Metadata: A Case Study at Western Sydney University

Assessment of metadata schema for active research data management

Western Sydney University

January 2023

Juan Cooper, Andrew Leahy, Mike Kmiec and Tracy Donelly

URL: https://doi.org/10.26183/t8x5-hf62

Overview

Western Sydney University has engaged Cloud Services provider Intersect to provide access to Mediaflux[1] on an Intersect research data management platform called Idea[2]. Idea has been designed with the intention of managing research data throughout the full data lifecycle from collection to archiving. WSU-derived research data are currently stored across many different silos and there is consideration of management of data through research data management plans. However, the use of metadata is not consistent.

As part of the Australian Research Data Commons (ARDC) National Data Assets Initiative, the ARDC funded a collaborative multi-university Institutional Underpinnings[3] project to develop a national Institutional Research Data Management Framework (the Framework). Several essential Elements to the Framework were developed, one of which is the Active Research Data Management Element[4]. The intention of this Element is to provide "*institutions with guidance to ensure that research practice is efficient and impactful, and that research data is managed according to requirements such as those outlined in The Australian Code for the Responsible Conduct of Research*".

Our approach at Western Sydney University in implementation of the Active Research Data Management Element of the Framework was to take the turnkey Idea platform, configure using a default metadata schema and engage with researchers in ingesting live research project datasets. This case study evaluates the metadata schema to reflect active research datasets uploaded directly from a data source to the Idea research data repository.

The Idea platform provides the ability for researchers to curate, manage, protect and disseminate research data. It will also provide the capacity for handling data from research instrument platforms as increasingly, the University's research infrastructure requires a component that automates metadata creation and storage for aspects of the research cycle.

Idea in the context of the researcher

Western Sydney University policy requires data arising from University-funded research projects to be described (and storage location noted) in the research data management plan provided on the ReDBox[5] platform. The generated metadata in ReDBox enables discoverability, indexing, and re-use of research data, which subsequently increases researcher profile, collaboration, institutional reputation, and ranking.

Framework model

University research is often conducted across multidisciplinary teams and locations, which can create data silos, security issues and oversight difficulties.

[1] Arcitecta Mediaflux
[2] Intersect.data.edu.au (Idea)
[3] ARDC Institutional Underpinnings Project
[4] ARDC Institutional Underpinnings Element: Active Research Data Management
[5] ReDBox Research Data Box

Idea uploads files to a unified, secure location that provides secure cloud storage and the ability to track active research data throughout the research life cycle. It automatically indexes content and applies metadata to data files to improve searchability and enhance collaborations.

Idea applies a consistent object model to distinguish between storage and workflow by managing the raw and processed data through detailed and consistent descriptions of:

- Representation of specific subject of study
- Representation of experimental method via a workflow
- Interdisciplinary research teams
- Access control
- Data types and storage

The object model also processes the protocols for research and includes how the data are acquired, relationships between data, and how the data can be distributed, analysed and interpreted.

Framework for discovery

The relationships between data, distribution and analysis will influence the search interface, providing a domain-specific layer that has metadata elements that reflect a researcher's search query. The elements contain:

- Framework: fixed metadata for object names and descriptions
- Domain: Configured specifically for domain, discipline, or specific package and that is relevant to a specific project

The flexibility to add and assign different metadata that is discipline-specific when uploading a project is needed due to the multitude of existing and evolving discipline-specific metadata schemas. This situation prevents a one-size-fits-all approach with research metadata and increases the complexity of discovery.

Idea and storage

Research institutions require large, flexible storage solutions internally and externally due to the volume of research data that is generated. Larger storage capacity, however, does not overcome all challenges for researchers in use and management of research data. Instead, it highlights the need to organise the data in the form of a framework that integrates an object model incorporating research methods, workflows, experimental data sets and the provenance of the data.

Metadata schema

Metadata enables the discoverability, interoperability and reusability of research data that is integral to any institution's successful adoption of a research data framework. The metadata schema should be defined by using existing standards and prescribe a minimum set of metadata elements for each object.

The metadata schema should also capture and track the metadata at all the stages of the research life cycle, containing relevant administrative metadata – things like how collections of data are organised, related services, parties, and activities. It provides the means to demonstrate the relationships between all entities.

