UK Research Data Registry Mapping Schemes

Project Information			
Project Identifier	PID TBC		
Project Title	UK Research Data (Metadata) Registry Pilot		
Project Hashtag	#jiscrdr		
Start Date	1 October 2013	**End Date**	31 March 2014
Lead Institution	Jisc		
Project Director	Rachel Bruce		
Project Manager	—		
Contact Email	—		
Partner Institutions	Digital Curation Centre (Universities of Edinburgh, Glasgow, Bath); UKDA (University of Essex)		
Project Webpage URL	http://www.dcc.ac.uk/projects/research-data-registry-pilot		
Programme Name	Jisc Capital Programme		

Document Information			
Author(s) and Role(s)	Alex Ball (Metadata Coordinator)		
Date	12 May 2014	**Project Refs**	T3.1; D3.1
Filename	uk-rdr-mapping-v09.pdf		
URL	http://www.dcc.ac.uk/sites/default/files/documents/registry/uk-rdr-mapping-v09.pdf		
Access	This report is for general dissemination		

Document History		
Version	**Date**	**Comments**
01	29 November 2013	Summary of RIF-CS. First draft of DDI mapping.
02	8 January 2014	Second draft of DDI mapping. First draft of DataCite mapping.
03	22 January 2014	New introduction. Expanded summary of RIF-CS. First draft of EPrints ReCollect, NERC Discovery, and OAI-PMH Dublin Core mappings.
04	23 January 2014	Section on tips for contributors.
05	31 January 2014	Second draft of NERC Discovery (UK GEMINI), EPrints mappings.
06	12 February 2014	Third draft of DDI mapping.
07	18 March 2014	Corrections to all mappings in the light of codification.
08	21 March 2014	First draft of MODS mapping. Registry quality levels included.
09	9 May 2014	Fixed small errors. Added MODS elements to tips section.

Contents

1 Introduction

This report documents the mappings created for importing metadata into the pilot UK Research Data Registry implementation. As the priority for the pilot is for populating the registry rather than surfacing content elsewhere, the mappings are all one-way, and convert metadata into the registry's internal data format, described in Chapter 2.

The target metadata schemas were determined by the capabilities of the project collaborators, i.e. those organizations who volunteered to provide the project with test metadata.

- The UK Data Archive provided records using the DDI standard, version 2.5.
- The NERC Data Catalogue Service provided records using UK GEMINI, version 2.2.
- The Archaeology Data Service provided records via the NERC Data Catalogue Service and DataCite.
- The University of Edinburgh (using DSpace) provided records using MODS.
- The University of Glasgow provided records using an adapted EPrints metadata profile.
- The University of Oxford (using DataBank) provided records via DataCite.
- Oxford Brookes University provided records using OAI-PMH Dublin Core.
- The University of Lincoln provided records using OAI-PMH Dublin Core.
- The University of Leeds provided records using an adapted EPrints metadata profile.
- The University of St Andrews provided records using MODS.
- The University of Southampton provided records using an adapted EPrints metadata profile.
- The University of Hull provided records using MODS.

1.1 Typographical conventions

In the mappings presented in this report, nesting levels are indicated using angle brackets, while attributes are presented in square brackets. For example, the full path of the type of citation date in a RIF-CS record would be written as registryObjects > registryObject > collection > citationInfo > citationMetadata > date[TYPE], though for brevity the beginning of the path may be omitted.

If the value of an attribute is relevant to identifying an element, or would be fixed for given instance of an element, it is presented alongside the attribute name with an equals sign, e.g. date[TYPE=created].

2

RIF-CS

The ANDS data registry software, used in the initial pilot phase of the UK Research Data Registry project, uses RIF-CS as its internal data model.[1] It is a profile of ISO 2146, tailored to the needs of a collection service registry.

A summary of the RIF-CS standard follows. It is presented as a structured list of elements and their attributes. Indentation in the list corresponds to the nesting of elements, so for example if an element contains two sub-elements, these would be displayed after the element at one greater level of indentation. Attributes immediately follow the elements to which they belong.

Each item in the list has the following structure.

