

By email: arc-researchevaluation@arc.gov.au

23 September 2026



Australian Academy of Science submission on the ARC FAIR Data Policy Consultation

The Australian Academy of Science (the Academy) welcomes the opportunity to comment on the Australian Research Council's (ARC) draft FAIR Data Policy which would apply to future National Competitive Grants Program (NCGP) schemes.

As a signatory to the FAIR (Findable, Accessible, Interoperable, Reuseable) access policy statement, the Academy has a direct interest in seeing the ARC's policy succeed in practice. Implementing the FAIR principles across Australia's research system is important to enhance the impact and longevity of research data, and support collaboration, through effective data stewardship.

This submission draws on consultation through the Academy's National Committees for Data in Science (NCDiS) whose remit spans both the advancement of the FAIR and [CARE](#) (Collective benefit, Authority to control, Responsibility, Ethics) data principles. This dual perspective has informed several of the recommendations below, particularly on managing sensitive data and Indigenous Data Sovereignty.

The Academy recommends the ARC strengthen the FAIR Data Policy by:

- Clarifying assessment mechanisms in Section 4, in particular how FAIR compliance will be measured.
- Narrowing the definition of Research Data, particularly for third-party and licensed data.
- Expanding metadata requirements to include organisational identifiers, creators versus contributors, versions, conditions of use, schemas, and controlled vocabularies.
- Providing guidance on data versioning, including for dynamic and continuously updated datasets.
- Clarifying management of sensitive data and Indigenous Data Sovereignty, including which framework takes precedence in conflicts.
- Addressing reproducibility, including how it will be supported alongside data reusability.

[Clearer assessment mechanisms for Data Management Plans are required to ensure stakeholders can meet expectations](#)

Section 4 on Policy Requirements does not provide a clear assessment mechanism beyond requiring a Data Management Plan and reporting all data outputs. This will make it difficult for stakeholders to assess whether they have met expectations. Therefore, FAIR maturity assessment criteria at different stages of the project should be considered.

There should be clear expectations around the FAIR requirements for Data Management Plans themselves. These should consider whether Data Management Plans should be archived alongside the data for which they were prepared, and whether they should be findable and accessible. Currently, there is no requirement to update the Data Management Plan as the project evolves, and if no one beyond the Administering Organisation ever assesses its quality, it could become a box-ticking exercise.

Section 4.2 requires the Data Management Plan "*prior to commencement*", but Section 5 requires the Administering Organisation to "*certify at award*" that a compliant Data Management Plan exists. Award precedes commencement, so the two deadlines are inconsistent.

The ARC Policy Consultation page states that "*the first version of the policy will focus on 'Findable and Accessible' with future updates to the policy expected to include additional measures that address 'Interoperable and Reusable'.*" However, section 4.2 states that a Data Management Plan must include "*how*

*the project will ensure the **interoperability** of all data collected and shared.*" The Data Policy should make it clearer which requirements are mandatory at commencement, and which are future maturity objectives.

The policy does not state when data must be made FAIR (e.g. at publication, at project end, or within a set period). Without a deadline the mandate seems unenforceable, and there are no embargo provisions for follow-on publications or patent filings. Compliance relies on certification at award and a final report, with no interim checkpoint, no audit, and no stated consequence for non-compliance. Datasets published after the final report, which is common, appear to escape reporting entirely.

Furthermore, no retention period is mentioned. Section 4.2 recommends that projects should consider whether data can be found both "now and in the future", which is indefinite. The policy does not state how long data and metadata must remain available, and who is responsible after the project ends. The policy does not state who will be responsible for paying for curation, repository fees, or long-term preservation. It is not clear whether data management costs are eligible grant expenditure.

The definition of Research Data should be narrowed to consider third-party and licensed data

In the draft policy, 'Research Data' is defined as *"Physical or digital data object partially or entirely **generated by, collected or accessed for, or used in** ARC funded research. These include, but are not limited to, digital and analogue data, both raw and processed, numerical scores, textual records, images and sounds used as a primary source in research, protocols, **analysis code and workflows.**"*

This scope seems to cover all research data arising from ARC-funded research. This could capture third-party and licensed data the researcher has no right to deposit or re-license, such as commercial datasets.