The specific metadata used, and the semantic interpretation of it can allow natural language searching, discovery through associated ontologies or full text searching.

Active data management at Western Sydney University is supported by the Library, Research Services and ITDS. These three professional support units combine diverse skills and knowledge to assist in identifying, describing, and enhancing discoverability of research data without duplicating effort or processing.

Institutional data governance is determined by roles and responsibilities with the key data management roles identified such as data owner, data custodian and data user. The typical searching behaviour such as free-text search or by specific fields (e.g., 'field of research' or 'license') highlights the importance of metadata associated with research data and discovery which could potentially increase research collaborations or reporting to funding agencies.

The Idea platform has a default metadata schema, based on Dublin Core[6], that specifies all metadata terms, including properties, vocabulary encoding schemes, syntax encoding schemes, and classes. It supports innovation in metadata design and applies a best practice approach across all metadata schemas as many repositories are based on XML (Extensible Markup Language) as it specifies the content of records as structured documents

Dublin Core is considered a generic metadata schema for many kinds of digital objects. However, this case study considers the metadata schema requirements to ensure that the descriptive metadata of research data is fit for purpose, based on common standards, and provides a level of futureproofing for emerging digital formats.

This case study recommends using the Registry Interchange Format – Collections and Services[7] (RIF-CS) for metadata schema. This XML-based schema is used as the underlying metadata format for both Research Data Australia (RDA) and the Australian Research Data Commons (ARDC).

RIF-CS is a layered and complex metadata schema covering the entirety of the research lifecycle from creation through preservation. It uses an 'object-oriented, relational model' which fits well with the Idea method of organisation, describing the relationship between the various research components.

Whilst the RDA Registry demonstrates flexibility and can accept other formats of metadata using metadata crosswalks to ingest the information, we feel that using RIF-CS allows more flexibility with research metadata ingest into other systems or services.

Assessment of on-boarding real-life research datasets

[6] The Dublin Core Metadata Initiative

[7] Registry Interchange Format – Collections and Services

Idea can ingest, analyse, package, quality-assure and store data, as well as automatically extract metadata for various data types, thus minimising the manual entry of metadata.

To best enable this, the University will first need to identify what elements of RIF-CS are mandatory/required for all University research datasets, regardless of their discipline. In this way, mapping of automated metadata collection within Idea becomes easier. Longer term, more specific metadata elements – either content-related or administrative – may be activated within Idea. This allows scalability rather than trying to address all aspects of research metadata at once.

The RIF-CS schema was compared with the fields from the University's Research Data Management Plan (RDMP) platform, ReDBox and [8]DataCite. The RDMP describes the data a researcher will acquire or generate during a research project and how it will be managed and stored. As the project approaches completion and archiving, the data publication process records licensing information, permanent storage location, and keywords, issuing the dataset with a persistent digital object identifier (DOI) for discovery in Research Data Australia.

The key information activities collected in ReDBox:

- Creating a new research data management plan
- Creating an archival data record
- Creating a data publication

The working party compared the 'creating a new research data management plan' and 'creating a data publication' metadata aspects with those of the default Idea metadata (Appendix 1). This was done to demonstrate how capturing quality metadata in a standardised way using controlled vocabularies would benefit the University.

Summarised learnings

- Idea allows for citable identifiers for all objects ingested, making tracking for impact and engagement possible (if data are made public)
- Embedding of new approaches to research data workflows requires appropriate technical infrastructure at the University and national levels, also requiring training, support, consultation, and engagement across the organisation
- At the discipline level, creation of 'champion' roles will help with adoption of any change to systems, workflows, or practice. This, in addition to greater automation will improve the quantity and quality of research data capture.
- Research Master, the University's system for managing research activity and grant processes may be used as a primary source of pre-existing authoritative metadata to describe research grants ('Activities' in RIF-CS) and researchers ('Parties' in RIF-CS) as this may improve coordination and enable better use of existing resources
- The RIF-CS schema system uses a 'trigger' field in the Collection Object to flag that the record has been approved for publishing to Research Data Australia. This approval trigger of the Collection Object automatically exports the associated Party and Activity records

[8] DataCite

- All records are exported in RIF-CS format with the appropriate relationships encoded. See Appendix 2 for an example.

