

RIF-CS Collection Types: expanded definitions

Background

In agreeing to the proposal to allow additional resource types to be described using RIF-CS, the RIF-CS Advisory Board (RAB) asked that ANDS provide greater clarity around the use of all collection types and provide examples.

The first five in the list of eight types below are existing collection types. The first four types ('catalogueOrIndex', 'collection', 'registry', 'repository'), and their definitions are taken directly from the ISO 2146 standard. These types and definitions remain unchanged to ensure compliance with the standard.

The 'dataset' type and definition were introduced to support business requirements for the ANDS Registry. To ensure consistency and continuity, this type and definition, also remain unchanged.

The final three collection types listed below have been recently approved by the RAB to support the business requirements of ANDS and other users of RIF-CS. They will be implemented with version 0.1.7 of the schema.

Explanatory text and examples have been provided for all collection types, noting that exemplar records in RDA do not yet exist for all types, in particular, the three newly approved types. The information provided here is intended to assist providers to answer the question "which collection type should I choose?"

Collection types and definitions

- **catalogueOrIndex:** collection of resource descriptions describing the content of one or more repositories or collective works
- **collection:** compiled content created as separate and independent works and assembled into a collective whole for distribution and use
- **registry:** collection of registry objects compiled to support the business of a given community
- **repository:** collection of physical or digital objects compiled for information and documentation purposes and/or for storage and safekeeping[RG(WC1)]
- **dataset:** collection of physical or digital objects generated by research activities
- **software:** one or more items that collectively represent a software product; including computer instructions (ranging in format from machine code to high-level programming languages); and associated non-executable items such as documentation.
- **classificationScheme:** a list or arrangement of terms used in a particular context eg. ontologies, thesauri
- **publication:** scholarly material consisting mainly of written text eg journal publications, book chapters

Definitions with explanatory text and examples

- **catalogueOrIndex:** a collection of resource descriptions describing the content of one or more repositories or collective works

This type should be used to describe the content of one or more repositories or collective works usually associated with an institution or subject discipline. A catalogue or index may be local, regional, national or international. Usually it will consist solely of resource descriptions but it may also contain full-text indexes to the digital content that it describes.

An example of this type is:

- [World Bank Data Catalog](#): a listing of available World Bank datasets, including databases, pre-formatted tables, reports, and other resources.

Catalogues and indexes may themselves be described in a registry.

- **collection:** compiled content created as separate and independent works and assembled into a collective whole for distribution and use

This type should be used to describe 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. See for example in Research Data Australia:

- [Shanghai Ancient Music Ensemble performing Tang Music reconstructions](#): a collection of recordings of the Shanghai Ancient Music Ensemble.

It may also be a collection of different object types brought together around a particular topic, subject or project. See for example in Research Data Australia:

- [Braided Channels](#): a collection of materials about Australia women, land and history in Queensland's Channel country. The collection is comprised of oral history files, archival film, transcripts, photos and music.

It should be noted 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 would enable each component of the collection (ie the dataset and the software) to be individually discoverable and citable. These separate descriptions should be connected via the relatedObject element to facilitate discovery of all related components.

- **registry**: a collection of registry objects compiled to support the business of a given community.

This type should be used to describe a collection that consists solely of resource descriptions or metadata records. These records may describe catalogues, indexes, repositories or collections.

Examples include:

- ANDS Registry: accessed through [Research Data Australia](#)
 - [Re3data](#): a registry of research data repositories
-

- **repository**: a collection of physical or digital objects compiled for information and documentation purposes and/or for storage and safekeeping.

This type should be used to describe a collection of digital or physical research objects sharing a managed storage location. A repository is usually associated with an institution or subject discipline.

An example of this type in Research Data Australia is:

- [Clinical Research Data Repository](#): contains clinical research data such as MRI brain imaging data, EEG data and neuropsychology test data

Other resources that could be described as collection type='repository' include:

- [The CSIRO Data Access Portal](#): the CSIRO's institutional data repository
- [PANGAEA](#): a repository for earth and environmental data

Repositories may store and provide access to other collection types such as *datasets*, *source code*, *collections and publications*.

