Australian Research Data Commons

ARDC Data Retention Project

Phase 2: Significant National Data Collections

Project Documentation

ARDC Storage and Compute

Release Mar 2021

Table of Contents

Statement of Intent 4
Project Purpose: 4
Key Terms: 5
Relationship with other ARDC programs and investments 6
Benefits 6
Benefit to Research Organisations 6
Benefit to Infrastructure providers 6
Benefit to Researchers 7
Benefit to ARDC 7
Eligibility 7
Eligible Organisations: 7
Eligible Data Collections: 7
Support 8
What we will support: 8
Who we will support: 8
How we will support: 8
Application process 8
Capacity: 8
Capability: 8
Authority: 9
Stage 1: Expression of Interest 9
Stage 2: Request for Proposal 9
Notification and Project Commencement 10
Project Closure 10
Project Timeline: 11
Partnership Model 11
Ideal state 11
FAIR Data Principles 12
Commitment to Full Compliance 12
ARDC Investment 12
Co-investment 13
Data Collection Assessment Criteria 13

Metadata related to Value Criteria 14
Commencement 15
Progression Reviews 15
Introducing Additional Collections 17
Flow Diagram 18
Worked Examples 19
Disclaimers 21
Appendix 1. NCRIS Principles 22
Appendix 2. Mapping required metadata to the DataCite metadata schema 23
Appendix 3. DataCite Metadata Schema tables + Project Specific Guidance 25
Controlled List Definitions 39
Standard values for unknown information 51

Statement of Intent

> *"ARDC is a transformational, sector-wide initiative, working with sector, government, and industry partners to build a coherent national and collaborative research data commons. This will deliver a world-leading data advantage, facilitate innovation, foster collaboration and enhance research translation."*

The ARDC believes that to deliver a data advantage to Australian researchers and maximise the impact of research output of meritorious research, researchers must have timely access to high quality data collections. To assess the impact of infrastructure investment in the data retention sector, we will leverage research data management practices that enrich data collections with controlled and consistent metadata.

This project will use a capacity subsidy model to support underpinning infrastructure needed to store important data collections that reflect the principles of NCRIS (see Appendix 1) and can be characterised by any or all of the following:

- Reusable data products of the national and/or collaborative research infrastructure
- Data output form meritorious research (e.g., funded via competitive grants)
- Data products that are evidence or cited in the scholarly record (e.g., peer reviewed academic publications)
- Data collections that can demonstrate enduring or special value to the research sector

Project Purpose:

The Data Retention project is a 3-year (2020-2023) partnership model to support the retention of valuable Australian research data collections and to embed controlled international metadata standards that enable access, reuse and impact assessment of these collections, ultimately defining a broad framework of nationally significant data collections. The project will drive an uplift in the description of data collections with 13 standard metadata elements that record concepts such as ownership, access conditions and location, amongst others. The ARDC will invite participation from all eligible organisations.

- Phase 1: 2020-2023. A co-investment partnership will be invited from eligible organisations who manage existing research data collections stored on legacy RDSI capacity and having a minimal 'eligibility' metadata. Partners are required to enrich these collections to a more FAIR state (defined as having 'eligibility' and 'value' metadata) during the course of the project.
- Phase 2: 2021-2023 **THIS PHASE**. A co-investment partnership will be invited from new Universities and NCRIS capabilities partners that manage significant research data collections not included in Phase 1. New partners are also required to enrich data collections to a more FAIR state (defined as having 'eligibility' and 'value' metadata) during the course of the project.
- Phase 3: Project closure will assess project effectiveness and build an impact and sustainability model to inform strategic vision and future investment targets.

From 2023 onwards, all ARDC partnerships seeking support for data collections will require those data collections to already possess ALL 13 metadata elements.

Intended Audience Phase 2

The Phase 2 project stage is available to whole-organisational units within universities and NCRIS capabilities that are responsible for managing large scale storage infrastructures and/or capacity to maintain important data collections. We will invite single applications from interested universities and NCRIS capabilities. Participation in Phase 1 does not exclude organisations from Phase 2.

ARDC will partner with eligible organisations to support underpinning capacity to maximise the impact of eligible data collections from nationally meritorious research and secure them for future use in research. This project will build a definition of nationally significant data collections and by enriching them with best-practice metadata ensure they are more findable and accessible and reusable as part of a robust scholarly record.

Key Terms:

Term	Context for the purposes of the Data Retention Project Phase 2
Eligible Stakeholders	Australian Universities and NCRIS capabilities who are storing eligible research data collections according to NCRIS principles, and: • Minimise barriers to access and re-use of data collections (i.e. not simply an institutional access) • Give due regard to infrastructure lifetimes and operational overheads • Demonstrate a commitment to the FAIR data principles while not compromising security or legislative obligations • A single application will be accepted from each interested Australian University and/or NCRIS capability.
Support	The ARDC will co-invest as partnerships to elevate important national collections to a more FAIR state. The ARDC will require at least 1:1 co-investment with matched cash contributions from each partner for the duration of the project.
Eligible data collections	Data collections that arise from, or are available for, nationally meritorious research. Such data collections are considered important to persist after initial project conclusion either as part of the scholarly record or to maximise reuse potential. Collectively they are significant data collections and must be able to progress to their most FAIR state.
Nationally meritorious research	Research funded via competitive research funding programs, or undertaken through the principles and/or resources of the NCRIS programme. Evidence for merit can be grant numbers, national allocation references or data citations in peer-reviewed academic publications.
Future use	Use of data collections that occurs after primary purpose and not necessarily by the same researchers, sometimes referred to as secondary use.
Nationally significant data collections	A clearer definition of 'nationally significant data collections' will be built by the ARDC and partners during this project. This definition

	will consider how to identify and measure impact of a data collection and what benefits such designation should confer.
Findable	A data collection that can be found independently, i.e. by persons unconnected to the owner, primary project or custodian.
Accessible	A data collection that can be accessed independently, i.e. by persons unconnected to the owner, primary project or custodian.
Reusable	A data collection that can be reused independently by persons unconnected to the owner, according to clear licensing conditions and minimal barriers.

Relationship with other ARDC programs and investments

The Data Retention Project supports other ARDC investments by investing in a secure and sustainable data infrastructure capacity and common metadata specification. This project establishes a coherent and consistent view of important data collections across the national research sector and quantifies the operational requirements to store and present those collections in line with the FAIR Data Principles, established academic conventions and the NCRIS Principles. The minimal but purposeful metadata requirements do not significantly impact the context, utility or cultural oversight of important data collections. Together with the ARDC Nectar Research Cloud, the Data retention Project underpins a robust and responsive national data commons.

Benefits

Benefit to Research Organisations

This project will help increase institutional compliance with R8 of the Australian Code for the Responsible Conduct of Research.

> *"Provide access to facilities for the safe and secure storage and management of research data, records and primary materials and, where possible and appropriate, allow access and reference[1]."*

Benefit to Infrastructure providers

Partners in this call will be Australian universities or NCRIS capabilities that can support both capital infrastructure and operational overheads to secure important data collections. This investment will underpin robust, reliable and high-quality data in a coherent national data commons. The international metadata standards used in this project facilitate data alignment and integration with other international initiatives such as the European Open Science Cloud, World Data System and DataOne.

[1] Australian Code for the Responsible Conduct of Research 2018. National Health and Medical Research Council, Australian Research Council and Universities Australia. Commonwealth of Australia, Canberra

Benefit to Researchers

Researchers will benefit from this investment through partnering infrastructure providers and will be able to more easily fulfil compliance requirements from major Australian (and international) funders for data management and the preservation of the evidence that underpins their research. They will also be able to draw from the growing body of data collections to fuel their future research with certainty and transparency.

This project will help to substantially increase a researcher's compliance with R22 of the Australian Code for the Responsible Conduct of Research.

> *"Retain clear, accurate, secure and complete records of all research including research data and primary materials. Where possible and appropriate, allow access and reference to these by interested parties[1]."*

Benefit to ARDC

Supporting underpinning infrastructure and layering the FAIR principles on data collections will secure a significant and valuable foundation for a coherent national and collaborative research data commons. Lowering the barriers to find and access quality data collections enables data driven innovation, collaboration and research translation to maximise the impact of sector investment.

Researchers will routinely be able to find and have access to high-quality research data assets.

Eligibility

Eligible Organisations:

Single applications for Phase 2 of the Data Retention Project are invited from NCRIS capabilities and Universities that can demonstrate the principles of NCRIS (see Appendix 1) with regard to the data collections they manage, namely they are national in scope, collaborative in nature, contribute to the wider research data infrastructure or possess enduring value to the research sector.

Eligible Data Collections:

For the purposes of the ARDC Data Retention Project, an "eligible data collection" is defined as a logical structure of data generated as a significant output from, or input to, meritorious research activity. Data collections may be the result of a single research undertaking, a national facility or NCRIS capability, or a collaborative effort within an international community.

The ARDC recognises that instantiation of this definition will be different across domains, disciplines and facilities. The goal of this investment is to establish a broad and minimal set of criteria which will build a coherent definition of "data collections of national significance" whose wider research impact can be championed and supported for maximum impact.

In all cases eligible data collections must fulfil either any or all of the following criteria;

- Value as irreplaceable collections, e.g., observational or unrepeatable collections
- Merit as the output from peer-reviewed funded activity
- Importance as evidence used to support traditional publication

Support

What we will support:

The ARDC will co-invest in capital or operational burdens of data storage capacity as a volume subsidy. Data storage infrastructures can be private, commercial or hybrid services, e.g., via on premise hardware, cloud, or third-party capacity, but the data collections they manage must be available and a national asset[2].

Who we will support:

Capacity will be managed as central enterprise infrastructure by an Australian university or NCRIS capability. In cases where single organisations administer other eligible applicants, for example a university that hosts an NCRIS capability, each application will be considered as independent of each other (despite ultimately being administered by a single organisation). In such cases due diligence will be implemented to minimise the risk of unintended duplication of investment.

How we will support:

The mechanism for support will be a capacity cost subsidy model that is structured as a compliance incentive to register controlled metadata over the duration of the project (ending 30 June 2023). A per TB $AUD subsidy will be calculated according to a nominated register of data collections from each applicant. The total subsidy calculation will be divided according to a pre-defined ratio and available at three stages of the project; commencement and two progression reviews (see the investment model described below).

Application process

To streamline our process, we will invite applications via a two-stage expression of interest (EOI), followed by a request for proposal (RFP) application. Applicants should manage data storage facilities or services and demonstrate the capability and authority to undertake the work requested in this call, specifically:

Capacity:

For the purposes of this investment partnership, a total single copy volume of nominated data collections should be calculated. For clarity this investment will not consider measures for resilience such as backup, duplication or redundancy.

Capability:

Applicants should manage the infrastructure or capacity that maintains eligible data collections in accordance with good RDM practice (e.g., via a data repository or other publication and access service) such that metadata can be collected and registered in the public metadata registry, DataCite MDS. This infrastructure can be operated in house or via third party provision (e.g., cloud services). Ideally, managed infrastructure or capacity would include a public facing data catalogue or registry, but use of Research Data Australia is encouraged if no such facility exists.

