

# Collection

[|| Collections in the RDA Registry](#) | [Metadata for Collection records in the RDA Registry](#) | [Collection attributes](#) | [Collection relationships](#) | [What makes a good Collection record?](#) | [Software](#) | [Exemplar](#) | [Change history](#)

## Collections in the RDA Registry

In the RDA Registry and [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, software and data. Objects from these collections provide context and meaning for each other.

A collection in Research Data Australia:

- must be understood as a single aggregation of resources within its research context;
- is not comprised exclusively of documents as the output of research, although they can certainly be documents as the subject matter of research; and
- has Australian relevance, either through involvement of Australian researchers, or Australian subject matter.

Research Data Australia can accommodate collections of [research data](#) resources as defined by the [Research Data Australia Collection Development Policy](#). Generally, stand-alone publication outputs, such as theses, journal articles or books, are not within the scope of collections for the RDA Registry (although valuable as [related information](#)). However, stand-alone publications would be considered for inclusion where the published material:

- has been integrated into a collection of unpublished items
- is integral to the use and understanding of other collection materials in Research Data Australia
- is part of a collection where significant value has been added to the collection through mark-up and hyperlinks.

## Metadata for Collection records in the RDA Registry

The table below specifies the mandatory and optional elements for creating a collection record for the RDA registry. Click on the label name to see how the element should be encoded.

Provide as many optional elements as you can, and follow the guidelines in the best practice sections to maximise discovery and reuse of your data.

Label in the RDA Registry	Schema element or attribute name	Obligation	Repeatable	Definition
<a href="#">registry object</a>	<registryObject>	<b>Mandatory</b>		A wrapper element for metadata records (registry objects). It has no relationship to objects being described, but exists solely as part of the interchange infrastructure.
<a href="#">originating source</a>	<originatingSource>	<b>Mandatory</b>	N	The entity holding the managed version of the registry object metadata, as represented by a URI. The primary source of truth for the metadata record.
<a href="#">group</a>	@group	<b>Mandatory</b>	N	The organisation that is contributing the metadata record (registryObject), that is, the metadata publisher.
<a href="#">key</a>	<key>	<b>Mandatory</b>	N	A unique string that persistently identifies a metadata record within the RDA Registry.
<a href="#">collection</a>	<collection>	<b>Mandatory</b>	N	“Collection” is one of four classes of things that may be described in a metadata record.
<a href="#">collection type</a>	@type	<b>Mandatory</b>	N	The type of collection selected from a predefined list.

name/title	<name>	<b>Mandatory</b>	Y	The name or title given to a collection.
description	<description>	<b>Mandatory</b>	Y	A plain text description of a collection.
location /address	<location>	Optional	Y	For collections published online, locations will usually be <a href="#">electronic addresses (URLs)</a> . For physical collections, this may be a <a href="#">street or postal address</a> or <a href="#">spatial location</a> .
related object ( party) OR related info (party)	<relatedObject> OR <relatedInfo>	Optional	Y	A related party linked to the collection using an object key or an identifier.
related object ( activity) OR related info (activity)	<relatedObject> OR <relatedInfo>	Optional	Y	A related activity linked to the collection using an object key or an identifier.
related object ( service) OR related info (service)	<relatedObject> OR <relatedInfo>	Optional	Y	A related service linked to the collection using an object key or an identifier.
related info	<relatedInfo>	Optional	Y	Additional information related to the collection, or providing context to the collection, including quality, provenance and reuse information.
rights	<rights>	Optional	Y	A wrapper element that contains descriptions of <a href="#">rights</a> , <a href="#">licences</a> and <a href="#">access rights</a> for a collection, in both text and URI formats.
citation	<citationInfo>	Optional	Y	The preferred form for citing a collection to enable data to be referenced.
dates	<dates>	Optional	Y	Dates associated with an <a href="#">event in the lifecycle of the collection</a> .
identifier	<identifier>	Optional	Y	A sequence of characters or words that uniquely identify a collection within a particular context or the domain of a specified authority.
subject	<subject>	Optional	Y	A term, keyword, classification code or phrase representing the primary topic or topics covered by a collection.
spatial coverage	<coverage>	Optional	Y	Spatial characteristics of a collection described using coordinates or text.
temporal coverage	<coverage>	Optional	Y	Temporal characteristics of a collection described using dates or text.
date modified	@dateModified	Optional	N	The date the collection record metadata was last changed in the source system.
date accessioned	@dateAccessioned	Optional	N	The date the collection record was accessioned into its source system.