- An expression in square brackets indicates how often the item should occur in a conforming instance of the metadata scheme. An element is represented by two numbers: a minimum and a maximum number of occurrences. An 'n' indicates that there is no upper bound to the number of times an element may occur. An attribute is either optional ('O') or mandatory ('M')
- The item name is given: element names in a sans serif typeface, attribute names in SANS SERIF SMALL CAPITALS.
- If the item has special formatting rules, this is given in roman type after the name.
  - 'encoded' means that the content of the element or the value of the attribute must conform to a particular syntax; where possible the syntax is given in parentheses.
  - 'controlled' means that the content of the element or the value of the attribute must be one of a number of pre-defined terms; where possible the authority list of terms is given in a footnote. Definitions of many of these terms are provided in Section 2.2.
- Where the meaning or usage of the item is not clear, additional information is provided *in italics*.

2.1 Elements

[1,1] registryObjects

· [0,n] registryObject

└ [M] GROUP – *Name identifying the organisation contributing the object's metadata.*

1. 'Registry Interchange Format – Collections and Services (RIF-CS) v1.5: Schema Guidelines', Australian National Data Service ⟨http://services.ands.org.au/documentation/rifcs/1.5/guidelines/rif-cs.html⟩, accessed 22 Nov. 2013.

- [1,1] key – *Unique identifier (used by Registry).*
- [1,1] originatingSource – *Identifying string or URI of the entity holding the managed version of the registry object metadata.*
  - [O] TYPE: controlled ('authoritative')
- [1,1] *one of*
  - activity
    - [M] TYPE: controlled (see Section 2.2.2)
    - [O] DATEMODIFIED: encoded (W3C dateTime,[2] UTC) – *Refers to the object's metadata.*
    - [0,n] identifier – *May repeat key or supply alternative identifiers.*
      - [M] TYPE: controlled (see Section 2.2.6)
    - [0,n] name
      - [O] TYPE: controlled (see Section 2.2.7) – *Used to distinguish primary and alternative versions of a name.*
      - [O] DATEFROM: encoded (W3C dateTime, UTC)
      - [O] DATETO: encoded (W3C dateTime, UTC)
      - [1,n] namePart – *Can be used for a whole name or for an element of a name.*
        - [O] TYPE: controlled (see Section 2.2.8)
    - [0,n] location – *Not used for recording the repository holding a collection, nor for data coverage.*
      - [O] TYPE – *Not currently used.*
      - [O] DATEFROM: encoded (W3C dateTime, UTC)
      - [O] DATETO: encoded (W3C dateTime, UTC)
      - [0,n] address
        - [0,n] electronic
          - [O] TYPE: controlled (see Section 2.2.12)
          - [1,1] value: encoded (URI) – *URI of collection, activity or party; base URL for HTTP service point; URL of WSDL file.*
          - [0,n] arg – *Argument for an electronic service (not used for activities, collections or parties).*
            - [M] REQUIRED: encoded (Boolean)
            - [M] TYPE: controlled (see Section 2.2.13)
            - [M] USE: controlled (see Section 2.2.14)
        - [0,n] physical
          - [O] TYPE: controlled (see Section 2.2.15)
          - [1,n] addressPart – *May contain full address or a meaningful fragment (e.g. postcode).*
            - [M] TYPE: controlled (see Section 2.2.16)
      - [0,n] spatial – *Geographical address information.*
        - [M] TYPE: controlled (see Section 2.2.17)
    - [0,n] relatedObject – *Used for other objects in the Registry (c.f. relatedInfo).*
      - [1,1] key – *Unique identifier used by Registry.*