[2] Noting that only data collections that are consumed or generated via Australian derived research or researchers will be eligible for support

Authority:

It is essential that applicants possess the authority from rights holders to collect and register the requested metadata criteria into a public metadata registry and in doing so indicate a commitment to the FAIR data principles.

Stage 1: Expression of Interest

Expressions of Interest for Phase 2 of the Data Retention Project are invited from 1st March 2021 until the 26th March 2021.

The EOI will require contact details, estimates on data collection capacity and a short description of anticipated benefits (to the researcher, the organisation and the Australian research sector.

The EOIs will be assessed by the Data Retention Project team. Some or all of the EOI submissions will be considered for full proposals based on the scope of disciplines, anticipated benefits/alignment with the NCRIS Principles, organisational representation, available funds and national capacity demand.

Applicants for the EOI should read the full project documentation and RFP when deciding to submit an EOI.

Submission of EOI will follow the format found in this ARDC Data Retention Project EOI template

Stage 2: Request for Proposal

Some or all of the EOIs will be invited to submit a full Request for Proposal (RFP) on or shortly after the 16th April 2021 with completed applications received no later than 7th May 2021.

Three (3) components are required for application:

1. Applicants will submit a MS Excel spreadsheet (from a supplied template) nominating data collections along with six (6) metadata values related to eligibility (see below). Applicants should be responsible for managing these data collections and making them available according to academic conventions.
2. Applicants will submit a concise project plan (max 500 words) which will describe how nominated data collections will be managed into the DataCite publicly accessible metadata registry with 13 mandatory metadata elements. For example, how capacity infrastructure is managed to permit access to data collections, use of any products or services that facilitate metadata management of data collections, dedicated teams and timelines of expected work, etc.
3. Applicants will also submit a financial budget detailing a proposed investment allocation for this partnership according to the requirements of ARDC investment and co-investment described below.

The general terms and conditions of ARDC contracts will be provided during the RFP, including a schedule of activity relevant to each applicant.

The ARDC will assess each application (register of data collections, project plan and allocation budget) against the following criteria,

- Completeness of application: All eligibility metadata for all nominated data collections must be complete
- Consistency with the EOI: The project plan must reflect the intention and estimates of the EOI.
- Organisational representation
- Distribution of research disciplines

Notification and Project Commencement

Successful applicants will be notified by 4th July 2021

Commencement may occur at any date between 6th August 2021 and 1st Sept 2021.

To be clear, contracts must be executed by the 1st September 2021 or any offer made will be withdrawn.

Project Closure

The Data Retention Project is scheduled to close on 30 June 2023.

At project completion all nominated data collections should be recorded in the DataCite MDR.

Project Timeline:

Date	Event
01/03/2021	Release EOI
9/03/2021	EOI Webinar 1
18/03/2021	EOI Webinar 2
26/03/2021	Deadline EOI
29/03/2021-09/04/2021	EOI Assessment
16/04/2021	Notification and RFP
22/04/2021	RFP Webinar 1
29/04/2021	RFP Webinar 2
7/05/2021	Deadline RFP
17/05/2021-04/06/2021	RFP Assessment
04/07/2021	Notification & Contracts
06/08/2021 to 01/09/2021	Commencement
08/07/2022	Review Cycle 1
05/05/2023	Review Cycle 2
30/06/2023	Project Closes

Partnership Model

The goals of this model is to support data infrastructure and provide a coherent view of important data collections by applying international metadata standards. Metadata used in the ARDC Data Retention Project will assess how nominated data collections support the FAIR data principles, in line with the ARDC strategic vision. This project will define and benchmark the effect of ARDC co-investment on eligible data collections stored by eligible institutions and work to provide a coherent national view of these data collections and their impact on infrastructure investment in the Australian research sector.

Ideal state

The ARDC will make use of contemporary metadata standards and controlled identifiers like DOI[3], ORCiD[4], ROR[5] with minimal use of uncontrolled or free-text values (but recognise that some textual narrative remains a valuable source of contextual information). In addition to clearer assessment of investment

[3] Provided via the International DataCite Digital Object Identifier (DOI) service

[4] Provided via the International ORCiD service

[5] Provided by the Research Organisation Registry

impact, this project will significantly advance efforts to build a realistic and transparent sustainability model for support of high value research data collections that will define their content, where they can be found and how they can be accessed and reused.

FAIR Data Principles

The ARDC Data Retention metadata criteria will be guided by the FAIR data principles.

To begin, the ARDC will support collections which are, or can be made more:

- Findable: Researchers, operating independently, are able to locate a metadata registry record of the data collection either via simple internet protocol searching tools, citations in traditional publications or other scholarly communications.
- Accessible: Researchers, operating independently, are able to access directly or with as few barriers as possible, the data collections described in the metadata registry record.
- Reuseable: Researchers, operating independently, are able to establish the conditions under which data reuse can occur and understand any obligations they must accept in such reuse.

Commitment to Full Compliance

The outcome of the Data Retention Project will be data collections enriched with thirteen (13) controlled and complete metadata elements as defined in the DataCite metadata schema. As such, entering a partnership with the ARDC by accepting this model signals that the applicant commits to extending the six (6) eligibility metadata criteria with a further seven (7) 'value' metadata criteria over the duration of the project. Further, each applicant must also acknowledge that any nominated data collections are not already part of this project, e.g., via Phase 1, nominated by other applicants or that they already exist as complete entries in the DataCite Metadata Registry.

ARDC Investment

The ARDC Storage and Compute theme has an annual budget of $6 million per annum for three years (2020-2023). The Data Retention Project has been allocated 50% of the total budget at $3 million per year for three years ($9 million in total). At its discretion, the ARDC Ltd Board may vary the theme or project allocation.

The ARDC Storage and Compute theme has allocated $4.4 million to this phase (Phase 2) of the Data Retention Project: Significant National Data Collections (2021-2023). The final phase, Phase 3, will focus on assessing impact.

Investment Calculation

A total ARDC investment, 'I_{ardc}', will be calculated from the capacity of nominated data collections in each register using the unit cost determined prior to RFP.

The total ARDC investment 'I_{ardc}' for each partner will be divided between a commencement allocation 'b_1' and a progression allocation 'b_2' according the ratio 30:70 (b_1:b_2)

The total ARDC allocated investment for each partner will fix a financial cap on each contract:

$$I_{ardc} = b_1 + b_2$$

The value b_1 will be fixed as 30% of total ARDC allocated investment.

The value b_2 will be fixed as 70% of total ARDC allocated investment.

Investment Distribution

The investment model distributes total ARDC investment across three milestones: 'commencement' and two progression 'reviews', described below.

The commencement investment 'b_1' will be available following successful application and on contract execution (see below).

The progression review investment 'b_2' will be used to draw-down investment following two (2) annual reviews on 8th July 2022 and 5th May 2023 (see below).

Co-investment

ARDC requires 1:1 matching co-investment for all infrastructure partnerships.

Co-investment must be in an auditable form (in accordance with reporting and accountability requirements as specified by the Commonwealth Department of Education) and can be cash (from project partners or other grants), or effort/labour. NOTE: if contributing effort/labour, work on the project must be a significant amount of the person's time, i.e., 25% or more).

Data Collection Assessment Criteria

To objectively evaluate nominated data collections, full compliance with standardised metadata is required. Metadata concepts are separated into two groups:

- eligibility metadata (the minimum requirement for application),
- value metadata (added to the eligibility metadata to enrich collections to a more FAIR state)

As the project progresses each application will register all metadata into the DataCite MDR to receive a DataCite DOI.

Metadata related to Eligibility Criteria - Required for RFP (not EOI)

This assessment determines the eligibility of the collection to be included in the Phase 2 partnership RFP.

Data collections will be selected by applicants as eligible and presented in application via a data collection register. They must contain all of the following metadata to be considered in this project.

Metadata ID	Metadata Concept	Metadata concept type	Accepted value types
E01	Collection capacity	Single copy	Terabyte-TB (1TB = 10^{12} bytes)
E02	Collection ID	ID at source/Local ID	User defined
E03	Applicable research disciplines (all that apply)	Fields of Research	Fields of research (FoR) codes. Can be 2-, 4- or 6-character codes

E04	Title	A semantically relevant title of the collection	Free text
E05	Description	Abstract style description of the data collection	Free text
E06	Data Controller/s[6] (all that apply)	Data controllers have responsibility for management of data and can apply licenses, access policies or other powers delegated from the data owner.	Research Organisation Registration (ROR) and/or ORCiD.

- For hierarchical collections, e.g., when multiple sub/or child collections exist as part of a higher 'parent' collection, applicants are able to nominate either parent OR child collections according to particular community convention for greatest reuse potential.
- For growing or versioned data collections, only 'versions of record' should be nominated, i.e., once they have reached a logical arrangement, are 'published' to a community and are immutable as records of research. For purposes of eligibility such collections can be accounted only once, with any additional growth being treated as an independent collection irrespective of whether it joins an existing collection.

Metadata related to Value Criteria

'Value metadata' are considered the minimum and achievable metadata required for data collections to be findable, accessible and reusable according to the FAIR data principles. Each partner will aim to increase criteria fulfilment over the contract period and the ARDC will conduct annual reviews to assess progress.

Metadata ID	Metadata Concept	Metadata concept type	Accepted value types
V01	Collection ID	Unique Persistent Identifier (UID or UPID)	DataCite Digital Object Identifier (DOI)
V02	Actionable Digital Registry entry	Landing page or registry entry that renders ALL metadata. Can be local data repository registry/repository or	URL/URI

[6] Taken from the EU GDPR regulation framework for responsibility in data protection, in the absence of a clear definition of similar entity in the (Australian Privacy Principles) or from the Australian OAIC and Australian Privacy Act 1988. Data Publisher or Data Custodian is sometimes used to describe a similar to Data Controller

		partners can make use of Research Data Australia[7]	
V03	Owner/s Contributors (all that apply)	Individual or organisational ownership declaration plus others involved	ORCiD for individuals or ROR for institutions.
V04	Merit (all that apply)	Peer reviewed grant authority, or data citation or other metric that quantifies reuse and/or impact	Grant number Citation Other metric
V05	License	Open licensing framework and waiver notices	Creative Commons licenses and waiver notices
V06	Collection type	Collection characterisation	Conferring some understanding of the data collection's characteristics, e.g. a dataset, image, model, workflow. Will be provided as a controlled list.
V07	Publication Year	The year collection was made available or if this is unknown, the year the collection was registered in DataCite.	YYYY

Commencement

To enter a partnership, data collections will be nominated by applicants and assessed as eligible by possessing ALL 6 eligibility criteria and a commitment to enrich them further according to the 7 value criteria.