<element>; @ = attribute

## Collection attributes

### Collection Type

A Collection Type must be specified, preferably from the Collection Type vocabulary below.

Defining the Collection Type can be complex, and there are some overlaps between Types.

T y pe	Explanation	Examples
--------------	-------------	----------

catalogue	<p>Describes the content of one or more repositories or collective works usually associated with an institution or subject discipline. Usually it will consist solely of resource descriptions (or metadata) but it may also contain full-text indexes to the digital content it describes. Catalogues and indexes may themselves be described in a registry.</p>	<ul style="list-style-type: none"> <li>• <a href="#">World Bank Data Catalog</a>: a listing of available World Bank datasets, including databases, pre-formatted tables, reports and other resources.</li> </ul>
classification scheme	<p>A list or arrangement of terms used in a particular context, e.g. thesauri, ontologies. Use this type to enable discovery of classification schemes such as controlled vocabularies, authority lists, ontologies and thesauri that may be reused by others.</p>	<ul style="list-style-type: none"> <li>• <a href="#">Aquatic Sciences and Fisheries Thesaurus</a></li> <li>• <a href="#">Confederation of Open Access Repositories (COAR) Resource Type Vocabulary</a></li> </ul>
collection	<p>A collection of objects, grouped according to a shared criteria, which are stored and managed as a collective group. This may be a collection of similar object types with a common theme such as a collection of music audio files; or it may be a collection of different object types brought together around a particular topic, subject or project.</p> <p>It's worth noting that where the appropriate collection types exist, a provider may also, or instead, choose to describe components of a collection separately. For example, where a collection is comprised of a dataset and software, separate descriptions of each component of the collection would allow the the dataset and the software to be individually discoverable and citable. These separate descriptions in Research Data Australia should be connected via the relatedObject element to facilitate discovery of all related components.</p>	<ul style="list-style-type: none"> <li>• <a href="#">Shanghai Ancient Music Ensemble performing Tang Music reconstructions</a>: a collection of records of the Shanghai Ancient Music Ensemble.</li> <li>• <a href="#">Braided Channels</a>: a collection of materials about Australian women, land and history in Queensland's Channel country. The collection is comprised of oral history files, archival film, transcripts, photos and music.</li> </ul>
dataset	<p>Structured data that is an input to, or output of research. This may include scientific observations, remote sensing data, survey transcripts and photographs.</p>	<ul style="list-style-type: none"> <li>• <a href="#">Water quality dataset for the Great Barrier Reef</a></li> <li>• <a href="#">Karawatha acoustic sensor raw audio data</a></li> </ul>
registry	<p>An object that consists solely of resource descriptions or metadata records at the collection level. The records in a registry may describe catalogues, indexes, repositories, collections or software.</p>	<ul style="list-style-type: none"> <li>• <a href="#">RDA Registry</a>: accessed through <a href="#">Research Data Australia</a></li> <li>• <a href="#">Re3data</a>: a registry of research data repositories.</li> </ul>
repository	<p>A collection of digital or physical research objects sharing a managed storage location. A repository is usually associated with an institution or subject discipline. Repositories may store and provide access to datasets, software and collections.</p>	<ul style="list-style-type: none"> <li>• <a href="#">Clinical Research Data Repository</a>: contains clinical research data such as MRI brain imaging data, EEG data and neuropsychology test data.</li> <li>• <a href="#">PANGAEA</a>: a repository for earth and environmental science data</li> <li>• <a href="#">CSIRO Data Access Portal</a>: the CSIRO's institutional data repository.</li> </ul>