2. http://www.w3.org/TR/xmlschema-2/#dateTime

- [1,n] relation
  - [M] TYPE: controlled (see Sections 2.2.18 to 2.2.21)
  - [0,1] description – *Required for 'hasAssociationWith' relations.*
  - [0,1] url: encoded (URL)
- [0,n] subject – *Keyword indicating field of activity, subject matter of collection, research interest of party.*
  - [M] TYPE: controlled[3] – *Name of controlled vocabulary used, if any.*
  - [O] TERMIDENTIFIER: encoded (URL) – *URL identifying the term.*
- [0,n] description
  - [M] TYPE: controlled (see Section 2.2.23)
- [0,n] coverage
  - [0,n] temporal
    - [0,n] date: encoded
      - [M] TYPE: controlled (see Section 2.2.10)
      - [M] DATEFORMAT: controlled ('W3CDTF')
    - [0,n] text – *Used if exact dates are not appropriate.*
  - [0,n] spatial – *Geographical coverage, e.g. co-ordinates, region information.*
    - [M] TYPE: controlled (see Section 2.2.17)
- [0,n] relatedInfo – *Used for objects outside the Registry (c.f. relatedObject).*
  - [O] TYPE: controlled (see Section 2.2.24)
  - [1,n] identifier
    - [M] TYPE: controlled (see Section 2.2.25)
  - [0,n] relation
    - [M] TYPE: controlled (see Sections 2.2.18 to 2.2.21)
    - [0,1] description – *Required for 'hasAssociationWith' relations.*
    - [0,1] url: encoded (URL) – *Can be used to record URL implementing a relationship, e.g. for a collection, the URL implementing the related service.*
  - [0,1] format
    - [1,n] identifier
      - [M] TYPE: controlled (see Section 2.2.25)
  - [0,1] title
  - [0,1] notes
- [0,n] rights
  - [0,1] rightsStatement – *IPR statement, licence, access rights/constraints statement.*
    - [O] RIGHTSURI: encoded (URI) – *Used in place of free text content.*
  - [0,1] licence – *Text of legal document.*
    - [O] TYPE: controlled (see Section 2.2.26)

3. A term from the Library of Congress' *Source Codes for Vocabularies, Rules, and Schemes* ⟨http://www.loc.gov/standards/sourcelist/⟩, or 'local' (indicating a controlled vocabulary not in the list, or an uncontrolled vocabulary). The UK registry has defined some additional terms: see Section 2.2.22.

└ [O] RIGHTSURI: encoded (URI) – *Used in place of free text content.*

· · · · · [0,1] accessRights – *Description of access rights and constraints.*

└ [O] TYPE: controlled

└ [O] RIGHTSURI: encoded (URI) – *Used in place of free text content.*

· · · · [0,n] existenceDates – *Refers to existence of registry object.*

· · · · · [0,1] startDate: encoded

└ [M] DATEFORMAT: controlled ('W3CDTF')

· · · · · [0,1] endDate: encoded

└ [M] DATEFORMAT: controlled ('W3CDTF')

· · · collection

└ [M] TYPE: controlled (see Section 2.2.3)

└ [O] DATEACCESSIONED: encoded (W3C dateTime, UTC)

└ [O] DATEMODIFIED: encoded (W3C dateTime, UTC) – *Refers to the object's metadata.*

As activity plus…

· · · · [0,n] dates – *(Typically provided after name and before location).*

└ [M] TYPE: controlled (see Section 2.2.9)

· · · · · [1,n] date: encoded

└ [M] TYPE: controlled (see Section 2.2.10)

└ [M] DATEFORMAT: controlled ('W3CDTF')

· · · · [0,n] citationInfo – *Used to provide a sample bibliographic reference.*

· · · · · [0,1] fullCitation – *Pre-formatted reference.*

└ [O] STYLE: controlled (see Section 2.2.27)

· · · · · [0,1] citationMetadata – *Elements from which to construct a reference.*

· · · · · · [1,1] identifier

└ [M] TYPE: controlled (see Section 2.2.28)

· · · · · · [1,n] contributor

└ [O] SEQ: encoded (integer) – *If order matters, the position of the contributor in the sequence.*

· · · · · · · [1,n] namePart – *Can be used for a whole name or for an element of a name.*

└ [O] TYPE: controlled (see Section 2.2.8)

· · · · · · [1,1] title

· · · · · · [0,1] version

· · · · · · [1,1] publisher

· · · · · · [0,1] placePublished

· · · · · · [1,n] date: encoded (W3C dateTime, UTC)

└ [M] TYPE: controlled (see Section 2.2.29)

· · · · · · [0,1] url: encoded (URL)