- ***DELIVERABLE 1: List of nominated collections with six (6) metadata related to eligibility criteria.***

The allocation value 'b_1' will be determined according to the investment calculation above and a commencement investment will be available.

$$I_{contract} = b_1$$

- ***MILESTONE 1: Commencement investment***

Progression Reviews

[7] https://ardc.edu.au/services/research-data-australia/

Over the contract term, partners will register data collections in the DataCite Metadata Registry (MDR). A DataCite .xsd (current version 4.3) is publicly available for the DataCite MDR[8] and this should be used to register via the DataCite Fabrica API according to your local service provider. ARDC currently maintains a subscription to the DataCite Fabrica service on behalf of publicly-funded Australian research organisations.

Two (2) reviews will require partners to submit to the ARDC a list of all data collection DataCite DOIs registered (or updated[9]) during the period. The ARDC Project Team will extract metadata from the DataCite MDR and those collections that possess all 13 valid metadata criteria will be considered compliant.

- Only data collections with required metadata records in the DataCite MDR will be considered compliant.
- Data collections with incomplete records in the DataCite MDR will be considered non-compliant resulting in a reduced investment for that period.

To be clear, a failure to register complete records into the DataCite MDR will result in a proportional reduction in investment.

Data collections that were incomplete after review 1 can be re-submitted as part of the 2nd review.

Appendix 2 provides a map between the value and eligibility criteria and the relevant DataCite metadata concepts. The full DataCite metadata schema is provided in Appendix 3 where additional schema elements are available for partners to use for even greater benefit, e.g. contributor, version, geolocation and format, but this will be left to the discretion of the partners.

Progress Review 1

➢ ***DELIVERABLE 2: Partners will submit a list of DataCite DOIs minted or updated for any of the collections nominated in deliverable 1 (or agreed exceptions, see below, 'Introducing New Collections').***

By the 8th July 2022, all partners will submit a report listing DOIs created or updated over the previous period as an extra column in the nominated collection register submitted in the RFP. Progress review 1 will calculate a value variable (v_{r1}) as the capacity percentage of collections that are or have progressed to full compliance during the period from commencement to this review.

This value will be applied to the progression component 'b_2' and an investment reflecting progress will be awarded: $I_{review1}$.

$$I_{review1} = v_{r1} \times b_2$$

➢ ***MILESTONE 2: Review 1 investment***

Progress Review 2

➢ ***DELIVERABLE 3: Partners will submit a list of DataCite DOIs minted or updated for any of the collections nominated in deliverable 1 (or agreed exceptions, see below 'Introducing New Collections').***

By the 5th May 2023, all partners will submit a report listing DOIs created or updated over the previous period as an extra column in the nominated collection register submitted in the RFP. Progress review 2 will

[8] https://schema.datacite.org/meta/kernel-4.3/

[9] If a DOI already exists for a particular collection then an update is acceptable.

calculate a value variable (v_{r2}) as the capacity percentage of eligible collections (excluding collections previously designated compliant) that progress to full compliance during the period from commencement to progress review 2.

Similarly, this value will be applied to the progression component 'b_2' and an investment reflecting progress will be awarded: $I_{review2}$.

$$I_{review2} = v_{r2} \times b_2$$

➢ ***MILESTONE 3: Review 2 investment***

Note that in all partnerships $(v_{r1} + v_{r2}) \leq 1$

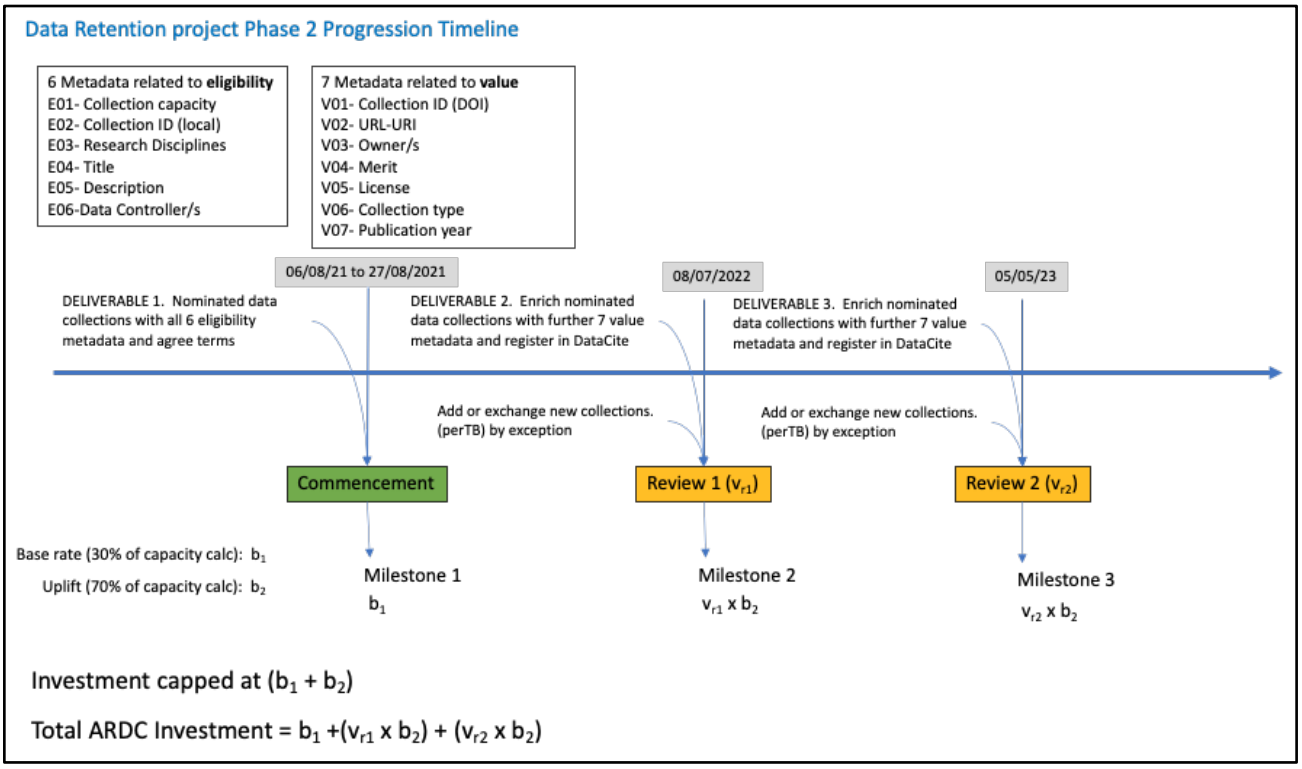
Introducing Additional Collections

Collections that were not included in the application register, e.g. if they were assessed as unable to be made fully compliant or were created during the period between commencement and review, will be considered as new collections during review cycles. New collections may be considered for inclusion into a partnership agreement after contract execution in two (2) circumstances:

- New collections can be 'exchanged' (per TB) for existing collections that are unable to meet metadata criteria.
- New collections can be 'added' (per TB) to existing data collection ledgers during review cycles.

In all cases, new collections are treated as exceptions to the agreement and acceptance will require justification, be at the discretion of the ARDC and subject to available funds. To be considered for inclusion, all new collections are required to be 100% compliant with all metadata criteria when nominated.

Flow Diagram



Worked Examples

NOTE: Until the EOI phase has determined a unit cost, these examples accurately reflect calculations but do not reflect eventual investment levels.

Example 1:

(Example assumes a unit cost @ AUD150/TB)

Applicant nominates 1000TB of data collections that they have been managing in an institutional data repository with a public facing registry/catalogue.

A calculated contract value will cap the ARDC investment at AUD150k over the partnership period.

A commencement allocation equal to 30% of AUD150k is awarded. ($I_{contract} = b_1$ = AUD 45k)

The remaining 70% (b_2 = AUD105k) is retained for drawdown awards following review cycles.

During year 1, 30% of the nominated collections have been elevated to full compliance with all 13 metadata elements: AUD31.5k is invested after Review 1.

$$I_{review1} = v_{r1} \times b_2$$

$v_{r1} = 0.30$ so $I_{review1} = 0.50 \text{ x } 105\text{k} = 31.5\text{k}$

During year 2, only a further 10% of the nominated collections are elevated to compliance with all 13 metadata elements. The final AUD10.5k is invested after Review 2.

$$I_{review2} = v_{r2} \times b_2$$

$v_{r2} = 0.10$ so $I_{review2} = 0.10 \text{ x } 105\text{k} = 10.5\text{k}$

The remaining AUD63K (105*0.6) is re-allocated to other parts of the project.

Example 2:

(Example assumes a unit cost @ AUD100/TB)

Applicant nominates 2500TB of data collections that they manage on behalf of a number of smaller data repositories from independent institutions.

A calculated contract value will cap the ARDC investment at AUD250k over the partnership period.

A commencement allocation equal to 30% of AUD375k is awarded ($I_{contract} = b_1$ = AUD 75k)

The remaining 70% (b_2 = AUD175k) is retained for drawdown awards following review cycles.

During year 1, 50% of the nominated collections have been elevated to full compliance with all 13 metadata elements: AUD 131.25k is invested after Review 1.

$$I_{review1} = v_{r1} \times b_2$$

$v_{r1} = 0.50$ so $I_{review1} = 0.50 \text{ x } 175\text{k} = 87.5\text{k}$

During year 2, the final 50% of the nominated collections are elevated to compliance with all 13 metadata elements. A further AUD131.25k is invested after Review 2.

$$I_{review2} = v_{r2} \times b_2$$

$v_{r2} = 0.50$ so $I_{review2} = 0.50 \text{ x } 175\text{k} = 87.5\text{k}$

Example 3:

(Example assumes a unit cost @ AUD70/TB)

Applicant nominates 3000TB of data collections that they have been managing in an institutional data repository with a public facing registry/catalogue.

A calculated contract value will cap the ARDC investment at AUD210k over the partnership period.

A commencement allocation equal to 30% of AUD450k is awarded ($I_{contract} = b_1$ = AUD63k).

The remaining 70% (b_2 = AUD147k) retained for drawdown awards following review cycles.

During year 1, 50% of the nominated collections have been elevated to full compliance with all 13 metadata elements: AUD73.5k is invested after Review 1.

$$I_{review1} = v_{r1} \times b_2$$

$v_{r1} = 0.50$ so $I_{review1} = 0.50 \text{ x } 147\text{k} = 73.5\text{k}$

During year 2, only a further 30% of the nominated collections are elevated to compliance with all 13 metadata elements. During the course of the year new collections are identified that fulfil all metadata criteria and application is made to exchange 10% of original capacity for these new collections. The exception is agreed with the ARDC Project Team and the review allocates 40% as an outcome, therefore $58.8k (44.1 + 14.7) is invested after Review 2.

$$I_{review2} = v_{r2} \times b_2$$

$v_{r2} = 0.30$ so $0.30 \text{ x } 147\text{k} = 44.1\text{k}$

plus

Agreed exception $= 0.10 \text{ x } 147\text{k} = 14.7\text{k}$

$I_{review2} = 44.1\text{k} + 14.7\text{k} = 58.8\text{k}$

Disclaimers

The ARDC will not in any way be bound by the timeline indicated in this project documentation. It will be under no obligation to respond to or accept any applications in whole or in part including any pricing, costs or funding requirements specified in the applications, from any Applicant.