Recommendations

1. We have determined that RIF-CS format based on ISO 2146:2010 is the preferred metadata schema for active research data.
2. The University should determine and document which aspects of the RIF-CS schema will be mandatory when describing research data. This will ensure that new datasets meet a common standard of descriptive practice and allow for mapping of existing metadata into a unified schema.
3. The University should then use the RIF-CS framework to determine the more specific aspects needed to describe University research, such as discipline-based metadata.
4. The University should consider the adoption of a method or system to enable the automated exchange of information between systems, such as Idea, for example, and research data management planning tools, so all data may be captured as part of the research lifecycle with a reduction of manual administrative tasks.

Appendix 1

The metadata schemas were compared against the current instance of the University's research data management plan platform, ReDBox 2.0.

Each form field that is presented to the researcher contains metadata in the following areas:

- Creating a new research data management plan
- Creating an archival data record
- Creating a data publication

Data contained in these fields transforms the descriptive elements (i.e., technical metadata generated by research equipment, or otherwise) into metadata that aligns to a metadata schema for discoverability.

Each field containing descriptive metadata within the 'creating a new research data management plan' and 'creating an archival data record' areas in ReDBox have been compared against the following metadata schemas for suitability and to ensure new emerging file formats are captured by the chosen metadata schema:

- Idea default metadata schema (Dublin Core)
- RIF-CS schema
- DataCite schema

Research Data Management Plan Form: Metadata schema comparison chart

Key Sections are:

- Project overview
- People
- Data and software
- Data storage
- Data retention
- Ethics and sensitives
- Ownership, licensing, and IP

Project Overview

	Dublin Core		RIF-CS format	DataCite
ReDBox Fields	**Idea:DCMI**	**Idea:Collection**		
Project name		name [0..1] string Name/title of the collection NLA-ID [0..1] string Globally unique National Library of Australia Identifier of the owner	<registryObjects xmlns="http://ands.org.au/standards/rif-cs/registryObjects"> <registryObject group="">{0,unbounded}</registryObject> </registryObjects>	<xs:element name="identifier">
Project ID			Not mapped	<xs:attribute name="subjectScheme" use="optional"/>
Project description	description [1..1] string Collection description	description [0..1] string A description of the collection (can be as short or as long as necessary)	<xsd:element name="description" type="descriptionType" minOccurs="0" maxOccurs="unbounded"> <xsd:annotation> <xsd:documentation xml:lang="en">A textual description or URI resolving to a description relevant to the collection.</xsd:documentation> </xsd:annotation> </xsd:element>	<xs:documentation>Description of the relationship of the resource being registered (A) and the related resource (B).</xs:documentation>

Start date	date-from [0..1] date Date From	date-created [0..1] string null	<xsd:element name="startDate" minOccurs="0" maxOccurs="1"> <xsd:complexType> <xsd:simpleContent> <xsd:extension base="xsd:string"> <xsd:attribute name="dateFormat" use="required" type="xsd:string"> <xsd:annotation> <xsd:documentation>The format of the date element.</xsd:documentation> </xsd:annotation> </xsd:attribute> </xsd:extension> </xsd:simpleContent> </xsd:complexType> </xsd:element>	<xs:include schemaLocation="include/datacite- dateType-v4.xsd"/>
End date	date-to [0..1] date Date To		</xsd:element> <xsd:element name="endDate" minOccurs="0" maxOccurs="1"> <xsd:complexType>	<xs:element name="publicationYear"> <xs:documentation>Year when the data is made publicly available. If an embargo period has been in effect, use the date when the embargo period ends.</xs:documentation> <xs:documentation>In the case of datasets, "publish" is understood to mean making the data available on a specific date to the community of researchers. If there is no standard publication year value, use the date that would be preferred from a citation perspective.</xs:documentation> <xs:documentation>YYYY</xs:documentation> </xs:annotation> <xs:simpleType> <xs:restriction base="yearType"/> </xs:simpleType>