It would be better to clarify how the policy applies to international collaborations, multi-institution projects, and data controlled by a third party. A clearer statement would help distinguish between different types of data:

1. Data relating to the researchers who are (partially) funded by the ARC to undertake the research
2. Data that are collected as part of the ARC funded research
3. Data relating to research outputs of the ARC funded research

This policy seems to focus on Item 2, but "Metadata" could apply to Item 2 or 3.

The definition for Research Data includes *"**analysis code and workflows**"*. However, Section 4.2, Item 8 states that "The appropriate open licensing to be applied to all openly available datasets to maximise reuse such as a Creative Commons CC BY licence." This licensing requirement is unsuited to software. Code needs [OSI](#)-style licenses and software citation practice, or an explicit exclusion. This definition also captures physical objects and all raw instrument data, with no appraisal or selection principle for what has lasting value.

The Metadata requirements should be expanded

The Requirements for Metadata (Section 4.3) provide a useful list of suggested items. This should be expanded to include organisational identifiers, creators versus contributors, versions, and conditions of use. The list should also include schemas (e.g. DataCite, Dublin Core, or discipline standards) and [controlled vocabularies](#). Without these, *"high-quality metadata"* will be unenforceable and varied. This section also states that a DOI or IGSN for the research output is mandatory, but these may not be suitable for every data output type.

Although the focus of the first version of the policy is on 'Findable and Accessible' data, the policy should acknowledge that metadata can and should be the gateway to interoperability. Readers should be encouraged at least to recognise and consider the possible value from maximising the machine interoperability of descriptions of assets and context (e.g. by following recommendations from CODATA's [Cross-Domain Interoperability Framework](#) efforts).

Guidance on data versioning should be provided

There are no explicit instructions for data versioning or dynamic data in projects. Continuously updated datasets, versioning, and streaming instrument data do not fit the assumed model of one static dataset with

one DOI and one date range. A note on how projects should manage corrections, updates, deprecated datasets, and links between data versions would be beneficial.

To manage this, the ARC could publish a versioned application profile aligned with established repository standards, using controlled vocabularies and qualified relations. An example of international standards is [DataCite Metadata Schema](#).

The draft policy does not mention data provenance requirements. Data provenance is [defined as](#) “the documentation of where a piece of data comes from and the processes and methodology by which it was produced”. Providing this information would be useful for any derived or processed data.

[Clearer guidance is needed on managing sensitive data and Indigenous Data Sovereignty](#)

Sensitive data is defined as: “*information that must be protected to prevent harm to **individuals, communities, or organisations.***” This wording does not clearly extend to categories of sensitive data that Section 4.5, Item 6 itself lists as sensitive, such as biodiversity data. Species and ecosystems are not individuals or organisations, and do not sit comfortably under “communities” either. The definition should be broadened to capture these cases explicitly.

Section 4.5, Item 2 on Indigenous Data Sovereignty merits further discussion. FAIR principles should be applied in such a way that they also support Indigenous data sovereignty and CARE principles. How to manage any conflicts between these should be explicitly stated in the policy.

Section 4.3 on Requirements for Metadata mandates publicly available metadata with a rich description for all outputs, with no exception. However, Section 4.5 on Sensitive Data flags locations of endangered species and archaeological remains as sensitive. Some metadata is sensitive, and the policy should provide clear advice on how this is handled.

Section 4.2 states Data Management Plans must include an “*appropriate open access data repository*”, yet Section 4.3 on Metadata and Section 4.5 on Sensitive Data consider restricted and mediated access. The policy should require a trusted repository and clearly state the underlying principle that data should be “[as open as possible, as closed as necessary](#)” to protect privacy and security matters.

[How the reproducibility of results will be implemented in the research should be considered](#)

The ARC should consider whether the scope of the FAIR principles also includes the ‘reproducibility (of results)’, not just the reusability of data. Reproducibility is not about finding or accessing results but should be part of any ARC-related research. These issues may be beyond the scope of FAIR, but a statement discussing how reproducibility will be implemented in the research could be considered.

To discuss or clarify any aspect of this submission, please contact Lauren Sullivan, Science Policy and Advice Manager at science.policy@science.org.au.