All information included in this partnership call is provided in good faith and believed to be reliable. Each Applicant must make its own enquiries about the information provided and shall be deemed to have satisfied itself as to the correctness and sufficiency of this partnership call.

Nothing in this partnership call requires ARDC to select an application and ARDC reserves the right to discontinue the application process (including any subsequent partnership call) at any time and for any reason.

By lodging an application, Applicants acknowledge and agree that:

- they will not make any public statement, or provide any information for publication in relation to the acceptance or otherwise of any RFP submission, without the prior written approval of ARDC;
- to the maximum extent permitted by law, neither the ARDC nor its employees, advisers or agents will in any way be liable to any person or entity for any cost, expense, loss, claim or damage arising out of or in connection with this partnership call;
- they have not relied on any express or implied warranty or representation made by or on behalf of the ARDC other than as expressly contained in this RFP or an addendum to this RFP;
- they have not received improper assistance from any staff member of the ARDC;
- ARDC may alter this partnership call, including its specifications / requirements, structure and timing, at any time and for any reason;
- ARDC may invite additional Applicants to submit an application at any time;
- they have not colluded with other organisations to inflate funding estimates;
- they understand that proposals will be treated as confidential by ARDC and that ARDC will not disclose application contents nor Applicant information, except:
    - as required by law;
    - for the purpose of investigations by the Australian Competition and Consumer Commission or other government authorities having relevant jurisdiction;
    - to the Department of Education and Training, at their request and if required, to enable transparency and accountability; and/or
    - general information from Applicants required to be disclosed by government policy and as part of the partnership approval process.

Appendix 1. NCRIS Principles

Taken from the 2018 NCRIS Guidelines - Appendix A

The key principles underpinning NCRIS are that:

- Australia's investment in research infrastructure should be planned and developed with the aim of maximising the contributions of the research and development system to foster innovation, economic development, national security, social wellbeing and environmental sustainability
- infrastructure resources should be focussed in areas where Australia is, or has the potential to be, world-class (in both discovery and application driven research) and provide international leadership
- major infrastructure should be developed on a collaborative, national, nonexclusive basis. Infrastructure funded through NCRIS should serve the research and innovation system broadly, not just the host/funded institutions. Funding and eligibility rules should encourage collaboration and co-investment. It should not be the function of NCRIS to support institutional level (or even small-scale collaborative) infrastructure
- access is a critical issue in the drive to optimise Australia's research infrastructure. In terms of NCRIS funding there should be as few barriers as possible to accessing major infrastructure for those undertaking meritorious research, including the use of preferential arrangements for meritorious researchers
- due regard be given to the whole-of-life costs of major infrastructure, with funding available for operational costs where appropriate
- NCRIS should seek to enable the fuller participation of Australian researchers in the international research system
- NCRIS should enable Government initiatives which seek to maximise opportunities for industry and international engagement and commercialisation of research
- data generated, created, captured or stored by NCRIS funded projects will be made available to the wider research community based on the F.A.I.R. principles, appropriately implemented for individual research communities. Data must be stored to an appropriate level of security
- new projects, and additional investment in existing projects, should be based on a robust business case.

Appendix 2. Mapping required metadata to the DataCite metadata schema

Notes on occurrence (Occ) values

0-n = optional and repeatable

0-1 = optional, but not repeatable

1-n = required and repeatable

1 = required, but not repeatable

Metadata ID	Occ	Metadata Concept	DataCite Schema ID	Comments
E01	1	Collection capacity	13 Size	Terabyte-TB (1TB = 10^{12} bytes)
E02	1-n	Local Collection ID	11 AlternateIdentifier	User defined local ID
E03	1-n	Applicable research disciplines	6 Subject	Fields of research (FoR) codes. an be 2-, 4- or 6-character codes
E04	1	Title	3 Title	Free text
E05	1	Description	17 Description	Free text
E06	1-n	Data Controller/s[10]	4 Publisher	Research Organisation Registration (ROR). The role of 'Publisher' is given prominence in the data citation and will generally represent a primary role with rights over data collections or delegated authority from rights holders. Partners may wish to include the optional term 'contributor' (DataCite schema ID) for other types of related roles, see V03 below.
V01	1	Collection ID	1 Identifier	DataCite Digital Object Identifier (DOI)
V02	1-n	Actionable Digital Location	12 RelatedIdentifier	Research Data Australia or another repository/registry URL

[10] Taken from the EU GDPR regulation framework for responsibility in data protection, in the absence of a clear definition of similar entity in the (Australian Privacy Principles) or from the Australian OAIC and Australian Privacy Act 1988.

V03	1-n	Owner/s (Contributors)	2 Creator (7 Contributor)	ORCiD for individuals or ROR for institutions. Contributor is provided as an optional term that can be used for interests other than Publisher and Owner/s (which are mandatory) terms, e.g. a data repository or data storage service.
V04	1-n	Merit	19 FundingReference 12 RelatedIdentifier	Grant number (for 19) Citation IsCitedBy or IsSupplementTo or isReferencedby etc (for 12)
V05	1	License	16 RIghts	ARDC recommends Creative Commons licenses and waiver notices.
V06	1	Collection type	10 ResourceType	Controlled list values, most commonly dataset or data collection.
V07	1	Publication Year	5 Publication year	YYYY, generally taken to be when collection was made available or the year it was registered in DataCite.

Appendix 3. DataCite Metadata Schema tables + Project Specific Guidance

DataCite - International Data Citation

XSD is available via https://schema.datacite.org/meta/kernel-4.3/

Notes on occurrence (Occ) values

0-n = optional and repeatable

0-1 = optional, but not repeatable

1-n = required and repeatable

1 = required, but not repeatable

ARDC Data Retention ID	ARDC Data Retention Concept	DataCite ID	DataCite-Property	Occ	Definition	Allowed values, examples, other constraints	ARDC Data Retention Project Comments
V01	Collection ID	1	Identifier	1	The Identifier is a unique string that identifies a resource. For software, determine whether the identifier is for a specific version of a piece of software, (per the Force11 Software Citation Principles11), or for all versions.	DOI (Digital Object Identifier) registered by a DataCite member. Format should be "10.1234/foo"	
		1.a	identifierType	1	The type of Identifier.	Controlled List Value: DOI	
V03	Owner/s	2	Creator	1-n	The main researchers involved in producing the data, or the authors of the publication, in priority order. To supply multiple creators, repeat this property.	May be a corporate/institutional or personal name. Note: DataCite infrastructure supports up to 8000-10000 names. For name lists above that size, consider attribution via linking to the related metadata.	
		2.1	creatorName	1	The full name of the creator.	Examples: Charpy, Antoine; Foo Data Center Note: The personal name, format	

[11] Smith AM, Katz DS, Niemeyer KE, FORCE11 Software Citation Working Group. (2016) Software citation principles. PeerJ Computer Science 2:e86 https://doi.org/10.7717/peerj-cs.86

						should be: family, given. Nonroman names may be transliterated according to the ALA-LC schemas12.	
		2.1.a	nameType	0-1	The type of name	Controlled List Values: Organizational Personal	
		2.2	givenName	0-1	The personal or first name of the creator.	Examples based on the 2.1 names: Antoine; Mae	
		2.3	familyName	0-1	The surname or last name of the creator.	Examples based on the 2.1 names: Charpy; Jemison	
		2.4	nameIdentifier	0-n	Uniquely identifies an individual or legal entity, according to various schemas.	The format is dependent upon schema.	
		2.4.a	nameIdentifierScheme	1	The name of the name identifier schema.	If nameIdentifier is used, nameIdentifierScheme is mandatory. Examples: ORCID13, ISNI14, ROR14, GRID15.	
		2.4.b	schemeURI	0-1	The URI of the name identifier schema.	Examples: http://www.isni.org/ https://orcid.org https://ror.org/ https://www.grid.ac/	
		2.5	affiliation	0-n	The organizational or institutional affiliation of the creator.	Free text. The creator's nameType may be Organizational or Personal. In case of an organizational creator, e.g. a research group, you can add here the name of the formal institution to which the creator belongs.	
		2.5.a	affiliationIdentifier	0-1	Uniquely identifies the organizational affiliation of the creator.	The format is dependent upon schema. Examples : https://ror.org/04aj4c181 grid.461819.3	
		2.5.b	affiliationIdentifierScheme	1	The name of the affiliation identifier schema.	If affiliationIdentifier is used, affiliationIdentifierScheme is mandatory. Examples : ROR, GRID	
		2.5.c	SchemeURI	1	The URI of the affiliation identifier schema	Examples : http://www.isni.org http:/orcid.org https://ror.org/ https://www.grid.ac/	
E04	Title	3	Title	1-n	A name or title by which a resource is known. May be the title of a dataset or the name of a piece of software.	Free text.	
		3.a	titleType	0-1	The type of Title.	Controlled List Values: AlternativeTitle	

[12] http://www.loc.gov/catdir/cpso/roman.html

[13] https://orcid.org/. When entering an ORCID, follow these style guidelines: https://support.orcid.org/knowledgebase/articles/116780-structure-of-the-orcid-identifier [14] http://www.isni.org/

[14] https://ror.org/

[15] https://www.grid.ac/

						Subtitle TranslatedTitle Other	
E06	Data Controller	4	Publisher	1	The name of the entity that holds, archives, publishes prints, distributes, releases, issues, or produces the resource. This property will be used to formulate the citation, so consider the prominence of the role. For software, use Publisher for the code repository. If there is an entity other than a code repository, that "holds, archives, publishes, prints, distributes, releases, issues, or produces" the code, use the property Contributor/contributorType/ hostingInstitution for the code repository.	Examples: World Data Center for Climate (WDCC); GeoForschungsZentrum Potsdam (GFZ); Geological Institute, University of Tokyo, GitHub	The Publisher is a primary role and used to generate data citation. To acknowledge related or enabling roles aligned to Publisher consider using 'Contributor' (7)
V07	Publication Year	5	PublicationYear	1	The year when the data was or will be made publicly available. In the case of resources such as software or dynamic data where there may be multiple releases in one year, include the Date/dateType/dateInformation property and sub-properties to provide more information about the publication or release date details.	YYYY *** If an embargo period has been in effect, use the date when the embargo period ends. 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.	Has been included as a mandatory element for DataCite DOI minting
V06	Collection Type	10	ResourceType	1	A description of the resource.	The format is open, but the preferred format is a single term of some detail so that a pair can be formed with the sub-property. Text formats can be free-text OR terms from the CASRAI Publications resource type list.16 *** Examples: Dataset/Census Data, where 'Dataset' is resourceTypeGeneral value and 'Census Data' is ResourceType value. Text/Conference Abstract, where 'Text' is resourceTypeGeneral value and 'Conference Abstract' is resourceType value aligned with CASRAI Publications term.	For the Data Retention project this would be either a dataset or a data collection but partners are free to choose which ever 'type' value is useful. IT is included as a mandatory element for DataCite DOI minting
		10.a	resourceTypeGeneral	1	The general type of a resource.	Controlled List Values: Audiovisual	