Funding source		funding_source [0..∞] string ARC/NHMRC funded project reference	Activity Relation Type (activity) hasAssociationWith: has an unspecified relationship with the related registry object hasOutput: delivers materials in the related collection hasPart: contains the related activity hasParticipant: is undertaken by the related party hasPrincipalInvestigator: is researched by the related party isFundedBy: receives monetary or in-kind aid from the related program isManagedBy: is organised and/or delivered by the related party isOwnedBy: legally belongs to the related party isPartOf: is contained in the related activities	<xs:documentation>Name of the funding provider.</xs:documentation>
Grant number(s)				<xs:element name="funderIdentifier" minOccurs="0"> <xs:element name="funderName" minOccurs="1" maxOccurs="1"> <xs:documentation>Information about financial support (funding) for the resource being registered.</xs:documentation> </xs:annotation> <xs:element name="fundingReferences" minOccurs="0">

People

	Dublin Core		RIF-CS format	DataCite
ReDBox Fields	**Idea:DCMI**	**Idea:Collection**		
Principal Investigator	creator [0..1] string The person primarily responsible for the creation of the data set	name [0..1] string Name/title of the collection	<xsd:element name="identifier" type="identifierType" minOccurs="1" maxOccurs="1"/> Primary: official name of the collection, activity, service or party (person or group)	<xs:documentation>The main researchers involved working on the data, or the authors of the publication in priority order. May be a corporate/institutional or personal name.</xs:documentation> <xs:documentation>Format: Family, Given.</xs:documentation> <xs:documentation>Personal names can be further specified using givenName and familyName.</xs:documentation> </xs:annotation> <xs:complexType> <xs:sequence> <xs:element name="creatorName">
Data Creator		role [0..1] enumeration AN optional role to be added when there are two or more owners, and these owners have a student/supervisor relationship. HDR Student Primary Supervisor Supervisor	<citationMetadata> Element Wrapper element containing metadata elements that describe a citation. Contained within: citationInfo May contain: identifier citationMetadata | contributor | title citationMetadata | version | placePublished | date citationMetadata | url citationMetadata | context	<xs:element name="givenName" minOccurs="0"/> <xs:element name="familyName" minOccurs="0"/> <xs:element name="nameIdentifier" xsi:type="nameIdentifier" minOccurs="0" maxOccurs="unbounded"/> <xs:element name="affiliation" xsi:type="affiliation" minOccurs="0" maxOccurs="unbounded"/>

School / Institute		faculty [0..∞] string Track the faculty (if concept relevant) of the owner.	<originatingSource>http://www.qut.edu.au</originatingSource>	<xs:documentation>The main researchers involved working on the data, or the authors of the publication in priority order. May be a corporate/institutional or personal name.</xs:documentation>
Data Manager		organisation [1..1] enumeration Organisation that owner belongs to (when the collection was created)	isManagedBy: is maintained and made accessible by the related party	<xs:annotation> <xs:documentation>The institution or person responsible for collecting, managing, distributing, or otherwise contributing to the development of the resource.</xs:documentation> </xs:annotation>
Collaborators (WSU)		organisation [1..1] enumeration Organisation that owner belongs to (when the collection was created)	<citationInfo> <citationMetadata> <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>	<xs:attribute name="contributorType" type="contributorType" use="required">
External Collaborators		organisation [1..1] enumeration Organisation that owner belongs to (when the collection was created)	hasParticipant: is undertaken by the related party	<xs:attribute name="contributorType" type="contributorType" use="required">

Principal Supervisor (students only)		role [0..1] enumeration AN optional role to be added when there are two or more owners, and these owners have a student/supervisor relationship. HDR Student Primary Supervisor Supervisor	<citationInfo> <citationMetadata> <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>	<xs:attribute name="subjectScheme" use="optional"/>

Data and software

	Dublin Core		**RIF-CS format**	**DataCite**
ReDBox Fields	**Idea:DCMI**	**Idea:Collection**		
What types of data will be produced and in what format will it be stored?	format [0..1] string Format of Media		<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>	<xs:simpleContent> <xs:extension base="xs:string"> <xs:attribute name="rightsURI" type="xs:anyURI" use="optional"/> <xs:attribute name="rightsIdentifier" use="optional"/> <xs:attribute name="rightsIdentifierScheme" use="optional"/> <xs:attribute name="schemeURI" type="xs:anyURI" use="optional"/> <xs:attribute ref="xml:lang"/> </xs:extension> </xs:simpleContent>