· · · · · · [0,1] context – *e.g. series title, database name.*

· · · party

└ [M] TYPE: controlled (see Section 2.2.4)

└ [O] DATEMODIFIED: encoded (W3C dateTime, UTC) – *Refers to the object's metadata.*

As activity

· · · service

└ [M] TYPE: controlled (see Section 2.2.5)

└ [O] DATEMODIFIED: encoded (W3C dateTime, UTC) – *Refers to the object's metadata.*

As activity plus…

· · · · [0,n] accessPolicy: encoded (URL) – *Web-accessible description of service access policies.*

2.2 Controlled vocabularies

The following represent the controlled terms defined for use in RIF-CS 2.5,[4] and are reproduced here for convenience. The headings indicate where, within a registryObject element, the vocabularies are used (vertical bars indicate alternatives at that part of the path hierarchy). They are presented in the order in which they would be first encountered in an XML record, if the element order in the previous section were used.

Substantive deviations from the published definitions have been marked in italic. These are intended as corrections or clarifications rather than local variations.

2.2.1 originatingSource[type]

authoritative The source holds the authoritative version of the metadata about the registry object.

2.2.2 activity[type]

award Something given to recognize excellence in a certain field.

course Education imparted in a series of lessons or meetings.

event Something that happens at a particular place or time as an organized activity with participants or an audience.

program System of activities intended to meet a public need.

project Piece of work that is undertaken or attempted, with a start and end date and defined objectives.

2.2.3 collection[type]

catalogueOrIndex Collection of resource descriptions describing the content of one or more repositories or collective works.

collection Compiled content created as separate and independent works and assembled into a collective whole for distribution and use.

registry Collection of registry objects compiled to support the business of a given community.

repository Collection of physical or digital objects compiled for information and documentation purposes and/or for storage and safekeeping.

dataset Collection of physical or digital objects generated by research activities.

4. 'Vocabularies for Registry Schema', Australian National Data Service ⟨http://services.ands.org.au/documentation/rifcs/1.5/vocabs/vocabularies.html⟩, accessed 22 Nov. 2013

2.2.4 party[type]

group One or more persons acting as a family, group, association, partnership or corporation.

person Human being or identity assumed by one or more human beings.

administrativePosition A kind of party where the position, name and contact information are present but the identity of the party filling the role is not specified.

2.2.5 service[type]

create Instrument.

generate Simulator.

report Visualisation, summary.

annotate *Collection of comments, reviews, ratings.*

transform Analysis, conversion.

assemble Aggregation.

harvest-oaipmh OAI-PMH Harvest.

search-http Search service over HTTP.

search-opensearch OpenSearch search.

search-sru SRU search.

search-srw SRW search.

search-z3950 Z39.50 search.

syndicate-atom ATOM syndication.

syndicate-rss RSS feed.

2.2.6 activity | collection | party | service > identifier[type]

abn Australian Business Number.

arc Australian Research Council identifier.

ark ARK Persistent Identifier Scheme.

AU-ANL:PEAU National Library of Australia identifier.

doi Digital Object Identifier.

handle Handle System Identifier.

infouri 'Info' URI scheme.

isil International Standard Identifier for Libraries.

local Identifer unique within a local context.

nhmrc National Health and Medical Research Council identifier.

orcid Open Researcher and Contributor Identifier.

purl Persistent Uniform Resource Locator.

uri Uniform Resource Identifier.

2.2.7 activity | collection | party | service > name[type]

primary Official name of the registry object.

abbreviated Shortened form of, or acronym for, the official name.

alternative Any other form of name used now or in the past as a substitute or alternative for the official name.

2.2.8 party > name > namePart[type], collection > citationInfo > citation-Metadata > contributor > namePart[type]

family Last name or surname.

given Forename or given or Christian name.

suffix Honours, awards, qualifications and other identifiers conferred.

title Word or phrase indicative of rank, office, nobility, honour, etc., or a term of address associated with a person.

superior Part of a name that describes a party (group) that contains one or more integral subordinate parties (sub-groups or sub-units).

subordinate Part of a name that describes a party (group) that is an integral sub-group or sub-unit of a superior party (group).