s of t w are	<p>One or more items that collectively represent a software product including computer instructions and associated non-executable items. Use this type for software that may be downloaded, compiled, executed and instantiated, as well as text-based models and workflows. Its scope may range from a single file to an entire code base of multiple files.</p> <p>Do not confuse the software type with:</p> <ul style="list-style-type: none"> <li>• repository: such as GitHub where software may be stored and made discoverable</li> <li>• service: the service class of registry object is used to describe a service delivered through an implemented software instance that enables users to 'do' something with data such as visualisations. Sometimes referred to as "software as a service".</li> </ul> <p>See <a href="#">Best practice software description</a> below.</p>	<ul style="list-style-type: none"> <li>• <a href="#">Edgar</a>: Australian bird species distribution now and in the future.</li> </ul>
--------------------------	---	--

Collection Type "publication" is included in the [RIF-CS vocabulary](#) but is not intended to be implemented in the RDA Registry. It has been added to the list of types to support business requirements of other systems that use RIF-CS. In the RDA Registry, contributors should continue to use [RelatedInfo](#) Type "publication" to describe a publication related to a collection.

## Date Accessioned and Date Modified (metadata) attributes

Dates that indicate the currency of a metadata record may be provided as collection attributes in a RIF-CS record, but are not displayed or searchable in Research Data Australia. The DateAccessioned attribute indicates the date that a collection was first registered in a managed environment such as a repository (the source system, not the RDA Registry). The DateModified attribute indicates the date when metadata describing a collection was last changed in the source system (not the RDA Registry). DateModified has no relation to the date of the last harvest of metadata from a data source. If a dataset is continually added to, but the metadata describing it doesn't change, there is no need to record a DateModified. On the other hand, if the underlying dataset changes its scope or nature, the metadata record describing it should change as well, and a DateModified attribute could be supplied. These dates will usually be system-generated by the source system and should be UTC and of one of the forms described in section 3.2.7 of the [W3C Schema Data Types document](#).

## Collection relationships

A collection may be described as a self-standing entity, or it may be related to other collections, activities, parties or services via the [RelatedObject](#) or [RelatedInfo](#) elements.

Hierarchical relationships, e.g. where a collection is derived from, or is part of another collection, are displayed in Research Data Australia in [browsable tree structures](#) to provide contextual information for this record and to facilitate discovery. Information on lateral relations between collections, e.g. these collections "are part of the same larger collection", "have come out of the same research activity", or "have the same primary collector", are automatically derived from the description of the hierarchical relationships, and do not need to be separately described.

The RDA Registry infers and displays bi-directional links in Research Data Australia between resources related via [RelatedObject](#). Links from collections to other related objects (collection, service, party or activity) within the same data source, will automatically generate an inferred reverse link in the RDA Registry which will display in Research Data Australia. If the related objects are from different data sources, the inferred reverse link will only be displayed if the receiving partner has opted in to allow bi-directional links. See [relationships between registry objects](#) for information on how the RDA Registry can automatically create relationships between objects, and bi-directional links between related objects.

Expand the links below to view an explanation of the relationships:

### ▼ [Related Collections...](#)

Hierarchical relations between collections may be described using the "hasPart"/"isPartOf" or "isDerivedFrom"/"hasDerivedCollection" relationships. Otherwise, collection to collection relationships may be described using "describes"/"isDescribedBy" (catalogue for, or index of, items in the related collection), or "isLocatedIn"/"isLocationFor" (repository where a related collection is held). If none of these relations are adequate, then use the generic "hasAssociationWith" together with a description to refine the relation.

### ▼ [Related Parties...](#)

Collections must be linked to at least one party through one of the following relations: "isManagedBy", "hasCollector", "isOwnedBy", "hasPrincipalInvestigator" or "isEnrichedBy". The most important of these are "hasCollector" (the party that takes credit for the collection), and "isManagedBy" (the party that is curating the collection, and can be contacted for further information). The relation type "isEnrichedBy", may be useful for aggregators, particularly of cultural collections. Use this relation type when a party's role goes beyond managing a collection to adding value to the collection, by, for example: creating linkages to relevant external sources, digitising hard-copy resources, changing the format of digital collections, indexing or providing additional search terms, or providing additional metadata to the collection.