What software if any is required to collect, collate or analyse your data?		description [1..1] document null	Electronic has a number of child elements: • a **Value** containing a URI representing the electronic address (required) • a **Title** for the electronic address target (optional) • **Notes** providing any information relevant to the electronic element (optional) • the **MediaType**(s) of the electronic target formatted in accordance with the IANA standard media-types (optional) • the **ByteSize** of the electronic target (optional) • Arg to describe the arguments for a machine-to-machine service (optional, services only). If Arg is used, the attributes @required and @type are required, and @use is optional.	<xs:attribute name="subjectScheme" use="optional"/>
What third party data are you using and are there any special requirements around the use of the data?	relation [0..1] string Related Object		As above	<xs:attribute name="subjectScheme" use="optional"/>

Data Storage

	Dublin Core		RIF-CS format	DataCite
ReDBox Fields	**Idea:DCMI**	**Idea:Collection**		
Expected size of the data collected			Subject Type (activity collection party service)	xs:element name="sizes" minOccurs="0"> <xs:complexType> <xs:sequence> <xs:element name="size" type="xs:string" minOccurs="0" maxOccurs="unbounded"> <xs:annotation> <xs:documentation>Unstructures size information about the resource.</xs:documentation>
Storage location during project (including physical)			<location> Element May contain: address | spatial (location) Contained within: activity | collection | party | service	<xs:element name="geoLocations" minOccurs="0"> <xs:complexType> <xs:sequence> <xs:element name="geoLocation" minOccurs="0" maxOccurs="unbounded"> <xs:complexType> <xs:choice maxOccurs="unbounded"> <xs:element name="geoLocationPlace" minOccurs="0"> <xs:annotation> <xs:documentation>Spatial region or named place where the data was gathered or about which the data is focused.</xs:documentation> </xs:annotat

Data Retention

	Dublin Core		RIF-CS format	DataCite
ReDBox Fields	**Idea:DCMI**	**Idea:Collection**		
Applicable minimum retention period (5,10, 10 or permanently)		retention [0..1] document The date of review for retaining the collection. This should trigger a review of the collection.	<rights> <accessRights type="Restricted" rightsUri="http://www.paradisec.org.au/downloads.html"> Access may be granted following approval of a written application.</accessRights> </rights>	<xs:attribute name="subjectScheme" use="optional"/>
What do you intend to do with the research data and primary materials at the end of the retention period? (retain or destroy)			<licence> Element Contained within: rights	<xs:attribute name="subjectScheme" use="optional"/>
Select the earliest possible disposal date of the data			<rights> <accessRights type="Restricted" rightsUri="http://www.paradisec.org.au/downloads.html"> Access may be granted following approval of a written application.</accessRights> </rights>	<xs:attribute name="subjectScheme" use="optional"/>

Ethics and Sensitives

Ethics Approval Number			Not mapped	
Type of sensitivity • **Commercial sensitivity** • **Culturally sensitive** • **Security classified** • **Participants identifiable** • **Information about private, confidential or sensitive data and how it will be managed**		description [0..1] string Any free text description you would like to add to qualify the rights.	<accessRights> Element Used to describe access rights and access constraints, including who may access and when access may occur (including any embargo). Restrictions may be based on security, privacy or other policies. A free text statement may be provided. Preferably, choose a type from the accessRights type vocabulary. Examples: Access is restricted to the project participants until six months after the completion of the project. Available from 11.30am AEST 9 February 2010 Not for release before 7.30pm AEST 13 May 1997 Contained within: rights Attributes:	<xs:include schemaLocation="include/datacite-descriptionType-v4.xsd"/>
Information about private, confidential or sensitive to data and how it will be manages?		description [0..1] string Any free text description you would like to add to qualify the rights.	<service> Element Wrapper element for descriptive and administrative metadata for service registry object. May contain: accessPolicy | description | identifier | location | name | relatedInfo | relatedObject | subject Contained within: registryObject	<xs:attribute name="subjectScheme" use="optional"/>

Can some or all the research dataset be made accessible to other researchers? • **Open** • **Mediated** • **Restricted**		Not mapped	collection/rights/licence[@rightsURI] AND/OR collection/rights/licence[@type] AND collection/rights/licence	<xs:documentation>Any rights information for this resource. Provide a rights management statement for the resource or reference a service providing such information. Include embargo information if applicable. Use the complete title of a license and include version information if applicable.</xs:documentation>