- **dataset**: a collection of physical or digital objects generated by research activities

This type should be used to describe structured data that is an input to, or output of research. This may include scientific observations, remote sensing data, survey transcripts and photographs.

Examples of this type in Research Data Australia include:

- [Water quality dataset for the Great Barrier Reef](#)
 - [Karawatha acoustic sensor raw audio data](#)
-

- **software**: one or more items that collectively represent a software product; including computer instructions (ranging in format from machine code to high-level programming languages); and associated non-executable items such as documentation.

This type should be used to describe software that supports research, including models and workflows. Use this type to publish and make discoverable software that may be downloaded, compiled, executed and instantiated. The aim is to enable such software to be reused and cited by others. Its scope may range from a single source file to an entire code base of multiple files.

Do not confuse with:

- *repository*: such as GitHub where sourceCode may be stored and made discoverable.
- *service*: the service registry object is used to describe a service delivered through a software instance that enables users to 'do' something with data such as visualisations. This is sometimes referred "software as a service".

The 'Edgar' web application run out of James Cook University can provide an example of this distinction.

This [Service record](#) in RDA describes the Edgar service that can be used to generate environmental suitability maps for Australian bird species under various climate change scenarios.

With the introduction of collection type='sourceCode', a collection record of type='sourceCode' could also be published in RDA to describe the code underpinning the Edgar service and provide access to the code in GitHub via the location element. See sample encoding below.