[16] http://dictionary.casrai.org/Output_Types

						Collection DataPaper Dataset Event Image InteractiveResource Model PhysicalObject Service Software Sound Text7 Workflow Other See Appendix for definitions and examples	
E03	Discipline	6	Subject	1-n	Subject, keyword, classification code, or key phrase describing the resource.	Free text.	For the Data Retention project, we will be using the 2, 4 or 6 FoR codes (AUSNZ Standard Research Classification 'Fields of Research Codes')
		6.a	subjectScheme	1-n	The name of the subject scheme or classification code or authority if one is used.	Free text.	AUSNZ Standard Research Classification Scheme
		6.b	schemeURI	1-n	The URI of the subject identifier scheme.		https://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/1297.0Main+Features12008?OpenDocument
		6.c	valueURI	1-n	The URI of the subject term.		Not required for the Data Retention Project
(E06)	(Data Controller)	7	Contributor	0-n	The institution or person responsible for collecting, managing, distributing, or otherwise contributing to the development of the resource. To supply multiple contributors, repeat this property. For software, if there is an alternate entity that "holds, archives, publishes, prints, distributes, releases, issues, or produces" the code, use the contributorType "hostingInstitution" for the code repository.	Note: DataCite infrastructure supports up to between 800010000 names. For name lists above that size, consider attribution via linking to the related metadata. Examples: Charpy, Antoine; Foo Data Center	For the Data Retention Project the 'Contributor' element is optional but can be used to recognise an enabling facility of the Publisher, e.g. the name of a data repository within a University.
		7.a	contributorType	1	The type of contributor of the resource.	If Contributor is used, then contributorType is mandatory. Controlled List Values: ContactPerson DataCollector DataCurator DataManager Distributor Editor	

						HostingInstitution Producer ProjectLeader ProjectManager ProjectMember RegistrationAgency RegistrationAuthority RelatedPerson Researcher ResearchGroup RightsHolder Sponsor Supervisor WorkPackageLeader Other See Appendix for definitions.	
		7.1	contributorName	1	The full name of the contributor.	If Contributor is used, then contributorName is mandatory. Examples: Patel, Emily; ABC Foundation The personal name format may be: family, given. Non-roman names should be transliterated according to the ALA-LC schemas .	
		7.1.a	nameType	0-1	The type of name	Controlled List Values: Organizational Personal (default)	
		7.2	givenName	0-1	The personal or first name of the contributor.	Examples based on the 7.2 names: Emily	
		7.3	familyName	0-1	The surname or last name of the contributor.	Examples based on the 7.2 names: Patel	
		7.4	nameIdentifier	0-n	Uniquely identifies an individual or legal entity, according to various schemes.	The format is dependent upon scheme.	
		7.4.a	nameIdentifierScheme	1	The name of the name identifier scheme.	If nameIdentifier is used, nameIdentifierScheme is mandatory. Examples: ORCID17, ISNI18,	

[17] https://orcid.org/ When entering an ORCID, follow these style guidelines: https://orcid.org/content/journalarticle-display-guidelines

[18] http://www.isni.org/

						ROR19 , GRID20	
		7.4.b	schemeURI	0-1	The URI of the name identifier scheme.	Examples: http://www.isni.org/ http://orcid.org https://ror.org/ https://www.grid.ac/	
		7.56	affiliation	0-n	The organizational or institutional affiliation of the contributor.	Free text. The contributor's nameType may be Organizational or Personal. In case of an organizational contributor, e.g. a research group, you can add here the name of the	
						formal institution to which the contributor belongs.	
		7.5.a	affiliationIdentifier		Uniquely identifies the organizational affiliation of the contributor.	The format is dependent upon schema. Examples : https://ror.org/04aj4c181 grid.461819.3	
		7.5.b	affiliationIdentifierScheme	1	Name of the affiliation identifier schema.	If affiliationIdentifier is used, affiliationIdentifierScheme is mandatory. Examples : ROR, GRID	
		7.5.c	SchemeURI	0-1	URI of the affiliation identifier schema.	Examples : http://www.isni.org/ https:/orcid.org https://ror.org/	
		8	Date	0-n	Different dates relevant to the work.	YYYY,YYYY-MM-DD, YYYYMM-DDThh:mm:ssTZD or any other format or level of granularity described in W3CDTF21. Use RKMSISO860122 standard for depicting date ranges. Example: 2004-03-02/200506-02. Years before 0000 must be prefixed with a - sign, e.g. -0054 to indicate 55 BC.	For the Data Retention Project the 'Date' element is not required but can be used to record a significant date or date range.
		8.a	dateType	1	The type of date.	If Date is used, dateType is mandatory. Controlled List Values: Accepted Available	

[19] https://ror.org/

[20] https://www.grid.ac/

[21] https://www.w3.org/TR/NOTE-datetime

[22] The standard is documented here: http://www.ukoln.ac.uk/metadata/dcmi/collection-RKMS-ISO8601/

						Copyrighted Collected Created Issued Submitted Updated Valid Withdrawn Other See Appendix for definitions and recommendations.	
		8.b	dateInformation	0-1	Specific information about the date, if appropriate.	Free text. May be used to provide more information about the publication, release or collection date details, for example. May also be used to clarify dates in ancient history. Examples: 55 BC, 55 BCE.	
		9	Language	0-1	The primary language of the resource.	Allowed values are taken from IETF BCP 47, ISO 639-1 language codes. Examples: en, de, fr	Language is not required for the Data Retention Project
E02	Local Collection ID	11	alternateIdentifier	1	An identifier or identifiers other than the primary Identifier applied to the resource being registered. This may be any alphanumeric string which is unique within its domain of issue. May be used for local identifiers. AlternateIdentifier should be used for another identifier of the same instance (same location, same file).	Free text. *** Example: E-GEOD-34814	For the Data retention Project this will be a locally generated ID, not to be confused with RelatedIdentifier (12) which links other controlled and globally unique IDs
		11.a	alternateIdentifierType	1	The type of the AlternateIdentifier.	Free text. *** If alternateIdentifier is used, alternateIdentifierType is mandatory. For the above example, the alternateIdentifierType would be “A local accession number”	
V02 (V04)	Registry URL (Merit)	12	RelatedIdentifier	1-n	Identifiers of related resources. These must be globally unique identifiers.	Free text. For the Data Retention can be used to record RDA entry or data citation or other characteristic in the controlled list. *** Use this property to indicate subsets of properties, as appropriate.	This element will record the RDA URL or other URL that registers the data collection Can be used to record citation of the collection as evidence of merit, e.g. in addition to or instead of a grant number, or to record versioning/obsolescence or other important characteristics.

						Note: DataCite Event Data[23] collects all references to related resources based on the relatedIdentifier property.	
		12.a	relatedIdentifierType	1	The type of the RelatedIdentifier	If relatedIdentifier is used, relatedIdentifierType is mandatory. Controlled List Values: ARK arXiv bibcode DOI EAN13 EISSN Handle IGSN ISBN ISSN ISTC LISSN LSID PMID PURL UPC URL URN w3id See Appendix for full names and examples.	
		12.b	relationType	1	Description of the relationship of the resource being registered (A) and the related resource (B).	If RelatedIdentifier is used, relationType is mandatory. Controlled List Values: IsCitedBy Cites IsSupplementTo IsSupplementedBy IsContinuedBy Continues IsDescribedBy Describes HasMetadata IsMetadataFor HasVersion IsVersionOf IsNewVersionOf IsPreviousVersionOf IsPartOf	

[23] https://support.datacite.org/docs/eventdata-guide

						HasPart IsReferencedBy References IsDocumentedBy Documents IsCompiledBy Compiles IsVariantFormOf IsOriginalFormOf IsIdenticalTo IsReviewedBy Reviews IsDerivedFrom IsSourceOf IsRequiredBy Requires IsObsoletedBy Obsoletes See Appendix for definitions, examples and usage notes.	
		12.c	relatedMetadataScheme	0-1	The name of the scheme.	Use only with this relation pair: (HasMetadata/ IsMetadataFor) See Appendix for example.	
		12.d	schemeURI	0-1	The URI of the relatedMetadataScheme.	Use only with this relation pair: (HasMetadata/ IsMetadataFor) See Appendix for example	
		12.e	schemeType	0-1	The type of the relatedMetadataScheme, linked with the schemeURI.	Use only with this relation pair: (HasMetadata/ IsMetadataFor) Examples: XSD, DDT, Turtle	
		12.f	resourceTypeGeneral	0-1	The general type of the related resource.	Use the controlled list values as stated in 10.1. See Appendix for definitions, examples and usage notes.	
E01	Capacity	13	Size	1	Size (e.g. bytes, pages, inches, etc.) or duration (extent), e.g. hours, minutes, days, etc., of a resource.	Free text. *** Examples: "15 pages", "6 MB", “45 minutes”	For the Data retention Project we request size in TB
		14	Format	0-n	Technical format of the resource.	Free text. ***	

						Use file extension or MIME type where possible, e.g., PDF, XML, MPG or application/pdf, text/xml, video/mpeg.	
		15	Version	0-1	The version number of the resource.	Suggested practice: track major_version.minor_version. Register a new identifier for a major version change. Individual stewards need to determine which are major vs. minor versions[24]. Software engineering practice follows this approach of tracking changes and giving new version numbers. 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.	Not required for the Data Retention project but can be used to register versioning.
V05	License	16	Rights	1-n	Any rights information for this resource. The property may be repeated to record complex rights characteristics.	Free text. *** 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. May be used for software licenses. Examples: Creative Commons Attribution 3.0 Germany License Apache License, Version 2.0[25]	The ARDC recommends the Creative Commons licenses and waivers but requires the authority of the rights owner/s
		16.a	rightsURI	1	The URI of the license.	Example: http://creativecommons.org/licenses/by/3.0/de/deed.en	
		16.b	rightsIdentifier	1	A short, standardized version of the license name.	Example: CC-BY-3.0 Note: It's suggested to use the identifiers from the SPDX licence list (https://spdx.org/licenses/).	
		16.c	rightsIdentifierScheme	1	The name of the scheme.	Example: SPDX	
		16.d	schemeURI	1	The URI of the rightsIdentifierScheme.	Example:	

[24] Based on the work of the Earth Science Information Partners (ESIP). For more guidance, see: http://wiki.esipfed.org/index.php/Interagency_Data_Stewardship/Citations/provider_guidelines#Note_on_Versioning_and_Locators

[25] http://www.apache.org/licenses/

						https://spdx.org/licenses/	
E05	Description	17	Description	1-n	All additional information that does not fit in any of the other categories. May be used for technical information.	Free text. *** It is a best practice to supply a description.	
		17.a	descriptionType	1	The type of the Description.	If Description is used, descriptionType is mandatory. Controlled List Values: Abstract Methods SeriesInformation TableOfContents TechnicalInfo Other See Appendix for definitions.	
		18	GeoLocation	0-n	Spatial region or named place where the data was gathered or about which the data is focused.	Repeat this property to indicate several different locations.	
		18.1	geoLocationPoint	0-1	A point location in space.	A point contains a single longitude-latitude pair.	
		18.1.1	pointLongitude	1	Longitudinal dimension of point.	If geolocationPoint26is used, pointLongitude is mandatory. Longitude of the geographic point expressed in decimal degrees (positive east). Example: -67.302 Domain: -180 <= pointLongitude <= 180	
		18.1.2	pointLatitude	1	Latitudinal dimension of point.	If geolocationPoint27 is used, pointLatitude is mandatory. Latitude of the geographic point expressed in decimal degrees (positive north) Example: 31.233 Domain: -90<= pointLatitude <= 90	
		18.2	geoLocationBox	0-1	The spatial limits of a box.	A box is defined by two geographic points. Left low corner and right upper corner.	
						Each point is defined by its longitude and latitude.	
		18.2.1	westBoundLongitude	1	Western longitudinal dimension of box.	If geolocationBox27 is used westBoundLongitude is mandatory. Longitude of the geographic point expressed in decimal degrees (positive east).	