Ownership, licensing, and IP

	Dublin Core		**RIF-CS format**	**DataCite**
ReDBox Fields	**Idea:DCMI**	**Idea:Collection**		
Copyright and intellectual property owners	licence [1..1] string Licence under which this dataset is published. If no licence, then explicitly state 'no licence'.		<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/"></licence> <accessRights type="Open" rightsUri="http://w3.unisa.edu.au/policies/policies/resrch/res20.asp">In accordance with the UniSA Open Access policy></accessRights> </rights>	<xs:simpleContent> <xs:extension base="xs:string"> <xs:attribute name="rightsURI" type="xs:anyURI" use="optional"/> <xs:attribute name="rightsIdentifier" use="optional"/> <xs:attribute name="rightsIdentifierScheme" use="optional"/> <xs:attribute name="schemeURI" type="xs:anyURI" use="optional"/> <xs:attribute ref="xml:lang"/> </xs:extension> </xs:simpleContent>

Please list any other owners		Not mapped	<subject> Element A term or phrase representing the primary topic(s) on which a registry object is focused. Use subject to associate activities with the field of activity, collection with the subject matter of items in the collection and party with field of activity or occupation. Services can be assigned subjects but may also be associated with a topic through the collections which support them. Contained within: activity | collection | party | service	<xs:attribute name="subjectScheme" use="optional"/>
Information about contractual obligations that apply to this data			other [0..1] string Any other information you have about the oragnisational affiliation of the contact. Contained within: activity | collection | party | service matter of items in the collection and party with field of activity or occupation. Services can be assigned subjects but may also be associated with a topic through the collections which support them.	<xs:attribute name="subjectScheme" use="optional"/>

Create data publication form: Metadata schema comparison chart:

Key Sections are:

- Coverage
- Select data
- Supplements
- Licence
- Citation
- Data Retention

Coverage

	Dublin Core		RIF-CS format	DataCite
ReDBox Fields	**Idea:DCMI**	**Idea:Collection**		
Temporal (time) coverage • **End Date** • **Time period** • **Geospatial location** • **Enter KML or GeoJSON**	spatial-coverage [0..1] string Geolocation of data collection temporal-coverage [0..1] string Temporal Coverage		<spatial> Element (location) Holds geographical address information such as co-ordinates or region information. Contained within: location NOTES: Spatial values encoded in XML markup (i.e. gml, gpx and/or kml) will cause RIF-CS schema validation errors. These values should either be escaped or referenced by providing a URL pointing to the markup. Do not use location to record the data coverage (temporal/spatial) of a registry object. Instead use the <coverage> elemen	<!-- definitions for geoLocation --> <xs:complexType name="point"> <xs:all> <xs:element name="pointLongitude" type="longitudeType" minOccurs="1"/> <xs:element name="pointLatitude" type="latitudeType" minOccurs="1"/> </xs:all> </xs:complexType>

Select Data

	Dublin Core		RIF-CS format	DataCite
ReDBox Fields	**Idea:DCMI**	**Idea:Collection**		
data manager is recorded as		contacts [0..1] document One or more contact persons (e.g. the Chief Investigator) to assist in decision making when the expiry time has lapsed.	<rightsStatement> Element A statement about the rights held in a collection. These may be intellectual property rights such as copyright or moral rights. Example: Copyright, John M. Smith, 2011 Contained within: rights	<xs:documentation>Any rights information for this resource. Provide a rights management statement for the resource or reference a service providing such information. Include embargo information if applicable. Use the complete title of a license and include version information if applicable.</xs:documentation>
• **Access conditions:** • **Open** • **Restricted** • **Conditional**		access_rights [1..1] enumeration Whether or not the dataset can be accessed openly (open), or if it requires permission or some other condition to be met (conditional), or is completely closed (restricted). open conditional restricted	<licence> Element A legal statement giving official permission to do something with a collection. Use this element to describe the type of licence that applies to the data. A free text statement may be provided. Preferably, choose a type from the licence type vocabulary. A licence is recommended for all data made available for reuse. Contained within: rights	<xs:documentation>Any rights information for this resource. Provide a rights management statement for the resource or reference a service providing such information. Include embargo information if applicable. Use the complete title of a license and include version information if applicable.</xs:documentation>

