Metadata for Impact: make RIF-CS work for you

Creating metadata descriptions involves some effort, so how should you decide which optional elements to include in your data descriptions? A good way to think about this is to consider what your institution wants to achieve by publishing data descriptions via Research Data Australia and how you expect people will search for and reuse your data. Richer metadata contains detailed and meaningful names, subject keywords, full descriptions, temporal and spatial coverage, citation information, rights information and meaningful relations that add information and context to the metadata document and support discovery and reuse. Contextual information such as information about the research program/project, data collection methods, researcher, or institution, helps a researcher decide if they want to reuse the data. Information about access such as access conditions and terms of use, restrictions on access, or contact information, enables the researcher to get to the data.

Some common institutional goals, with examples of associated RIF-CS encoding, are provided below:

For more information about maximising the impact of your data:

- Read the Best Practice advice for the RIF-CS elements in the Research Data Australia Content Providers Guide
- Contact your Outreach Officer or email services@ardc.edu.au for assistance or to discuss your requirements.

1. We want to highlight our open data - the transparency of our research is important to our reputation

As open data is citable and reusable it brings many benefits to individuals, institutions and the broader research community. The 'openness' of data is defined by several characteristics, so describing this requires more than one element. Providing content for these elements will not only provide users with clarity around access and reuse conditions, it will also ensure the record is prominently displayed in Research Data Australia.

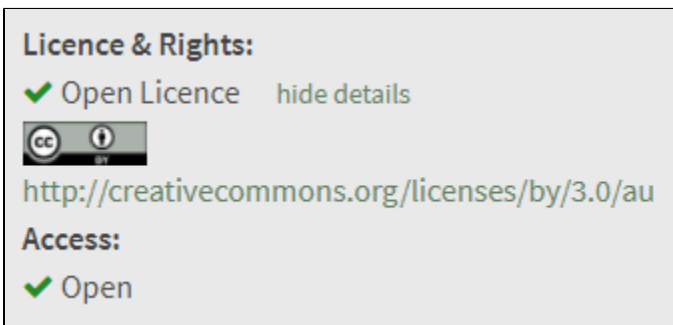To highlight open data, provide content for these RIF-CS elements:

- Rights:AccessRights - specify an Access Rights Type of **open** to indicate that the data is publicly accessible online, and to ensure that the record is returned when Research Data Australia users limit their search to open data.
- Rights:Licence - specify an open Licence Type such as **CC-BY**. If no licence is assigned, reuse of the data is not permitted, so be sure to assign an appropriate licence and describe it in this element.
- Location:Address : Electronic - specify an Electronic Address Type of **url**, and provide the URI to download the data in the Value element. The link may trigger a direct download of the data (specify an Electronic Target of **directDownload)** or link to a metadata landing page from which the data may be downloaded (specify an Electronic Target of **landingPage**). You can also provide additional information such as Title, MediaType, ByteSize and Notes.

A defining characteristic of open data is that it is well described, so include as much rich information as possible in collection records to describe what the data is, how it was collected and for what purpose.

Example 1.1 xml encoding of the Rights element incorporating Licence and AccessRights information

```
<rights>
        <licence type="CC-BY" rightsUri="http://creativecommons.org
/licenses/by/3.0/au"></licence>
        <accessRights type="open"></accessRights>
</rights>
```

Example 1.1 Research Data Australia display of Licence and AccessRights

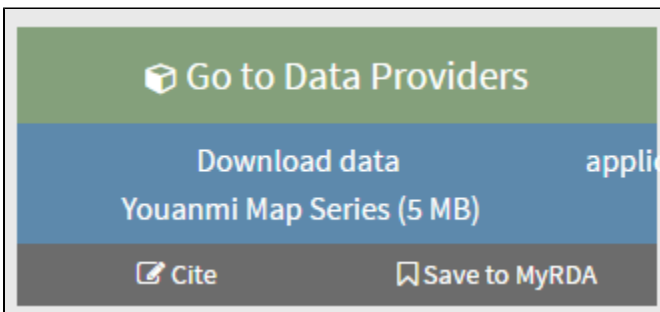**Example 1.2 xml encoding of the Location element incorporating Electronic Address**

```
<location>
        <address>
                    <electronic type="url" target="directDownload">
                            <value>http://www.ga.gov.au/corporate_data/64739
    /64739_sh50-04_kml.zip</value>
                            <title>Youanmi Map Series</title>
                            <mediaType>application/zip/kml</mediaType>
                            <byteSize>5 MB</byteSize>
                    </electronic>
        </address>
</location>
```

Example 1.2 Research Data Australia display of Electronic Address (direct download)

2. We want citation metrics for our data - as we have for our publications - to demonstrate the impact of our research and how we might identify new collaborators

Collection records can be harvested from a data source in Research Data Australia to the Clarivate Analytics (formerly Thompson Reuters) Data Citation Index using this process. This will enable metrics for citation of data to be counted in the same way as for publications. As similar indices emerge, we will seek to ensure records in Research Data Australia are included.

To enable citation metrics for data collections, provide content for these RIF-CS elements:

- CitationInfo:CitationMetadata - enables you to describe how the data should be cited, in a structured, machine-readable way; that is readily identifiable by citation indexing services.
- Identifier - if possible, assign a DOI to your data using a DOI-minting service. DOIs uniquely identify a dataset and are considered best practice for the accurate capture of citation metrics.