[26] Use WGS 84 (World Geodetic System) coordinates. Use only decimal numbers for coordinates. Longitudes are -180 to 180 (0 is Greenwich, negative numbers are west, positive numbers are east), Latitudes are -90 to 90 (0 is the equator; negative numbers are south, positive numbers north).

						Domain: -180.00 ≤ westBoundLongitude ≤ 180.00	
		18.2.2	eastBoundLongitude	1	Eastern longitudinal dimension of box.	If geolocationBox27 is used eastBoundLongitude is mandatory. Longitude of the geographic point expressed in decimal degrees (positive east) Domain: -180.00 ≤ eastBoundLongitude ≤ 180.00	
		18.2.3	southBoundLatitude	1	Southern latitudinal dimension of box.	If geolocationBox27 is used southBoundLatitude is mandatory. Latitude of the geographic point expressed in decimal degrees (positive north). Domain: -90.00 ≤ southBoundingLatitude ≤ 90.00	
		18.2.4	northBoundLatitude	1	Northern latitudinal dimension of box.	If geolocationBox27 is used northBoundLatitude is mandatory. Latitude of the geographic point expressed in decimal degrees (positive north). Domain: -90.00 ≤ northBoundingLatitude ≤ 90.00	
		18.3	geoLocationPlace	0-1	Description of a geographic location	Free text. Use to describe a geographic location.	
		18.4	geoLocationPolygon	0-n	A drawn polygon area, defined by a set of points and lines connecting the points in a closed chain.	A polygon is delimited by geographic points. Each point is defined by a longitudelatitude pair. The last point should be the same as the first point.	
		18.4.1	polygonPoint	4-n	A point location in a polygon.	If geoLocationPolygon27 is used, polygonPoint must be used as well. There must be at least 4 non-aligned points to make a closed curve, with the last point described the same as the first point.	
		18.4.1 .1	pointLongitude	1	Longitudinal dimension of point.	If polygonPoint is used pointLongitude is mandatory. Longitude of the geographic point expressed in decimal degrees (positive east). Domain: -180 <= pointLongitude <= 180	
		18.4.1 .2	pointLatitude	1	Latitudinal dimension of point.	If polygonPoint is used pointLatitude is mandatory. Latitude of the geographic point expressed in decimal degrees (positive north). Domain: -90<= pointLatitude <= 90	
		18.4.2	inPolygonPoint27	0-1	For any bound area that is larger than half the earth, define a (random) point inside.	inPolygonPoint is only necessary to indicate the "inside" of the polygon if the polygon is larger than half the earth. Otherwise the smallest of the two areas bounded by the polygon will be used.	

[27] A polygon that crosses the anti-meridian (i.e. the 180th meridian) can be represented by cutting it into two polygons such that neither crosses the anti-meridian.

		18.4.2.1	pointLongitude	1	Longitudinal dimension of point.	If inPolygonPoint30 is used pointLongitude is mandatory. Longitude of the geographic	
						point expressed in decimal degrees (positive east).	
		18.4.2.2	pointLatitude	1	Latitudinal dimension of point.	If inPolygonPoint is used, pointLatitude is mandatory. Latitude of the geographic point expressed in decimal degrees (positive north).	
V04	Merit	19	FundingReference	1-n	Information about financial support (funding) for the resource being registered.	It is a best practice to supply funding information when financial support has been received.	For the Data Retention Project a Grant ID is ideal but internal allocation reference can be accepted.
		19.1	funderName	1	Name of the funding provider.	Example: Gordon and Betty Moore Foundation	
		19.2	funderIdentifier	0-1	Uniquely identifies a funding entity, according to various types.	Example: https://doi.org/10.13039/100 000936	
		19.2.a	funderIdentifierType	0-1	The type of the funderIdentifier.	Controlled List Values: Crossref Funder ID[28] GRID ISNI ROR Other	
		19.2.b	SchemeURI	0-1	The URI of the funder identifier schema.	Examples: https://www.crossref.org/ser vices/funder-registry/ https://ror.org/	
		19.3	awardNumber	1	The code assigned by the funder to a sponsored award (grant).	Example: GBMF3859.01	
		19.3.a	awardURI	0-1	The URI leading to a page provided by the funder for more information about the award (grant).	Example: https://www.moore.org/grants/list/GBMF3859.01	
		19.4	awardTitle	1	The human readable title or name of the award (grant).	Example: Socioenvironmental Monitoring of the Amazon Basin and Xingu	The name of the award or allocation is required

[28] The Crossref service is called “Funder Registry” (https://www.crossref.org/services/funder-registry/) and Crossref Funder ID is the name for a Crossref identifier.

Controlled List Definitions

controlled list values that enhance the prospect that the resource's metadata will be found, cited and linked are indicated by shading.

contributorType

Description of contributorType

Option	*Description*	*Usage Notes*
ContactPerson	Person with knowledge of how to access, troubleshoot, or otherwise field issues related to the resource	May also be "Point of Contact" in organisation that controls access to the resource, if that organisation is different from Publisher, Distributor, Data Manager
DataCollector	Person/institution responsible for finding, gathering/collecting data under the guidelines of the author(s) or Principal Investigator (PI)	May also use when crediting survey conductors, interviewers, event or condition observers, person responsible for monitoring key instrument data.
DataCurator	Person tasked with reviewing, enhancing, cleaning, or standardizing metadata and the associated data submitted for storage, use, and maintenance within a data centre or repository	While the "DataManager" is concerned with digital maintenance, the DataCurator's role encompasses quality assurance focused on content and metadata. This includes checking whether the submitted dataset is complete, with all files and components as described by submitter, whether the metadata is standardized to appropriate systems and schema, whether specialized metadata is needed to add value and ensure access across disciplines, and determining how the metadata might map to search engines, database products, and automated feeds.
DataManager	Person (or organisation with a staff of data managers, such as a data centre) responsible for maintaining the finished resource.	The work done by this person or organisation ensures that the resource is periodically "refreshed" in terms of software/hardware support, is kept available or is protected from unauthorized access, is stored in accordance with industry standards, and is handled in accordance with the records management requirements applicable to it.
Distributor	Institution tasked with responsibility to generate/disseminate copies of the resource in either electronic or print form.	Works stored in more than one archive/repository may credit each as a distributor.
Editor	A person who oversees the details related to the publication format of the resource.	Note: if the Editor is to be credited in place of multiple creators, the Editor's name may be supplied as Creator, with "(Ed.)" appended to the name.
HostingInstitution	Typically, the organisation allowing the resource to be available on the internet through the provision of its hardware/software/operating support.	May also be used for an organisation that stores the data offline. Often a data centre (if that data centre is not the "publisher" of the resource.)
Producer	Typically a person or organisation responsible for the artistry and form of a media product.	In the data industry, this may be a company "producing" DVDs that package data for future dissemination by a distributor.
ProjectLeader	Person officially designated as head of project team or subproject team instrumental in the work necessary to development of the resource.	The Project Leader is not "removed" from the work that resulted in the resource; he or she remains intimately involved throughout the life of the particular project team.
ProjectManager	Person officially designated as manager of a project. Project may consist of one or many project teams and sub-teams.	The manager of a project normally has more administrative responsibility than actual work involvement.
ProjectMember	Person on the membership list of a designated project/project team.	This vocabulary may or may not indicate the quality, quantity, or substance of the person's involvement.
RegistrationAgency	Institution/organisation officially appointed by a Registration Authority to handle specific tasks within a defined area of responsibility.	DataCite is a Registration Agency for the International DOI Foundation (IDF). One of DataCite's tasks is to assign DOI prefixes to the allocating agents who then assign the full, specific character string to data clients, provide metadata back to the DataCite registry, etc.
RegistrationAuthority	A standards-setting body from which Registration Agencies obtain official recognition and guidance.	The IDF serves as the Registration Authority for the International Standards Organisation (ISO) in the area/domain of Digital Object Identifiers.
RelatedPerson	A person without a specifically defined role in the development of the resource, but who is someone the author wishes to recognize.	This person could be an author's intellectual mentor, a person providing intellectual leadership in the discipline or subject domain, etc.

Researcher	A person involved in analyzing data or the results of an experiment or formal study. May indicate an intern or assistant to one of the authors who helped with research but who was not so "key" as to be listed as an author.	Should be a person, not an institution. Note that a person involved in the gathering of data would fall under the contributorType "DataCollector." The researcher may find additional data online and correlate it to the data collected for the experiment or study, for example.
ResearchGroup	Typically refers to a group of individuals with a lab, department, or division; the group has a	May operate at a narrower level of scope; may or may not hold less administrative responsibility than a project team.
	particular, defined focus of activity.	
RightsHolder	Person or institution owning or managing property rights, including intellectual property rights over the resource.	
Sponsor	Person or organisation that issued a contract or under the auspices of which a work has been written, printed, published, developed, etc.	Includes organisations that provide in-kind support, through donation, provision of people or a facility or instrumentation necessary for the development of the resource, etc.
Supervisor	Designated administrator over one or more groups/teams working to produce a resource or over one or more steps of a development process.	
WorkPackageLeader	A Work Package is a recognized data product, not all of which is included in publication. The package, instead, may include notes, discarded documents, etc. The Work Package Leader is responsible for ensuring the comprehensive contents, versioning, and availability of the Work Package during the development of the resource.	
Other	Any person or institution making a significant contribution to the development and/or maintenance of the resource, but whose contribution does not "fit" other controlled vocabulary for contributorType.	Could be a photographer, artist, or writer whose contribution helped to publicize the resource (as opposed to creating it), a reviewer of the resource, someone providing administrative services to the author (such as depositing updates into an online repository, analysing usage, etc.), or one of many other roles.

dateType

NOTE: To indicate a date range, follow the RKMS-ISO8601 standard for depicting date ranges.