Supplements

	Dublin Core		RIF-CS format	DataCite
ReDBox Fields	**Idea:DCMI**	**Idea:Collection**		
Related publications	relation [0..1] string Related Object		<description> Element (relation) A plain text description further refining or describing a relation. Contained within: relation	</xs:complexType> </xs:element> <xs:element name="relatedItems" minOccurs="0"> <xs:complexType> <xs:sequence> <xs:element name="relatedItem" minOccurs="0" maxOccurs="unbounded"> <xs:annotation> <xs:documentation>Information about a resource related to the one being registered e.g. a journal or book of which the article or chapter is part.</xs:documentation> </xs:annotation> <xs:complexType> <xs:sequence> <xs:element name="relatedItemIdentifier" minOccurs="0"> <xs:annotation> <xs:documentation>The identifier for the related item.</xs:documentation> </xs:annotation>
Related website	relation [0..1] string Related Object		<description> Element (relation) A plain text description further refining or describing a relation. Contained within: relation	As Above

Related metadata (including standards, codebooks, vocabularies, thesauri, ontologies)	relation [0..1] string Related Object		<description> Element (relation) A plain text description further refining or describing a relation. Contained within: relation	As above
Related data	related_data [0..∞] string Record related research data sets/collections (and optionally specify their relationship)		<description> Element (relation) A plain text description further refining or describing a relation. Contained within: relation	As above
Related services			<description> Element (relation) A plain text description further refining or describing a relation. Contained within: relation	<description> Element (relation) A plain text description further refining or describing a relation. Contained within: relation

Licence

	Dublin Core		RIF-CS format	DataCite
ReDBox Fields	**Idea:DCMI**	**Idea:Collection**		
The data will be licensed under	rights-info [1..1] string Rights Information		<licence> Element A legal statement giving official permission to do something with a collection. Use this element to describe the type of licence that applies to the data. A free text statement may be provided. Preferably, choose a type from the licence type vocabulary. A licence is recommended for all data made available for reuse. Contained within: rights	<xs:documentation>Technical format of the resource.</xs:documentation> <xs:documentation>Use file extension or MIME type where possible.</xs:documentation> </xs:annotation> </xs:element> </xs:sequence> </xs:complexType> </xs:element> <xs:element name="version" type="xs:string" minOccurs="0"> <xs:annotation> <xs:documentation>Version number of the resource. If the primary resource has changed the version number increases.</xs:documentation> <xs:documentation>Register a new identifier for a major version change. Individual stewards need to determine which are major vs. minor versions. May be used in conjunction with properties 11 and 12 (AlternateIdentifier and RelatedIdentifier) to indicate various information updates. May be used in conjunction with property 17 (Description) to indicate the nature and file/record range of version.</xs:documentation> </xs:annotation> </xs:element> <xs:element name="rightsList" minOccurs="0"> <xs:complexType> <xs:sequence> <xs:element name="rights" minOccurs="0" maxOccurs="unbounded">

				<xs:annotation> <xs:documentation>Any rights information for this resource. Provide a rights management statement for the resource or reference a service providing such information. Include embargo information if applicable. Use the complete title of a license and include version information if applicable.</xs:documentation>
Statement of rights in data	rights-info [1..1] string Rights Information		A value taken from the Licence Type vocabulary.	As above

Citation

	Dublin Core		**RIF-CS format**	**DataCite**
ReDBox Fields	**Idea:DCMI**	**Idea:Collection**		
Request a DOI checkbox	doi [0..1] string DOI identifying the dataset		<identifier> Element (citationMetadata) This element is used to enter a string literal which uniquely identifies the publication being cited within a <citationMetadata> element. (E.g. ISSN/ISBN/ReportNumber) Contained within: citationMetadata	<xs:annotation> <xs:documentation>The type of the Identifier for the related item e.g. DOI.</xs:documentation> </xs:annotation>
Title	title [1..1] string Name of the collection or dataset		<url> Element (citationMetadata) Use this element to specify the URL of a publication being cited within a <citationMetadata> element. Contained within: citationMetadata	<xs:element name="titles" minOccurs="0">