Class:	Collection												
Type:	software												
Key:	CuFR4hgtJLgDgjjtr56nAevoM8nhKDBcY6TSe0coBifRhr01JWXWm												
Source:	ANDS Training												
Originating Source:	https://demo.ands.org.au/registry/orca/register_my_data												
Group:	Demonstration University												
Names:	<table border="1"> <tr> <td>type:</td> <td>primary</td> </tr> <tr> <td>Name Part:</td> <td>Value: Edgar: climate change impact on Australian bird species distribution</td> </tr> </table>	type:	primary	Name Part:	Value: Edgar: climate change impact on Australian bird species distribution								
type:	primary												
Name Part:	Value: Edgar: climate change impact on Australian bird species distribution												
Location:	<table border="1"> <tr> <td>location:</td> <td>address:</td> <td>electronic:</td> <td> <table border="1"> <tr> <td>type:</td> <td>url</td> </tr> <tr> <td>target:</td> <td>landingPage</td> </tr> <tr> <td>value:</td> <td>https://github.com/jcu-eresearch/Edgar</td> </tr> <tr> <td>title:</td> <td>Edgar code base</td> </tr> </table> </td> </tr> </table>	location:	address:	electronic:	<table border="1"> <tr> <td>type:</td> <td>url</td> </tr> <tr> <td>target:</td> <td>landingPage</td> </tr> <tr> <td>value:</td> <td>https://github.com/jcu-eresearch/Edgar</td> </tr> <tr> <td>title:</td> <td>Edgar code base</td> </tr> </table>	type:	url	target:	landingPage	value:	https://github.com/jcu-eresearch/Edgar	title:	Edgar code base
location:	address:	electronic:	<table border="1"> <tr> <td>type:</td> <td>url</td> </tr> <tr> <td>target:</td> <td>landingPage</td> </tr> <tr> <td>value:</td> <td>https://github.com/jcu-eresearch/Edgar</td> </tr> <tr> <td>title:</td> <td>Edgar code base</td> </tr> </table>	type:	url	target:	landingPage	value:	https://github.com/jcu-eresearch/Edgar	title:	Edgar code base		
type:	url												
target:	landingPage												
value:	https://github.com/jcu-eresearch/Edgar												
title:	Edgar code base												
Related Objects:	<table border="1"> <tr> <td>Edgar: Climate Change Impact on the Distributions of Australian Bird Species (service)</td> </tr> <tr> <td>key: jcu.edu.au/dh/service/australian-birds</td> </tr> <tr> <td>relation: type: supports</td> </tr> <tr> <td>James Cook University (party)</td> </tr> <tr> <td>key: http://hia.gov.au/hia party-642603</td> </tr> <tr> <td>relation: type: isOwnedBy</td> </tr> <tr> <td>Jeremy James Vanderwal (party)</td> </tr> <tr> <td>key: http://hia.gov.au/hia party-1507093</td> </tr> <tr> <td>relation: type: hasPrincipalInvestigator</td> </tr> <tr> <td>Title: Demonstration University (contributor)</td> </tr> <tr> <td>Key: Contributor: Demonstration University</td> </tr> <tr> <td>Relation: type: (Automatically generated contributor page link)</td> </tr> </table>	Edgar: Climate Change Impact on the Distributions of Australian Bird Species (service)	key: jcu.edu.au/dh/service/australian-birds	relation: type: supports	James Cook University (party)	key: http://hia.gov.au/hia party-642603	relation: type: isOwnedBy	Jeremy James Vanderwal (party)	key: http://hia.gov.au/hia party-1507093	relation: type: hasPrincipalInvestigator	Title: Demonstration University (contributor)	Key: Contributor: Demonstration University	Relation: type: (Automatically generated contributor page link)
Edgar: Climate Change Impact on the Distributions of Australian Bird Species (service)													
key: jcu.edu.au/dh/service/australian-birds													
relation: type: supports													
James Cook University (party)													
key: http://hia.gov.au/hia party-642603													
relation: type: isOwnedBy													
Jeremy James Vanderwal (party)													
key: http://hia.gov.au/hia party-1507093													
relation: type: hasPrincipalInvestigator													
Title: Demonstration University (contributor)													
Key: Contributor: Demonstration University													
Relation: type: (Automatically generated contributor page link)													
Description:	<table border="1"> <tr> <td>type:</td> <td>brief</td> </tr> <tr> <td>Value:</td> <td> <p><p>Edgar is a website where visitors can explore the future impact of climate change on Australian birds. </p> <p>The Edgar source code is well suited to other projects that want to display and vet species occurrences, and show projections of future climate change and its effects. </p> <p>There are a few parts to Edgar: </p> An importing section, mostly written in Python A modeling section, also mostly in Python A mapping section using MapServer A web application section using CakePHP <p>Full documentation, installation instructions and license information are all available on GitHub</p> <p></p></p></td> </tr> </table>	type:	brief	Value:	<p><p>Edgar is a website where visitors can explore the future impact of climate change on Australian birds. </p> <p>The Edgar source code is well suited to other projects that want to display and vet species occurrences, and show projections of future climate change and its effects. </p> <p>There are a few parts to Edgar: </p> An importing section, mostly written in Python A modeling section, also mostly in Python A mapping section using MapServer A web application section using CakePHP <p>Full documentation, installation instructions and license information are all available on GitHub</p> <p></p></p>								
type:	brief												
Value:	<p><p>Edgar is a website where visitors can explore the future impact of climate change on Australian birds. </p> <p>The Edgar source code is well suited to other projects that want to display and vet species occurrences, and show projections of future climate change and its effects. </p> <p>There are a few parts to Edgar: </p> An importing section, mostly written in Python A modeling section, also mostly in Python A mapping section using MapServer A web application section using CakePHP <p>Full documentation, installation instructions and license information are all available on GitHub</p> <p></p></p>												

- **classificationScheme:** a list or arrangement of terms used in a particular context eg. ontologies, thesauri

This type should be used to describe a research asset that is comprised of terms used in a particular context. Use this type to publish and make discoverable classification schemes such as controlled vocabularies, authority lists, ontologies and thesauri that may be reused by others. Examples include:

- [Aquatic Sciences and Fisheries Thesaurus](#)

- [Confederation of Open Access Repositories \(COAR\) Resource Type Vocabulary](#)
-

- **publication:** scholarly material consisting mainly of written text eg journal publications, book chapters.

It is not intended that collection type="publication" be implemented in the ANDS Registry. It has been added to the list of types to support business requirements of other systems that use RIF-CS. In the ANDS Registry, providers should continue to use relatedInfo type="publication" to describe a publication related to a collection.

Those organisations intending to implement type="publication", may wish to consider whether sub-types such as 'journal article', or 'conference paper' are required. These should be implemented in accordance with a controlled list of terms such as the [Confederation of Open Access Repositories \(COAR\) Resource Type Vocabulary](#)