For example:

<date dateType="created">2012-03-01/2012-03-05</date>

Description of dateType

Option	*Description*	*Usage Notes*
Accepted	The date that the publisher accepted the resource into their system.	To indicate the start of an embargo period, use Submitted or Accepted, as appropriate.
Available	The date the resource is made publicly available. May be a range.	To indicate the end of an embargo period, use Available.
Copyrighted	The specific, documented date at which the resource receives a copyrighted status, if applicable.	
Collected	The date or date range in which the resource content was collected.	To indicate precise or particular timeframes in which research was conducted.
Created	The date the resource itself was put together; this could refer to a timeframe in ancient history, be a date range or a single date for a final component, e.g. the finalised file with all of the data.	Recommended for discovery.
Issued	The date that the resource is published or distributed e.g. to a data centre	
Submitted	The date the creator submits the resource to the publisher. This could be different from Accepted if the publisher then applies a selection process.	Recommended for discovery. To indicate the start of an embargo period, use Submitted or Accepted, as appropriate.
Updated	The date of the last update to the resource, when the resource is being added to. May be a range.	
Valid	The date or date range during which the dataset or resource is accurate.	
Withdrawn	The date the resource is removed.	It's good practice to indicate the reason for retraction or withdrawal in the descriptionType.

resourceTypeGeneral

Description of resourceTypeGeneral

Option	*Description*[29]	*Examples and Usage Notes*	*Suggested Dublin Core Mapping*
Audiovisual	A series of visual representations imparting an impression of motion when shown in succession. May or may not include sound.	May be used for films, video, etc, Example: https://data.datacite.org/applicatio n/vnd.datacite.datacite+xml/10.17608/k6.auckland.4620790.v1	MovingImage
Collection	An aggregation of resources, which may encompass collections of one resourceType as well as those of mixed types. A collection is described as a group; its parts may also be separately described.	A collection of samples, or various files making up a report. Example: https://data.datacite.org/applicatio n/vnd.datacite.datacite+xml/10.1594/pangaea.877589	Collection
DataPaper	A factual and objective publication with a focused intent to identify and describe specific data, sets of data, or data collections to facilitate discoverability.	A data paper describes data provenance and methodologies used in the gathering, processing, organizing, and representing the data. Example: https://data.datacite.org/application/vnd.datacite.datacite+xml/10.17912/w2mw2d	Text
Dataset	Data encoded in a defined structure.	Data file or files. Example: https://data.datacite.org/applicatio n/vnd.datacite.datacite+xml/10.1594/pangaea.804876	Dataset
Event	A non-persistent, timebased occurrence.	Descriptive information and/or content that is the basis for discovery of the purpose, location, duration, and responsible agents associated with an event such as a webcast or convention. Example: https://data.datacite.org/applicatio n/vnd.datacite.datacite+xml/10.7269/p3rn35sz	Event
Image	A visual representation other than text.	Digitised or born digital images, drawings or photographs. Example: https://data.datacite.org/applicatio n/vnd.datacite.datacite+xml/10.6083/m4qn65c5	Image, StillImage
InteractiveResource	A resource requiring interaction from the user to be understood, executed, or experienced	Training modules, files that require use of a viewer (e.g., Flash), or query/response portals. Example: https://data.datacite.org/applicatio n/vnd.datacite.datacite+xml/10.7269/p3tb14tr	InteractiveResource
Model	An abstract, conceptual, graphical, mathematical or visualization model that represents empirical objects, phenomena, or physical processes.	Modelled descriptions of, for example, different aspects of languages or a molecular biology reaction chain. Example: https://data.datacite.org/applicatio n/vnd.datacite.datacite+xml/10.5285/4d866cd2-c907-4ce2-b070084ca9779dc2	N/A

[29] Where there is direct correspondence with the Dublin Core Metadata, DataCite definitions have borrowed liberally from the DCMI definitions. See: http://dublincore.org/documents/dcmi-terms/index.shtml

Option	*Description*[30]	*Examples and Usage Notes*	*Suggested Dublin Core Mapping*
PhysicalObject	An inanimate, threedimensional object or substance.	Artifacts, specimens. Example: https://data.datacite.org/applicatio n/vnd.datacite.datacite+xml/10.72 99/X78052RB	PhysicalObject
Service	An organized system of apparatus, appliances, staff, etc., for supplying some function(s) required by end users.	Data management service, or longterm preservation service. Example: https://data.datacite.org/applicatio n/vnd.datacite.datacite+xml/10.21 938/3I01ISNUCODNH1ZJBCVUWA	Service
Software	A computer program in source code (text) or compiled form. Use this type for all software components supporting scholarly research.	Software supporting scholarly research. Example: https://data.datacite.org/applicatio n/vnd.datacite.datacite+xml/10.42 25/03/5954F738EE5AA	Software
Sound	A resource primarily intended to be heard.	Audio recording. Example: https://data.datacite.org/applicatio n/vnd.datacite.datacite+xml/10.72 82/T3J67F05	Sound
Text	A resource consisting primarily of words for reading.	Grey literature, lab notes, accompanying materials, data management plan, conference poster. Example: https://data.datacite.org/applicatio n/vnd.datacite.datacite+xml/10.56 82/9786065914018	Text
Workflow	A structured series of steps which can be executed to produce a final outcome, allowing users a means to specify and enact their work in a more reproducible manner.	Computational workflows involving sequential operations made on data by wrapped software and may be specified in a format belonging to a workflow management system, such as Taverna	N/A
		(http://www.taverna.org.uk/). More.[30]	
Other	If selected, supply a value for ResourceType.		

[30] An education module on workflows prepared by DataONE is available at http://www.dataone.org/sites/all/documents/L10_AnalysisWorkflows.pptx

relatedIdentifierType

Description of relatedIdentifierType

Option	*Full Name*	*Example*
ARK	Archival Resource Key; URL designed to support long-term access to information objects. In general, ARK syntax is of the form (brackets indicate [optional] elements: [http://*NMA*/]ark:/*NAAN*/*Name* [*Qualifier*]	<relatedIdentifier relatedIdentifierType="ARK" relationType="IsCitedBy">ark:/13030/tqb3kh97gh8w </relatedIdentifier>
arXiv	arXiv identifier; arXiv.org is a repository of preprints of scientific papers in the fields of mathematics, physics, astronomy, computer science, quantitative biology, statistics, and quantitative finance.	<relatedIdentifier relatedIdentifierType="arXiv" relationType="IsCitedBy">arXiv:0706.0001 </relatedIdentifier>
bibcode	Astrophysics Data System bibliographic codes; a standardized 19 character identifier according to the syntax yyyyjjjjjvvvvmppppa. See http://infouri.info/registry/OAIHandler?verb=GetRecord&metadataPrefix= reg&identifier=info:bibcode/	<relatedIdentifier relatedIdentifierType="bibcode" relationType="IsCitedBy"> 2014Wthr...69...72C </relatedIdentifier> Note: bibcodes can be resolved via http://adsabs.harvard.edu/abs/bibcode
DOI	Digital Object Identifier; a character string used to uniquely identify an object. A DOI name is divided into two parts, a prefix and a suffix, separated by a slash.	<relatedIdentifier relatedIdentifierType="DOI" relationType="IsSupplementTo"> 10.1016/j.epsl.2011.11.037 </relatedIdentifier>
EAN13	European Article Number, now renamed International Article Number, but retaining the original acronym, is a 13-digit barcoding standard which is a superset of the original 12-digit Universal Product Code (UPC) system.	<relatedIdentifier relatedIdentifierType="EAN13" relationType="Cites">9783468111242 </relatedIdentifier>
EISSN	Electronic International Standard Serial Number; ISSN used to identify periodicals in electronic form (eISSN or eISSN).	<relatedIdentifier relatedIdentifierType="eISSN" relationType="Cites">1562-6865 </relatedIdenfifier>
Handle	A handle is an abstract reference to a resource.	<relatedIdentifier relatedIdentifierType="Handle" relationType="References">10013/epic.10033 </relatedIdentifier>
IGSN	International Geo Sample Number; a 9-digit alphanumeric code that uniquely identifies samples from our natural environment and related sampling features.	<relatedIdentifier relatedIdentifierType="IGSN" relationType="References">IECUR0097 </relatedIdentifier>
ISBN	International Standard Book Number; a unique numeric book identifier. There are 2 formats: a 10-digit ISBN format and a 13digit ISBN.	<relatedIdentifier><relatedIdentifier relatedIdentifierType="ISBN" relationType="IsPartOf">978-3-905673-82-1 </relatedIdentifier>
ISSN	International Standard Serial Number; a unique 8-digit number used to identify a print or electronic periodical publication.	<relatedIdentifier relatedIdentifierType="ISSN" relationType="IsPartOf">0077-5606 </relatedIdentifier>

ISTC	International Standard Text Code; a unique "number" assigned to a textual work. An ISTC consists of 16 numbers and/or letters.	<relatedIdentifier relatedIdentifierType="ISTC" relationType="Cites">0A9 2002 12B4A105 7 </relatedIdentifier>
LISSN	The linking ISSN or ISSN-L enables collocation or linking among different media versions of a continuing resource.	<relatedIdentifier relatedIdentifierType="LISSN" relationType="Cites">1188-1534</relatedIdentifier>
LSID	Life Science Identifiers; a unique identifier for data in the Life Science domain. Format: urn:lsid:authority:namespace:id entifier:revision	<relatedIdentifier relatedIdentifierType="LSID" relationType="Cites"> urn:lsid:ubio.org:namebank:11815</relatedIdentifier>
PMID	PubMed identifier; a unique number assigned to each PubMed record.	<relatedIdentifier relatedIdentifierType="PMID" relationType="IsReferencedBy">12082125</relatedId entifier>
PURL	Persistent Uniform Resource Locator. A PURL has three parts: (1) a *protocol*, (2) a *resolver address*, and (3) a *name*.	<relatedIdentifier relatedIdentifierType="PURL" relationType="Cites"> http://purl.oclc.org/foo/bar</relatedIdentifier>
UPC	Universal Product Code is a barcode symbology used for tracking trade items in stores. Its most common form, the UPC-A, consists of 12 numerical digits.	<relatedIdentifier relatedIdentifierType="UPC" relationType="Cites"> 123456789999</relatedIdentifier>
URL	Uniform Resource Locator, also known as web address, is a specific character string that constitutes a reference to a resource. The syntax is: schema://domain:port/path?qu ery_string#fragment_id	<relatedIdentifier relatedIdentifierType="URL" relationType="IsCitedBy">http://www.heatflow.und.edu/i ndex2.html</relatedIdentifier>
URN	Uniform Resource Name; is a unique and persistent identifier of an electronic document. The syntax is: urn:< NID>:<NSS> The leading urn : sequence is case-insensitive, <NID> is the namespace identifier, <NSS> is the namespace-specific string.	<relatedIdentifier relatedIdentifierType="URN" relationType="IsSupplementTo">urn:nbn:de:101:1201102033592</relatedIdentifier>
w3id	Permanent identifier for Web applications. Mostly used to publish vocabularies and ontologies. The letters 'w3' stand for "World Wide Web".	<relatedIdentifier relatedIdentifierType="w3id" relationType="IsCitedBy">https://w3id.org/games/spec/co il#Coil_Bomb_Die_Of_Age</relatedIdentifier>

relationType

Description of the relationship of the resource being registered (A) and the related resource (B).