If none of these relations are adequate, then use the generic "hasAssociationWith" together with a description to refine the relation. If multiple parties have made a substantial contribution to the collection, the collection is related to all those parties.

#### ▼ Related Services...

A collection may be supported by a service using the relationship "isAvailableThrough" (for harvest, search and syndicate), "isPresentedBy" (for report), "isProducedBy" (for create, generate and assemble), "isOperatedOnBy" (for transform), or "hasValueAddedBy" (for annotate). The URL which implements the related service in the collection's context can be recorded in the Relation child element "url".

Relating a collection to a service allows the service to be discovered and also allows the discovery of collections that are available via a particular type of service - see more at [Beyond RIF-CS: Metadata for Services and related Collections: Good Practice Guide](#).

#### ▼ Related Activities...

Collections that are the output of a related project, grant or program should link to the activity using the relationship "isOutputOf". A globally unique, persistent and resolvable identifier (PURL) exists for every grant in Research Data Australia and this identifier should always be used when describing research outputs resulting from the grant.

## What makes a good Collection record?

Ideally, collection records will include accurate, concise and authoritative descriptive content that facilitates discovery, access and reuse of the data being described. They will also connect to information about related people, projects, software and publications that give context to the data being described.

In practice, the actual content of individual collection records will depend on the type of data being described, the source of the metadata for the description (machine generated or human created), the goals and resources of the institution providing the records, and the information needs of researchers accessing the records. While there is no 'one size fits all' for collection descriptions, a 'good' collection record might:

- have a globally unique persistent **identifier** such as a DOI
- provide access to, or information about **how to access**, the data being described
- include **citation information** that clearly indicates how the data should be cited when reused
- include **licence** information that specifies how the data may be reused by others
- be connected via an identifier to **related outputs** such as publications and software that give context to the data
- be connected via an identifier or key to **people** and **projects** associated with the data to improve discovery
- be connected to **services** that can be used to access or manipulate the data
- include a **description** of how the data were created and how to interpret the data, to enable determination of the value of data, and reuse
- contain **subject** information to enhance discovery
- provide **spatial and temporal coverage** information that positions the data in space and time, and helps researchers find data that relates to a geographical area or time period of interest

See the individual elements and attributes for best practice information, and find out how to create RIF-CS [metadata with impact](#).

## Software

More specifically, when describing collections of type software, the following is recommended:

#### ▼ Best practice software description...

While there is no 'one size fits all' for software descriptions, a 'good' software record might:

- have a globally unique persistent **identifier** such as a DOI
- provide access to, or information about **how to access**, the software being described
- include **citation information** that clearly indicates how the software should be cited when reused, including the relevant version (see: [Software Citation](#) on the ANDS website)
- include **licence** information that specifies how the software may be reused by others (the Open Source Initiative publishes a list of common [open source licences](#) for software)
- be connected via an identifier to **related outputs** such as publications
- be connected via an identifier or key to data, **people** and **projects** associated with the software to improve discovery
- include a **description** of how the software was created including information about the programming language, the operating system and any other environment and dependency requirements
- contain **subject** information to enhance discovery