2.2.9 collection > dates[type]

dc.available Date (often a range) that the resource became or will become available.

dc.created Date of creation of the resource.

dc.dateAccepted Date of acceptance of the resource.

dc.dateSubmitted Date of submission of the resource.

dc.issued Date of formal issuance (e.g. publication) of the resource.

dc.valid Date (often a range) of validity of a resource.

2.2.10 collection > dates > date[type], activity | collection | party | service > coverage > temporal > date[type]

dateFrom Start date for *the applicability of the date, or* a temporal coverage period.

dateTo End date for *the applicability of the date, or* a temporal coverage period.

2.2.11 collection > dates > date[dateFormat], activity | collection | party | service > coverage > temporal > date[dateFormat]

W3CDTF W3C Date Time Format.

2.2.12 activity | collection | party | service > location > address > electronic[type]

Service

wsdl Web Service Definition Language.

All

email String used to receive messages by means of a computer network.

other Other electronic address.

url Uniform Resource Locator.

2.2.13 service > location > address > electronic > arg[type]

string Indicates the value of an argument is a plain text string

object indicates the value of an argument is an object, most likely in serialized form

2.2.14 service > location > address > electronic > arg[use]

inline Indicates the argument forms part of the base URL.

keyValue Indicates the argument is passed using key = value pairings in the query component of a URL.

2.2.15 activity | collection | party | service > location > address > physical[type]

streetAddress Address where an entity is physically located.

postalAddress Address where mail for an entity should be sent.

2.2.16 activity | collection | party | service > location > address > physical > addressPart[type]

addressLine An address part that is a separate line of a structured address.

text A single address part that contains the whole address in unstructured form.

telephoneNumber An address part that contains a telephone number, including a mobile telephone number.

faxNumber An address part that contains a fax (facsimile) number.

2.2.17 activity | collection | party | service > location | coverage > spatial[type]

gmlKmlPolyCoords A set of KML long/lat co-ordinates derived from GML (OpenGIS Geography Markup Language) defining a polygon as described by the KML coordinates element but without the altitude component.

gpx GPS Exchange Format.

iso31661 ISO 3166-1 Codes for the representation of names of countries and their subdivisions - Part 1: Country codes.

iso31662 Codes for the representation of names of countries and their subdivisions - Part 2: Country subdivision codes.

iso31663 ISO 3166-3 Codes for country names which have been deleted from ISO 3166-1 since its first publication in 1974.

iso19139dcmiBox DCMI Box notation derived from bounding box metadata conformant with the ISO 19139 schema.

kmlPolyCoords A set of KML (Keyhole Markup Language) long/lat co-ordinates defining a polygon as described by the KML coordinates element.

dcmiPoint spatial location information specified in DCMI Point notation.

text Free-text representation of spatial location.

2.2.18 activity > relatedObject | relatedInfo > relation[type]

To activity

hasPart Contains the related activity.

isPartOf Is contained in the related activity.

To collection

hasOutput Delivers materials in the related collection.

To party

hasParticipant Is undertaken by the related party.

hasPrincipalInvestigator Is researched by the related party.

isFundedBy Receives monetary or in-kind aid from the related program.

isManagedBy Is organised and/or delivered by the related party.

isOwnedBy Legally belongs to the related party.

To any

hasAssociationWith Has an unspecified relationship with the related *object*.

2.2.19 collection > relatedObject | relatedInfo > relation[type]

To activity

isOutputOf Is a product of the related activity.

To collection

describes Is a catalogue for, or index of, of items in the related collection.

hasPart Contains the related collection.

isDescribedBy Is catalogued or indexed by the related *catalogue or index*.

isLocatedIn Is held in the related repository.

isLocationFor Is the repository where the related collection is held.

isPartOf Is contained within the related collection.

isDerivedFrom Collection is derived from the related collection, e.g. through analysis.

hasDerivedCollection The related collection is derived from the collection, e.g. through analysis.

To party

hasCollector Has been aggregated by the related party.

hasPrincipalInvestigator Is researched by the related party.

isManagedBy Is maintained and made accessible by the related party.

isOwnedBy Legally belongs to the related party.

isEnrichedBy Additional value provided to a collection by a party.

