

FAIR DATA Guidelines for Project Outputs

DATE EFFECTIVE	19 June 2020
REVIEW DATE	19 June 2023 - extended until 19 June 2024 unchanged
TYPE	Guideline
OWNER	Manager, Skilled Workforce Development, ARDC
APPROVED BY	Keith Russell, Director Outreach
CATEGORY	Projects
VERSION NUMBER	1.4
AUDIENCE	External
CONTENT ENQUIRIES	contact@ardc.edu.au
RELATED DOCUMENTS	FAIR ARDC or ARDC-funded Materials Policy Persistent Identifiers for ARDC Investments
PURPOSE	Provide guidelines to partners for ARDC Investments on how data outputs involved in the partnership will be made more FAIR in the course of the project.

Context

This document provides guidelines and resources for ARDC projects on how data outputs involved in the project will be made more FAIR. The ARDC is supportive and encourages the adoption of the FAIR principles as a valuable way in making research outputs more reusable, both for humans and machines. This also aligns with the requirements for research institutions as the FAIR principles are referenced in the [Guide for Management of Data and Information in Research](#) that accompanies the Code for the Responsible Conduct of Research.

ARDC is an NCRIS facility and it is the [policy](#) of the Commonwealth NCRIS program that “data generated, created, captured or stored by NCRIS funded projects will be made available to the wider research community based on the F.A.I.R. principles.”

This document describes our expectation of how FAIR will be addressed in ARDC projects.

The [FAIR data principles](#) present a number of key areas in which data can be made FAIR, without being prescriptive on how this should be achieved. The Research Data Alliance (with global input) has published a [model on](#) how to assess FAIR. The current document translates these into a practical set of actions for ARDC projects, that will make the data more FAIR. We consider these realistic aims based on the state of maturity of services and standards currently available. The links provide guidance on how the different aspects can be achieved. Please consult with your ARDC liaison for the project as they will be able to provide guidance, support and organise training on all these aspects.

If you believe there is a good reason why a certain aspect cannot be addressed, please discuss this with your project liaison as early on as possible. The reason for it not being addressed will need to be documented and provided in the progress and final report.

Findable

	Action	Required/ Recommended	Links
1	Assign all data outputs appropriate PIDs (DOIs are the default identifier for this purpose)	Required	PIDs guidelines
2	All data outputs have metadata to enable discovery. The discovery metadata schema and fields to be completed for all data outputs should be discussed by partners	Required	See Appendix 1 for required fields List of mandatory and optional collection elements

	and ARDC at project initiation.		for RDA Guide to What makes a good collection record
3	All data outputs generated by the project must have a record in Research Data Australia	Required	Providing metadata records to Research Data Australia Defining a data collection
4	Records for all data outputs are registered with relevant discipline-specific discovery aggregators (if one exists)	Recommended	For a list of potential discovery mechanisms see Re3Data
5	The persistent identifier for the data being described must be included in the metadata	Required	For example, providing the dataset DOI in the <identifier> element in RIF-CS

Accessible

6	All data outputs are made as openly available as possible; they are only closed where necessary	Required	For guidance on what Open means see e.g. Open Data Handbook
7	All data outputs are made available through a repository	Required	See guidance on How to select a repository
8	All data outputs are available as a download and/or accessible through an open, documented API	Required where data is not closed	See guidance on APIs
9	If the data outputs are not openly available there is a clear description on the landing page on how to request access to the data outputs and conditions that need to be met	Required where relevant	See Guidance on Access Rights
10	The persistent identifier for the data output will point to a landing page about the data output, even if the data output is not public.	Required	See Best Practice for landing pages

11	If the data output is not openly available there is an authorisation and authentication procedure to provide access to the data.	Required where relevant	
12	The persistent identifier for the data output will continue to point to a landing page, even if the data output is no longer available. There is a policy to maintain these landing pages.	Required	See Best Practices for Tombstone pages

Interoperable

13	Data outputs use community-agreed standard data formats.	Required where research community agreed standard formats exist	For standards in specific disciplines see FAIRsharing.org
14	Metadata uses community-agreed standards	Required where research community agreed metadata standards exist	ARDC Metadata guide
15	Data and metadata use community-agreed vocabularies, data models and ontologies (preferably internationally agreed ones where they exist)	Recommended	Good practice in vocabulary creation Research Vocabularies Australia Vocabularies and research data
16	Metadata contains persistent identifiers for research objects and entities (people, organisations) linked to the dataset(s). This includes links to ORCIDs, grantIDs, RAIDs, DOIs, IGSNs of related materials	Required	Which PIDs to use: ARDC Persistent Identifiers (PIDs): Guidelines for ARDC Investments ARDC Citation and Identifiers guide

Reusable

17	All data outputs are assigned a machine readable (default choice should be a CC-BY 4.0) licence.	Required	ARDC Research Data Rights Management Guide
18	The licence information is available in a machine readable form on the landing page that the persistent identifier refers to.	Required	Guidance on Machine readable reference to CC licence
19	The landing page contains a citation statement for the data output	Required	ARDC Data Citation guide DataCite DOI Display Guideline
20	Provenance information on the data output is attached alongside the data	Recommended	ARDC guide on data provenance
21	Relevant discipline-specific metadata to enable reuse is captured and presented alongside the data output following research community best practice.	Recommended	See links in ARDC Metadata guide

Appendix 1 - Minimum metadata requirements: Collections

Metadata field			Required/ Recommended	Meaning
RIF-CS	Dublin Core	DataCite		
Identifier	dc:identifier	1 Identifier	Required	A unique identifier for the resource, preferably a DOI.
Collection Type	dc:type	10 ResourceType	Required	The type of collection being described, i.e. collection, dataset, software, etc.
Title	dc:title	3 Title	Required	The name or title of the collection, should be descriptive and unique, avoid acronyms.
Creator	dc:creator	2 Creator	Required	The creator (person or organisation) of the collection (include ORCID if possible).
Location	dc:identifier ('doi')	12 RelatedIdentifier	Required	Online location of the resource

Access Rights	dcterms:accessRights	16 Rights	Required	Collection access conditions. Specify one of: open, conditional or restricted .
Licence	dcterms:license	16 Rights 16.a rightsURI 16.b rightsIdentifier	Required	License conditions associated with the collection; a standard, machine understandable licence should be the default choice, e.g. Creative Commons BY-4.0.
Description	dc:description	17 Description	Required	A summary description of the collection. Provide sufficient information to enable a user to assess suitability of the data for reuse for their purpose.
Subject	dc:subject	6 Subject	Recommended	Keywords or terms to describe the topic of the resource. Include at least one ANZSRC-FOR code.
Spatial coverage	dcterms:spatial	18 GeoLocation	Recommended where relevant	The geometry for the spatial location the resource relates to.
Temporal coverage	dcterms:temporal	8 Date 8.a dateType = Collected 8.b dateInformation	Recommended where relevant	The time period during which data was collected or observations were made; or a time period that an activity or collection is linked to intellectually or thematically (for example, 1997 to 1998; the 18th century), in W3C Date/Time Format .
Publisher	dc:publisher	4 Publisher	Recommended	The entity that holds, archives, published, prints, distributes, releases, issues or produces the resource. In the case of datasets, "publish" is understood to mean making the data available to the community of researchers.
Related information	dc:relation	12 RelatedIdentifier 19 FundingReference	Recommended	Links to related resources such as publications (via DOIs), websites (via URLs), funding info (via grant IDs), provenance information (via URLs), etc