## Exemplar

Example collection record of type "dataset"

```
<registryObject group="The University of South Australia">
  <key>DWgy6aps4TcHxM5GMKAZuANEChYyie7M0eaXeegg89iFwVB3xbtmv</key>
  <originatingSource>https://demo.and.s.org.au/registry//orca
/register_my_data</originatingSource>
  <collection type="dataset" dateModified="2014-06-09T23:30:12Z">
    <name type="primary">
      <namePart>Surface water run-off measurements in the City of
Salisbury, South Australia, during the period June 2012 to December
2012</namePart>
    </name>
    <description type="brief">The Parafield stormwater harvesting and
Managed Aquifer Recharge (MAR) facility is operated by the City of
Salisbury in South Australia. This data collection is made up of three
datasets that record measurements of surface water run-off at three
data collection points within the study area, specifically Parafield,
Ayfield and Cobbler Creek. The data were collected during the period
June 1 to Dec 31, 2012. It is provided as three files in .csv format.
This study was supported by The Goyder Institute for Water Research.
The data were collected with support from the South Australian Water
Corporation and the City of Salisbury.</description>
    <rights>
      <rightsStatement rightsUri="http://unisa.edu.au/About-UniSA
/Governance-and-management-structure/Copyright-at-UniSA/">Copyright
2014, The University of South Australia</rightsStatement>
      <licence type="CC-BY" rightsUri="http://creativecommons.org
/licenses/by/3.0/au"/>
      <accessRights type="open" rightsUri="http://w3.unisa.edu.au
/policies/policies/resrch/res20.asp">In accordance with the UniSA Open
Access policy</accessRights>
    </rights>
    <identifier type="doi">10.4225/13/50BBFCFE08A12</identifier>
    <identifier type="uri">http://YourRepositoryID.html</identifier>
    <dates type="dc.issued">
      <date type="dateFrom" dateFormat="W3CDTF">2014-07-01</date>
    </dates>
    <location>
      <address>
        <electronic type="url" target="landingPage">
          <value>http://www.datadryad.org/resource/doi:10.4225/13
/50BBFCFE08A12</value>
        </electronic>
      </address>
    </location>
    <coverage>
      <spatial type="kmlPolyCoords">138.629130,-34.797870</spatial>
    </coverage>
    <coverage>
      <temporal>
        <date type="dateFrom" dateFormat="W3CDTF">2012-07-01</date>
        <date type="dateTo" dateFormat="W3CDTF">2012-12-31</date>
      </temporal>
    </coverage>
    <relatedObject>
```

```

    <key>Goyder-NC-1</key>
    <relation type="isOutputOf"/>
  </relatedObject>
  <relatedObject>
    <key>Contributor:University of South Australia</key>
    <relation type="isManagedBy"/>
  </relatedObject>
    <relatedObject>
      <key>http://research.unisa.edu.au/person/10107</key>
      <relation type="hasPrincipalInvestigator"/>
    </relatedObject>
  <subject type="anzsrc-for">040603</subject>
  <subject type="local">urban water</subject>
  <subject type="local">surface water</subject>
  <subject type="local">water quality monitoring</subject>
  <relatedInfo type="publication">
    <title>Surface water quality in the City of Salisbury, South
Australia</title>
    <identifier type="doi">10.4225/08/53EC60AB0DD1B</identifier>
    <relation type="isCitedBy"/>
    <notes> Myers, B., Oliver, R. & Pezzaniti, D. (2013) Surface
water quality in the City of Salisbury, South Australia. Australian
Journal of Water Quality, vol.12, no.5, pp.4-9 </notes>
  </relatedInfo>
  <citationInfo>
    <citationMetadata>
      <identifier type="doi">10.4225/13/50BBFCFE08A12</identifier>
      <title>Surface water run-off measurements in the City of
Salisbury, South Australia during the period June 2012 to December 2012<
/title>
      <publisher>University of South Australia</publisher>
      <url>http://www.datadryad.org/resource/doi:10.4225/13
/50BBFCFE08A12</url>
      <context/>
      <contributor seq="1">
        <namePart type="superior">University of South Australia<
/namePart>
      </contributor>
      <date type="publicationDate">2013</date>
    </citationMetadata>
  </citationInfo>
</collection>
</registryObject>

```

## Change history

▼ [Click here to view...](#)

Date	Change history
April 2010	Consultation draft
	First web publication

26 October 2010	
2 November 2012	Added "metadata" as a type
20 November 2012	Added dates (collections)
26 November 2013	Updated Related Information to include changes with RIF-CS v1.5.0
15 May 2014	Incorporated information about what best practice means
17 May 2017	New Collection page created replacing the "Best practice for creating collection records" and "RIF-CS in Practice: Describing a Collection" pages. Content completely revised and updated.  New table providing an overview of schema requirements for collections added to replace the Metadata Content Requirements page on the ANDS website. Providing a title and a description now mandatory in the RDA Registry.
24 Aug 2018	Added information on best practice software description