To publication

isCitedBy Indicates that B includes A in a citation.

isReferencedBy Indicates A is used as a source of information by B.

isDocumentedBy Indicates B is documentation about/explaining A.

isSupplementedBy Indicates that A is a supplement to B.

isSupplementTo Indicates that B is a supplement to A.

isReviewedBy The cited entity presents statements, ideas or conclusions that are reviewed by the citing entity.

isSupportedBy The cited entity receives intellectual or factual support from the citing entity.

To service

supports Can be contributed to, accessed or used through the related service.

isAvailableThrough Specialisation of 'supports' type, for Harvest, Search and Syndicate.

isProducedBy Specialisation of 'supports' type, for Create, Generate and Assemble.

isPresentedBy Specialisation of 'supports' type, for Report.

hasValueAddedBy Specialisation of 'supports' type, for Annotate.

isOperatedOnBy Specialisation of 'supports' type, for Transform.

To any

hasAssociationWith Has an unspecified relationship with the related *object*.

2.2.20 party > relatedObject | relatedInfo > relation[type]

To activity

isFundedBy Receives monetary or in-kind aid from the related program.

isFunderOf Provides monetary or in-kind aid to the related activity.

isParticipantIn Is enrolled in the related activity.

isPrincipalInvestigatorOf Is a *researcher involved* in the related activity.

To collection

enriches Provides additional value to the related collection.

isCollectorOf Has aggregated the related collection.

isManagerOf Administers the related collection.

isPrincipalInvestigatorOf Is a *researcher involved* in the related collection.

To party

hasMember Is a group that has enrolled the related party.

hasPart Is a group that contains the related group.

isFundedBy Receives monetary or in-kind aid from the related party.

isFunderOf Provides monetary or in-kind aid to the related party.

isManagedBy Is overseen by the related party.

isManagerOf Oversees the related party.

isMemberOf Is enrolled in the related group.

isOwnedBy Legally belongs to the related party.

isPartOf Is a group that is contained in the related group.

To any

hasAssociationWith Has an unspecified relationship with the related *object*.

isOwnerOf Legally possesses the related activity, collection, service or group.

2.2.21 service > relatedObject | relatedInfo > relation[type]

To collection

isSupportedBy Enables contribution and access to and use of the related collection.

makesAvailable Specialisation of 'isSupportedBy' type, for Harvest, Search and Syndicate.

produces Specialisation of 'isSupportedBy' type, for Create, Generate and Assemble.

presents Specialisation of 'isSupportedBy' type, for Report.

operatesOn Specialisation of 'isSupportedBy' type, for Transform.

addsValueTo Specialisation of 'isSupportedBy' type, for Annotate.

To party

isManagedBy Is overseen by the related party.

isOwnedBy Legally belongs to the related party.

To service

hasPart Contains the related service.

isPartOf Is contained in the related service.

To any

hasAssociationWith Has an unspecified relationship with the related *object*.

2.2.22 activity | collection | party | service > subject[type]

The following are additional controlled terms used by the UK registry but not recognized in the RIF-CS standard.

gemet General Environmental Multi-Lingual Thesaurus.[5]

hasset Humanities and Social Science Electronic Thesaurus.[6]

ipsv2 Integrated Public Sector Vocabulary, version 2.0.[7]

iso19115topic ISO 19115 Topic Category Code; see for example *UK GEMINI*[8] for a list.

jacs3 Joint Academic Coding System (JACS), version 3.0.[9]

rcukrc Research Councils UK Research Classifications.[10]

ukdasc UK Data Archive Subject Categories.

2.2.23 activity | collection | party | service > description[type]

Collection

significanceStatement A statement describing the significance of a collection within its domain or context.

lineage Text describing the collection lineage

Party

researchAreas Text describing a contributor organisation's distinctive research portfolio and research strengths.

researchDataProfile Text describing and highlighting the research data (and related parties, projects and services) whose description the organisation has contributed to Research Data Australia.

researchSupport Text describing specific data-related support services offered by the contributor organisation such as archives, repositories, data centres, metadata stores, high performance computing facilities, data-intensive instruments, e-research support centres, data management support services, etc.