Creators	creator [0..1] string The person primarily responsible for the creation of the data set		<namePart> Element (name) A name can be represented by either a single namePart (as in the case of an organization or group) or may be split into a more granular structure (as in the case of a person) through use of multiple namePart elements (e.g. for title, first name, surname, etc). Contained within: name	<xs:annotation> <xs:documentation>Uniquely identifies a creator or contributor, according to various identifier schemes.</xs:documentation> </xs:annotation
Publisher	publisher [0..1000] string The person responsible for the publication of the data set		<publisher> Element Use this element to specify the publisher of a publication within a <citationMetadata> element. Contained within: citationMetadata	<xs:element name="publisher">
Publication date	date-created [0..1] string null		<date> Element (dates) Use this element to record a date as a string. When entering a date range use two <date> elements with dateType attributes of 'dateFrom' and 'dateTo'. Both dates should be entered using the same format where possible. Use the 'dateFormat' attribute on the <date> element to specify the format of the date being entered. Contained within: dates	<xs:documentation>Year when the data is made publicly available. If an embargo period has been in effect, use the date when the embargo period ends.</xs:documentation> <xs:documentation>In the case of datasets, "publish" is understood to mean making the data available on a specific date to the community of researchers. If there is no standard publication year value, use the date that would be preferred from a citation perspective.</xs:documentation> <xs:documentation>YYYY</xs:documentation>

Embargo			<accessRights> Element Used to describe access rights and access constraints, including who may access and when access may occur (including any embargo). Restrictions may be based on security, privacy or other policies. A free text statement may be provided. Preferably, choose a type from the accessRights type vocabulary. Examples: Access is restricted to the project participants until six months after the completion of the project. Available from 11.30am AEST 9 February 2010 Not for release before 7.30pm AEST 13 May 1997 Contained within: rights	<xs:documentation>Year when the data is made publicly available. If an embargo period has been in effect, use the date when the embargo period ends.

Data Retention

	Dublin Core		**RIF-CS format**	**DataCite**
ReDBox Fields	**Idea:DCMI**	**Idea:Collection**		
Applicable minimum retention period		retention [0..1] document The date of review for retaining the collection. This should trigger a review of the collection.	<dates> Element Wrapper element used to describe dates associated with events in the life cycle of a data collection. Must contain: date (dates) Contained within: collection Attributes:	

Justification for extended retention period		Not mapped	Not mapped	Not mapped
Select the earliest possible disposal date of the data		Not mapped	Not mapped	Not mapped

Sources: https://services.ands.org.au/documentation/rifcs/guidelines/

Sources: https://schema.datacite.org/meta/kernel-4.4/doc/DataCite-MetadataKernel_v4.4.pdf

Appendix 2

Example of RIF-CS in Research Data Australia

RIFCS:

```
<?xml version="1.0" encoding="UTF-8"?>
<registryObjects xmlns="http://ands.org.au/standards/rif-cs/registryObjects" xmlns:xsi="http://
  <registryObject group="Western Sydney University">
    <key>10.26183/7d43-ze13</key>
    <originatingSource type="">https://research-data.westernsydney.edu.au</originatingSource>
    <collection type="dataset">
      <name type="primary">
        <namePart>Proteomics data for androgen receptor splice variant-7 in circulating tumour
      </name>
      <description type="full">&lt;p&gt;The dataset contains the raw data for samples 7 and 8,
      <rights>
        <rightsStatement>Copyright Western Sydney University</rightsStatement>
        <licence type="CC-BY" rightsUri="http://creativecommons.org/licenses/by/4.0">CC BY 4.
        <accessRights type="open">Open</accessRights>
      </rights>
      <identifier type="local">https://research-data.westernsydney.edu.au/published/cc59f230dc
      <identifier type="doi">10.26183/7d43-ze13</identifier>
      <dates type="dc.created">
        <date type="dateFrom" dateFormat="W3CDTF">2022-05-26</date>
      </dates>
      <location>
        <address>
          <electronic type="url" target="landingPage">
            <value>https://doi.org/10.26183/7d43-ze13</value>
            <title>Western Sydney ResearchDirect</title>
```