Description of relationType

Option	*Definition*	*Example and Usage Notes*
IsCitedBy	indicates that B includes A in a citation	Recommended for discovery. <relatedIdentifier relatedIdentifierType="DOI"relationType="IsCited By">10.4232/10.ASEAS-5.2-1 </relatedIdentifier>
Cites	indicates that A includes B in a citation	Recommended for discovery. <relatedIdentifier relatedIdentifierType="ISBN" relationType="Cites">0761964312 </relatedIdentifier>
IsSupplementTo	indicates that A is a supplement to B	Recommended for discovery. <relatedIdentifier relatedIdentifierType="URN" relationType="IsSupplementTo">urn:nbn:de:0168-ssoar13172 </relatedIdentifier>
IsSupplementedBy	indicates that B is a supplement to A	Recommended for discovery. <relatedIdentifier relatedIdentifierType="PMID" relationType="IsSupplementedBy">16911322/ </relatedIdentifier>
IsContinuedBy	indicates A is continued by the work B	<relatedIdentifier relatedIdentifierType="URN" relationType="IsContinuedBy">urn:nbn:de:bsz:21-opus4967 </relatedIdentifier>
Continues	indicates A is a continuation of the work B	<relatedIdentifier relatedIdentifierType="URN" relationType="Continues">urn:nbn:de:bsz:21-opus-4966 </relatedIdentifier>
Describes	indicates A describes B	<relatedIdentifier relatedIdentifierType="DOI" relationType="Describes">10.6084/m9.figshare.c.3288407< /relatedIdentifier>
IsDescribedBy	indicates A is described by B	<relatedIdentifier relatedIdentifierType="DOI" relationType="IsDescribedBy">10.1038/sdata.2016.123</r elatedIdentifier>
HasMetadata	indicates resource A has additional metadata B	<relatedIdentifier relatedIdentifierType="DOI" relationType="HasMetadata" relatedMetadataSchema="DDI-L" schemeURI="http://www.ddialliance.org/Specification/DDILifecycle/3.1/XMLSchema/instance.xsd">10.1234/567890</rel atedIdentifier>
IsMetadataFor	indicates additional metadata A for a resource B	<relatedIdentifier relatedIdentifierType="DOI" relationType="IsMetadataFor "relatedMetadataSchema="DDI-L" schemeURI="http://www.ddialliance.org/Specification/DDILifecycle/3.1/XMLSchema/instance.xsd">10.1234/567891</rel atedIdentifier>

HasVersion	indicates A has a version (B)	The registered resource such as a software package or code repository has a versioned instance (indicates A has the instance B) e.g. it may be used to relate an un-versioned code repository to one of its specific software versions. <relatedIdentifier relatedIdentifierType="DOI" relationType="HasVersion">10.5281/ZENODO.832053 </relatedIdentifier>
IsVersionOf	indicates A is a version of B	The registered resource is an instance of a target resource (indicates that A is an instance of B) e.g. it may be used to relate a specific version of a software package to its software code repository. <relatedIdentifier relatedIdentifierType="DOI" relationType="IsVersionOf">10.5281/ZENODO.832054 </relatedIdentifier>
IsNewVersionOf	indicates A is a new edition of B, where the new edition has been modified or updated	<relatedIdentifier relatedIdentifierType="DOI" relationType="IsNewVersionOf">10.5438/0005 </relatedIdentifier>
IsPreviousVersion Of	indicates A is a previous edition of B	<relatedIdentifier relatedIdentifierType="DOI" relationType="IsPreviousVersionOf">10.5438/0007 </relatedIdentifier>
IsPartOf	indicates A is a portion of B; may be used for elements of a series	Primarily this relation is applied to container-contained type relationships. Note: May be used for individual software modules; note that code repository-to-version relationships should be modeled using IsVersionOf and HasVersion Recommended for discovery. <relatedIdentifier relatedIdentifierType="DOI" relationType="IsPartOf">10.5281/zenodo.754312 </relatedIdentifier>
HasPart	indicates A includes the part B	Primarily this relation is applied to container-contained type relationships. Note: May be used for individual software modules; note that code repository-to-version relationships should be modeled using IsVersionOf and HasVersion Recommended for discovery. <relatedIdentifier relatedIdentifierType="URL" relationType="HasPart">https://zenodo.org/record/16564/f iles/dune-stuff-LSSC_15.zip</relatedIdentifier>
IsReferencedBy	indicates A is used as a source of information by B	<relatedIdentifier relatedIdentifierType="URL" relationType="IsReferencedBy">http://www.testpubl.de </relatedIdentifier>
References	indicates B is used as a source of information for A	<relatedIdentifier relatedIdentifierType="URN" relationType="References">urn:nbn:de:bsz:21-opus963</relatedIdentifier>
IsDocumentedBy	indicates B is documentation about/ explaining A; e.g. points to software documentation	<relatedIdentifier relatedIdentifierType="URL" relationType="IsDocumentedBy">http://tobias-lib.unituebingen.de/volltexte/2000/96/ </relatedIdentifier>
Documents	indicates A is documentation about B; e.g. points to software documentation	<relatedIdentifier relatedIdentifierType="DOI" relationType="Documents">10.1234/7836 </relatedIdentifier>
IsCompiledBy	indicates B is used to compile or create A	<relatedIdentifier relatedIdentifierType="URL" relationType="isCompiledBy">http://dnb.info/gnd/4513749-3 </relatedIdentifier>

		Note: may be used for software and text, as a compiler can be a computer program or a person.
Compiles	indicates B is the result of a compile or creation event using A	<relatedIdentifier relatedIdentifierType="URN" relationType="Compiles">urn:nbn:de:bsz:21-opus-963</relatedIdentifier> Note: may be used for software and text, as a compiler can be a computer program or a person.
IsVariantFormOf	indicates A is a variant or different form of B	<relatedIdentifier relatedIdentifierType="DOI" relationType="IsVariantFormOf">10.1234/8675</relatedIdentifier> Use for a different form of one thing. May be used for different software operating systems or compiler formats, for example.
IsOriginalFormOf	indicates A is the original form of B	<relatedIdentifier relatedIdentifierType="DOI" relationType="IsOriginalFormOf">10.1234/9035</relatedIdentifier> May be used for different software operating systems or compiler formats, for example.
IsIdenticalTo	indicates that A is identical to B, for use when there is a need to register two separate instances of the same resource	<relatedIdentifier relatedIdentifierType="URL" relationType="IsIdenticalTo">http://oac.cdlib.org/findaid/ar k:/13030/c8r78fzq </relatedIdentifier> IsIdenticalTo should be used for a resource that is the same as the registered resource but is saved on another location, maybe another institution.
IsReviewedBy	indicates that A is reviewed by B	<relatedIdentifier relatedIdentifierType="DOI" relationType="IsReviewedBy">10.5256/F1000RESEARCH.4288.R4745</relatedIdentifier>
Reviews	indicates that A is a review of B	<relatedIdentifier relatedIdentifierType="DOI" relationType="Reviews">10.12688/f1000research.4001.1</relatedIdentifier>
IsDerivedFrom	indicates B is a source upon which A is based	<relatedIdentifier relatedIdentifierType="DOI" relationType="IsDerivedFrom">10.6078/M7DZ067C</relatedIdentifier> IsDerivedFrom should be used for a resource that is a derivative of an original resource. In this example, the dataset is derived from a larger dataset and data values have been manipulated from their original state.
IsSourceOf	indicates A is a source upon which B is based	<relatedIdentifier relatedIdentifierType="URL" relationType="IsSourceOf"> http://opencontext.org/projects/81204AF8-127C-4686-E9B01202C3A47959 </relatedIdentifier> IsSourceOf is the original resource from which a derivative resource was created. In this example, this is the original dataset without value manipulation, and the source of the derived dataset.
IsRequiredBy	Indicates A is required by B	<relatedIdentifier relatedIdentifierType="DOI"

		relationType="IsRequiredBy">10.1234/8675 </relatedIdentifier> Note: May be used to indicate software dependencies.
Requires	Indicates A requires B	<relatedIdentifier relatedIdentifierType="DOI" relationType="Requires">10.1234/8675 </relatedIdentifier> Note: May be used to indicate software dependencies.
Obsoletes	Indicates A replaces B	<relatedIdentifier relatedIdentifierType="DOI" relationType="Obsoletes">10.5438/0007 </relatedIdentifier>
IsObsoletedBy	Indicates A is replaced by B	<relatedIdentifier relatedIdentifierType="DOI" relationType="IsObsoletedBy">10.5438/0005 </relatedIdentifier>

descriptionType

Description of descriptionType

Option	*Definition*	*Usage Notes*
Abstract	A brief description of the resource and the context in which the resource was created.	Recommended for discovery. Use " " to indicate a line break for improved rendering of multiple paragraphs, but otherwise no html markup. Example: https://data.datacite.org/application/vnd.datacite.datacit e+xml/10.1594/PANGAEA.771774
Methods	The methodology employed for the study or research.	Recommended for discovery. Example: https://data.datacite.org/application/vnd.datacite.datacit e+xml/10.6078/D1K01X
SeriesInformation	Information about a repeating series, such as volume, issue, number.	For use with grey literature. If providing an ISSN, use property 12 (RelatedIdentifier), relatedIdentifierType=ISSN. For dataset series, use property 12 (RelatedIdentifier) and describe the relationships with isPartOf or HasPart. Example: https://data.datacite.org/application/vnd.datacite.datacit e+xml/10.4229/23RDEUPVSEC2008-5CO.8.3
TableOfContents	A listing of the Table of Contents.	Use " " to indicate a line break for improved rendering of multiple paragraphs, but otherwise no html markup. Example: https://data.datacite.org/application/vnd.datacite.datacit e+xml/10.5678/LCRS/FOR816.CIT.1031

TechnicalInfo	Detailed information that may be associated with design, implementation, operation, use, and/or maintenance of a process or system.	For software description, this may include the contents of a readme.txt, and necessary environmental information (hardware, operational software, applications/programs with version information, a human-readable synopsis of software purpose) that cannot be described using other properties (e.g. Language (software)). For other uses, this can include specific and detailed information as necessary and appropriate.
Other	Other description information that does not fit into an existing category.	Use for any other description type.

Standard values for unknown information

Below provides a set of standard values that may be used when mandatory property values are not available for various reasons.

Examples of usage:

```
<creatorName>:unkn</creatorName>

<title>:unas</title>

<publisher>:null</publisher>
```

Table 11: Standard values for unknown information

Code	***Definition***
:unac	temporarily inaccessible
:unal	unallowed, suppressed intentionally
:unap	not applicable, makes no sense
:unas	value unassigned (e.g., Untitled)
:unav	value unavailable, possibly unknown
:unkn	known to be unknown (e.g., Anonymous, Inconnue)
:none	never had a value, never will
:null	explicitly and meaningfully empty
:tba	to be assigned or announced later
:etal	too numerous to list (et alia)