5. 'GEMET Thesaurus', European Environment Information and Observation Network ⟨http://www.eionet.europa.eu/gemet/⟩, accessed 31 Jan. 2014.
6. 'Our HASSET Thesaurus', UK Data Archive ⟨http://data-archive.ac.uk/find/hasset-thesaurus⟩, accessed 31 Jan. 2014.
7. UK Government, e-Government Unit, IPSV Board, *Integrated Public Sector Vocabulary*, version 2.0 (London: Porism, 2006) ⟨http://id.esd.org.uk/IPSV/2.00⟩, accessed 31 Jan. 2014.
8. Association for Geographic Information, *UK GEMINI: Specification for discovery metadata for geospatial data resources*, version 2.2 (Dec. 2012) ⟨http://www.agi.org.uk/storage/standards/uk-gemini/GEMINI2.2.pdf⟩, accessed 20 Jan. 2014, 18.
9. 'Joint Academic Coding System (JACS) Version 3.0', Higher Education Standards Authority ⟨http://www.hesa.ac.uk/content/view/1776/649/⟩, accessed 31 Jan. 2014.
10. 'Summary of Peer Review harmonisation activities', Research Councils UK ⟨http://www.rcuk.ac.uk/research/efficiency/researchadmin/harmonisation/⟩, accessed 31 Jan. 2014.

Service

deliveryMethod Information about how the service is delivered. Should be one of: web-service, software, offline, workflow.

All

brief Short account for selection purposes.

full Full account.

logo Symbol used as an identifying mark.

note A brief informational message, not object metadata, to notify the record consumer of some important aspect regarding the object or its metadata.

2.2.24 activity | collection | party | service > relatedInfo[type]

activity An undertaking or process related to the creation, update, or maintenance of a collection.

collection An aggregation of physical and/or digital resources which has meaning in a research context.

dataQualityInformation Data quality statements or summaries of data quality issues affecting the data.

metadata An alternative metadata format for the Object. This is most likely to be a discipline or system-specific format, e.g. NetCDF or ANZLIC.

party A person, group or role related to the creation, update, or maintenance of a collection, to an activity, or to the provision of a service.

publication Any formally published document, whether available in digital or online form or not.

reuseInformation Information that supports reuse of data, such as data definitions, instrument calibration or settings, units of measurement, sample descriptions, experimental parameters, methodology, data analysis techniques, or data derivation rules.

service A system (analogue or digital) that provides one or more functions of value to an end user.

website Any publicly accessible web location containing information related to the collection, activity, party or service.

2.2.25 activity | collection | party | service > relatedInfo > identifier[type], activity | collection | party | service > relatedInfo > format > identifier[type]

abn Australian Business Number.

arc Australian Research Council identifier.

ark ARK Persistent Identifier Scheme.

AU-ANL:PEAU National Library of Australia identifier.

doi Digital Object Identifier.

ean13 International Article Number.

eissn Electronic International Standard Serial Number.

handle Handle System Identifier.

infouri 'Info' URI scheme.

isbn International Standard Book Number.

isil International Standard Identifier for Libraries.

issn International Standard Serial Number.

istc International Standard Text Code.

lissn *Linking International Standard Serial Number.*

local Identifer unique within a local context.

mediaType The Media Type (MIME type) of the information. Values should be taken from IANA Media Type assignment listing. You may choose to use application/x-name if it is well known within the relevant discipline.

nhmrc National Health and Medical Research Council identifier.

orcid Open Researcher and Contributor Identifier.

purl Persistent Uniform Resource Locator.

researcherID Thomson Reuters ResearcherID.

upc Universal Product Code.

uri Uniform Resource Identifier.

urn Uniform Resource Name.

2.2.26 activity | collection | party | service > rights > licence[type]

CC-BY Creative Commons Attribution.

CC-BY-SA Creative Commons Attribution-ShareAlike.

CC-BY-ND Creative Commons Attribution-NoDerivs.

CC-BY-NC Creative Commons Attribution-NonCommercial.

CC-BY-NC-SA Creative Commons Attribution-NonCommercial-ShareAlike.

