ANDS Guide

Defining a data collection

Level: Awareness

Last updated: 11 January 2017

Web link: **www.ands.org.au/guides/defining-a-collection**

Who should read this?

This document is intended to be of interest to researchers, research data stewards and data management professionals involved in providing records of collections to Research Data Australia

What is a collection?

While any grouping of objects could be referred to as a collection, the concept can be refined by considering what kind of objects a collection might contain, and what the criteria for including them in the collection might be.

The following definition comes from the Dublin Core Collections Application Profile:

The term 'collection' can be applied to any aggregation of physical and/or digital resources. Those resources may be of any type, so examples might include aggregations of natural objects, created objects, "born-digital" items, digital surrogates of physical items, and the catalogues of such collections (as aggregations of metadata records).'

Selection criteria for inclusion in a collection are extremely variable, and in an ideal world would be based on the needs of those for whom the collection is created. Collections may be assembled based on one or more of the following criteria: location, type or form, provenance, source, ownership, audience, temporal or geographical scope, value or significance, format, intellectual theme, subject or topic, associated administrative structures, funding source, genre, author, or curation and preservation requirements. This is unlikely to be an exhaustive list.

Collections may vary in size, and may be permanent or temporary. In the digital world, groupings of resources brought together by a search in a search engine can be regarded as a temporary virtual collection that meets the

specific needs of the searcher. However, more permanent groupings can provide a context for managing, interpreting and presenting information, and assist with the discovery of resources.

What is a collection in the context of Research Data Australia?

Research Data Australia contains descriptions of collections produced by or relevant to Australian researchers. For Research Data Australia, the concept of a collection means an aggregation of physical and/or digital resources which has meaning in a research context.

This context includes the research process itself, any resources which support that process, and the linked scholarly communications cycle with its research outputs of publications and published data. Objects from these collections provide context and meaning for each other.

Research Data Australia can accommodate collection descriptions of resources based on any criteria which have meaning to the research community.

Some common groupings have been identified as follows:

- Collections of mixed objects based on a research project. The collection might include any kind of object, including experimental observations, photographs or digital images, maps, structured data in spreadsheets or databases, lab notebooks, or any other materials used in the research process. These collections should link to other entities in the Research Data Australia mesh, including parties (authors, contributors, funders, institutions) and activities (course, research project), and point to related information outside Research Data Australia, such as professional publications based on the collection.
- Example: Braided Channels Collection
- Collections of particular object types based on intellectual themes together with curatorial requirements. An example might be a collection of photographs or music with a common theme that are stored and managed as a group. (Research Data Australia collection type of 'collection'). These collections should link to other entities in Research Data Australia, including parties (creators, institutions).
- Example: Kartomi Collection of Traditional Music Arts in Sumatra
- Collections of digital data. A collection might include scientific observations in a digital format, together with information about scientific equipment and methods used to compile them. (Research Data Australia collection type of 'dataset').
- Example: Water Quality dataset, Great Barrier Reef
- Collections of digital data or physical objects based on a temporal range such as time series data. (Research Data Australia collection type of 'dataset').
- Example: Time-series of seagrass maps from 1972 to 2010, in Moreton Bay, Australia
- Collections comprised of software and related materials. A software collection may be comprised of numerous resources (source code, executables, documentation) that together constitute a software product; for example, a model or workflow that may be downloaded, compiled, executed and instantiated. (Research Data Australia collection type of 'software')
- Collections comprised of a list or arrangement of terms used in a particular context. May describe classification schemes such as controlled vocabularies, authority lists, ontologies and thesauri.
- (Research Data Australia collection type of 'classificationScheme)
- Collections of digital data or physical objects based on storage location (in association with an institution or subject discipline) (Research Data Australia collection type of 'repository').

- Example: CSIRO Data Access Portal
- Collections of resource descriptions (metadata) describing the content of one or more repositories or collective works (in association with an institution or subject discipline) (Research Data Australia collection type of 'catalogue or index').
- Example: The Australian Government's Spatial Information Gateway
- Collections of descriptions (metadata) of one or more collections, parties, activities and services (Research Data Australia collection type of 'registry').
- Example: Australian National Data Service Collections Registry, accessed through the RDA

Bringing together items into collections

Managers of digital data currently stored in repositories may wish to aggregate items in order to create a collection record for inclusion in Research Data Australia. There are many ways in which items might be aggregated, and repository managers should decide what will be most effective.

The main objective of Research Data Australia is discovery and reuse; therefore, collection criteria which further this goal, such as thematic, project, temporal or spatial collections are desirable. Criteria that meet local administrative or curatorial needs, such as department or organisational units or by author, are less desirable to Research Data Australia.

There is no single right way to create collections. Groupings which make the best use of curatorial time and effort are recommended. In particular, collections should be large enough to allow good metadata to be created and to be managed effectively. Choose a level of granularity which allows meaningful description without having a large number of objects which are both difficult to manage, and would return too many undifferentiated search results when displayed in Research Data Australia.

Research Data Australia is not intended to display collections of parties, services or activities. These kinds of objects should be described individually so that they can be linked to other registry objects appropriately. The concept of a collection applies only to resources, whether physical or digital.

Collection names and descriptions should be consistent with the assigned collection type. Collection descriptions should include descriptions of the kinds of objects in the collection and the basis of selection for objects included in the collection as well as describing what the collection is about.

Further information

- ANDS Content Providers Guide - Describing Collections
- Research Data Australia explained

Feedback?

We welcome your feedback on this guide. Please email contact@ands.org.au with any comments or questions.

About ANDS

The Australian National Data Service (ANDS) makes Australia's research data assets more valuable for researchers, research institutions and the nation.

ANDS is a partnership led by Monash University in collaboration with the Australian National University (ANU) and the Commonwealth Scientific and Industrial Research Organisation (CSIRO). It is funded by the Australian Government through the National Collaborative Research Infrastructure Strategy (NCRIS).