If you'd like the records in your data source to be harvested to the Data Citation Index take a look at the DCI encoding notes to ensure your records are ready to be harvested and can start accruing metrics. See also 3. Linking data to related publications.

Example 2.1 xml encoding of CitationMetadata, with DOI as the Identifier

```
<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>
            <version>1</version>
            <publisher>The University of South Australia</publisher>
            <contributor seq="1">
                    <namePart type="family">Oliver</namePart>
                    <namePart type="given">R</namePart>
            </contributor>
            <contributor seq="2">
                        <namePart type="family">Myers</namePart>
                    <namePart type="given">B</namePart>
                 </contributor>
            <date type="publicationDate">2013</date>
        </citationMetadata>
    </citationInfo>
```

Example 2.1 Research Data Australia display of the citation

3. We want to link our published data to related publications - it may help drive up the citation count for our publications

While it is not possible to create a registry object (or separate RIF-CS record) to describe a publication in Research Data Australia, it is possible to provide information about a publication using the RelatedInfo element in the relevant collection record. Providing this information can bring multiple benefits including:

1. providing rich contextual information to support data reuse.
2. enabling the linking of data records in the Data Citation Index to related publications in the Web of Science citation indices, providing enhanced discovery of both data and publications.
3. increased citations for the related publication.

To link data collections to related publications, provide content for this RIF-CS element:

- RelatedInfo:type="publication" - enables you to describe a publication associated with a data collection. Where possible, include the DOI for the publication and provide the full citation as a note.

Example 3.1 xml encoding of RelatedInfo to link a data collection to a related publication

```
<relatedInfo type="publication">
        <title>Preconception risk factors and SGA babies: Papilloma virus,
omega 3 and fat soluble vitamin deficiencies</title>
        <identifier type="doi">https://doi.org/10.1016/j.earlhumdev.
2011.06.002</identifier>
        <relation type="isCitedBy"/>
</relatedInfo>
```

Example 3.1 Research Data Australia display of the related publication

Related Publications

📕 Is cited by Preconception risk factors and SGA babies: Papilloma virus, omega 3 and fat soluble vitamin deficiencies
doi : https://doi.org/10.1016/j.earlhumdev.2011.06.002

4. We want people to know they can access and use our data via a service

The service may enable a user to download the data being described, or it may allow a user to "do something" with the data, e.g. create a visualisation of or analyse the data. The ARDC, in consultation with the community, has developed a schema-agnostic, best practice guide for service metadata, primarily for enhancing machine-to-machine discovery. Refer to the Metadata for Services and Related Collections: Best Practice Guide for further information.

Two options exist to describe a service related to a data collection in the RDA Registry:

1. Create a separate registry object (or RIF-CS record) to describe the service and link that record to the relevant collection record(s) using the RelatedObject element. Where possible, include a direct link to the data described in the Relation URL element. This option will give greater exposure to the data service by ensuring it is readily discoverable in Research Data Australia. This option is preferred where the description is being provided by the service owner or maintainer. The Service Discovery tool in the RDA Registry may also be used to auto-generate RIF-CS service records from OGC service URLs in RIF-CS collection records.
2. Provide information about the service within a collection record using the RelatedInfo element. Include a URI for the service in the Identifier element and where possible, a direct link to the data described in the Relation URL element. This option provides sufficient information for a user to access a collection via a service without providing a description of the service itself. Use this option where option 1 is not feasible or the service is owned and maintained by another organisation.

Example 4.1 xml encoding of RelatedInfo to link a data collection to a related service (option 2 above)

```
<relatedInfo type="service">
        <title>Marine Virtual Laboratory Information System</title>
        <identifier type="uri">http://marvlis.aodn.org.au/marvlis<
/identifier>
        <relation type="supports">
                <url>http://marvlis.aodn.org.au/marvlis/ACQ_SurfPlt
/MAPWaterTemps.png</url>
        </relation>
        <notes>Data visualisation</notes>
</relatedInfo>
```

Example 4.1 Research Data Australia display of the related service

Related Services

⚗ Supports Marine Virtual Laboratory Information System

5. We want to optimise our data descriptions for display in Google Dataset Search

There is widespread and growing use of structured metadata by web search engines, such as Google Dataset Search. To help the discoverability of metadata harvested into Research Data Australia, Schema.org metadata has been added to all Collection and Service records.The RIF-CS elements which are mapped to Schema.org and display in Google Dataset Search (beta) are:

- Name/title
- Description
- SpatialCoverage (some issues with this currently, but should be resolved in time)
- TemporalCoverage (some issues with date ranges currently, but should be resolved in time)
- Identifier (only DOIs display currently)
- Licence
- Related parties (author, creator, contributor, principal investigator, coinvestigator, collector, owner, manager)
- Collection dates

If you include these elements in the RIF-CS collection records you provide to Research Data Australia, then you will maximise the discoverability and display of your records in Google Dataset Search.
(Note: this advice may change over time).

Like to know more? Refer to our mapping of RIF-CS elements to the Schema.org metadata standard used by Google Dataset Search.