (the collection or dataset is derived from the related collection e.g. through analysis), Has version (the collection or dataset is versioned by the related collection), Is version of (the collection or dataset is a version of the related collection), Associated (has an undefined relationship with the related collection).

Click here to learn more about the Handle.Net Registry run by Corporation for National Research Initiatives.

Name:	
DOI:	
Handle:	
Relationship: *Select from the drop-down menu*	☐ Has part ☐ Is part of ☐ Has derived collection ☐ Is derived from ☐ Has version ☐ Is version of ☐ Associated

21. Provide the title and identifier for any publications and specify how they are related to the collection or dataset. Repeat this table as needed.

Include details of related publications or outputs (including research papers, posters, websites, blog posts, GRDC Updates, GroundCover articles or media interviews) that reference the collection or dataset or the work done to generate it. Include DOIs or URLs where possible.

Relationships: Is cited by (the publication includes the collection or dataset in a citation), Is referenced by (indicates the collection or dataset is used as a source of information by the publication), Is documented by (indicates the publication is documentation about/explaining the collection or dataset), Is supplemented by (indicates that the collection or dataset is a supplement to the publication), Is supplement to (indicates that the publication is a supplement to the collection or dataset), Is reviewed by (the collection or dataset presents statements, ideas or conclusions that are reviewed by the publication), Is supported by (the collection or dataset receives intellectual or factual support from the publication).

Title:	
DOI:	
Handle:	
IGSN:	
Other identifier	
Relationship: *Select from the dropdown menu*	☐ Is cited by ☐ Is documented by ☐ Is supplement to ☐ Is supported by ☐ Is referenced by ☐ Is supplemented by ☐ Is reviewed 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, in order of preference:

1. Automated harvesting of Project Metadata from an Approved Repository.
2. Import of XML metadata file that is compatible with the Data Catalogue XML harvester.
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.

For guidance on how to interact with the GRDC Data Catalogue, please see Instructional videos on this link Instructional videos

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, a harvester must exist in the GRDC Data Catalogue platform that can communicate with the selected Approved Repository. If your organisation hosts its own data repository, check with your organisation to see if the GRDC Data Catalogue is capable of harvesting metadata from your repository.

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 (See the GRDC Metadata Collection Form v3.2).

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.

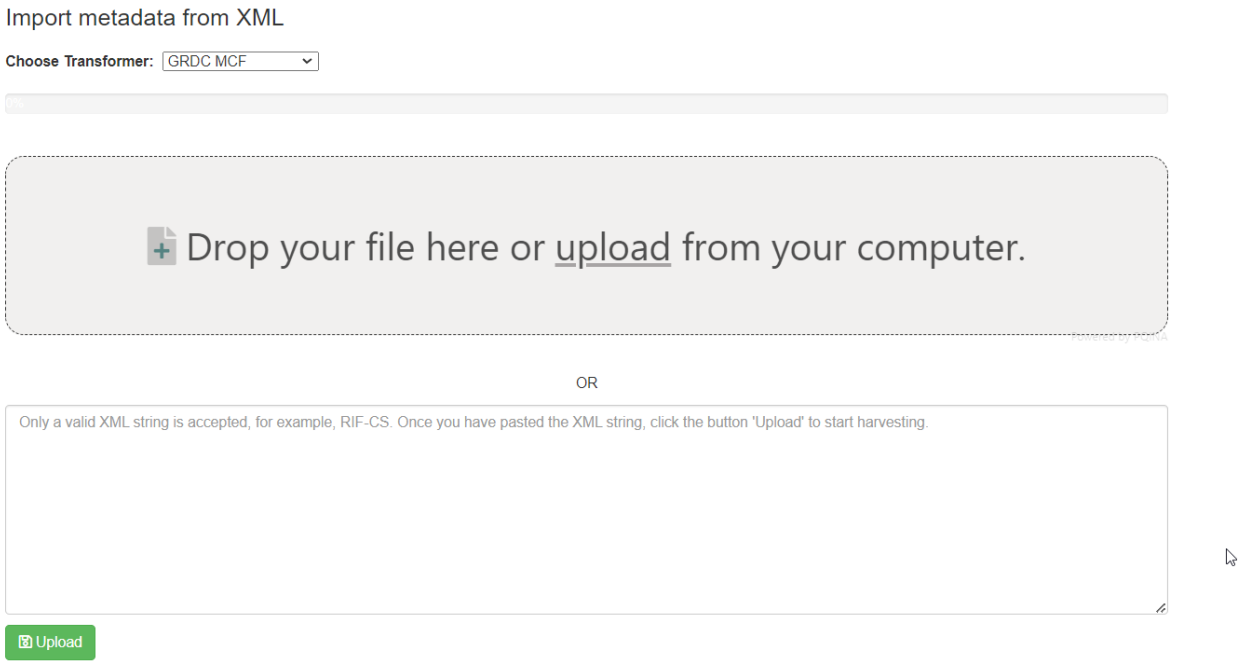
If in doubt about how to create a suitable metadata record, please contact your data steward or librarian for assistance.

D-2.2. Import of XML metadata file that is compatible with the Data Catalogue XML harvester

If there is no harvester for the Approved Repository that you have chosen for storing Project Data, it is possible to transfer metadata records from the repository to the GRDC Data Catalogue in an XML file.

The GRDC Data Catalogue can import a metadata record in the XML file, provided it is structured in the right way. An exemplar **GRDC Metadata Record Upload Template.xml** is available for download.

This file can then be imported into the GRDC Data Catalogue, replicating the record from the metadata record from the Approved Repository



If you would like to upload metadata to the GRDC Data Catalogue in this way, you can provide an example XML file to the support services for the Approved Repository. They may be able to organise an export that matches this format.

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

Where an automatic harvester is not available for the chosen repository and it is not possible transfer the metadata via an XML file, manual metadata creation via the GRDC Data Catalogue is possible.

For this, make sure you have an account with the GRDC Data Catalogue. Also make sure that you your organisation is registered as a Stewardship Organisation. Please contact data.catalogue@grdc.com.au for assistance.

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 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 organisation) 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 use GRDC Guideline for Research Partners for data upload and metadata creation in Zenodo

D-4. GRDC Guideline for Research Partners for data upload and metadata creation in Zenodo

D-4.1. Adding a new dataset and metadata creation to the Zenodo 'GRDC' Community

Communities are groupings of records that are associated with a given organisation. On the Zenodo platform (https://zenodo.org/), there is a GRDC community. When you upload a GRDC RD&E Dataset you will need to add it to the GRDC community so that it is easily findable alongside other GRDC RD&E Datasets on Zenodo and that the record can be found by, and transferred to, the GRDC Data Catalogue.

This guideline provides detailed steps on how to enter data into the Zenodo Repository upload form on the GRDC community to ensure a high-quality metadata record is harvested into the GRDC Data Catalogue.

See: GRDC Guideline for data upload and metadata creation in Zenodo for Research Partners v1.0.