CC-BY-NC-ND Creative Commons Attribution-NonCommercial-NoDerivs.

GPL GNU General Public Licence.

AusGoalRestrictive AusGoal Restrictive Licence.

NoLicence No licence.

Unknown/Other Unknown licence or *custom licence defined by the rights holder*.

2.2.27 collection > citationInfo > fullCitation[style]

Harvard *Generic author-date style.*

APA American Psychological Association.

MLA Modern Language Association of America.

Vancouver *US National Library of Medicine/International Committee of Medical Journal Editors.*

IEEE Institute of Electrical and Electronic Engineers.

CSE Council of Science Editors.

Chicago Chicago Manual of Style.

AMA American Medical Association.

AGPS-AGIMO Australian Style Manual.

AGLC Australian Guide to Legal Citation.

ACS American Chemical Society.

DataCite *DataCite Consortium.*

2.2.28 collection > citationInfo > citationMetadata > identifier[type]

ark ARK Persistent Identifier Scheme.

doi Digital Object Identifier.

ean13 International Article Number.

eissn Electronic International Standard Serial Number.

handle Handle System Identifier.

infouri 'Info' URI scheme

isbn International Standard Book Number.

issn International Standard Serial Number.

istc International Standard Text Code.

lissn *Linked International Standard Serial Number.*

local Identifer unique within a local context.

purl Persistent Uniform Resource Locator.

upc Universal Product Code.

uri Uniform Resource Identifier.

urn Uniform Resource Name.

2.2.29 collection > citationInfo > citationMetadata > date[type]

publicationDate The date when the data was or will be made publicly available.

available The date the resource is made publicly available. May be a range, or indicate the end of an embargo period

created The date the resource itself was put together; this could be a date range or a single date for a final component, e.g. the finalised file with all of the data

date Any relevant date not otherwise specified

dateAccepted The date that the publisher accepted the resource into their system

dateSubmitted The date the creator submits the resource to the publisher. This could be different from dateAccepted if the publisher then applies a selection process

endPublicationDate Use when publicationDate is a range.

issued The date that the resource is published or distributed, e.g. to a data centre.

modified The date of the last update to the resource, when the resource is being added to. (Equivalent to DataCite 'Updated'.)

startPublicationDate Use when publicationDate is a range.

valid The date or date range during which the dataset or resources are accurate.

3

Internally managed RIF-CS elements

Several of the elements in a RIF-CS record are managed internally by the Registry and are not populated by harvested metadata. These are listed below.

registryObject[GROUP] This is set to the name of the organization from which the metadata has been harvested.

registryObject > key This is the Registry's internal unique identifier for the object. It may be based on an existing identifier for the object or newly generated.

registryObject > originatingSource This is set to an identifier or URI for the repository that originally supplied the metadata. This is assumed to be the organization from which the metadata has been harvested unless more accurate information can be inferred.

registryObject > collection[TYPE] This is 'dataset' for harvested dataset records.

registryObject > collection[DATEMODIFIED] This is the date on which the Registry record was last modified (or created).

registryObject > collection > existenceDates > startDate This is the date on which the record was first harvested.

registryObject > collection > existenceDates > endDate This would be the date on which the record was withdrawn.

4

Mapping from DDI to RIF-CS

The following table provides a mapping to populate a RIF-CS Collection record from a UKDA DDI record. The value of the UKDA ID is recorded in the DDI record at codeBook > stdyDscr > citation > titlStmt > IDNo[AGENCY=UKDA].

RIF-CS 1.5 element	Source (using DDI 2.5 record)
collection[dateAccessioned]	codeBook > stdyDscr > citation > distStmt > depDate[DATE]
identifier[TYPE=doi]	codeBook > stdyDscr > citation > titlStmt > IDNo[AGENCY=datacite]
identifier[TYPE=local]	codeBook > stdyDscr > citation > titlStmt > IDNo[agency] + ':' + IDNo
name[TYPE=primary] > namePart	codeBook > stdyDscr > citation > titlStmt > titl
name[TYPE=alternative] > namePart	codeBook > stdyDscr > citation > titlStmt > altTitl
dates[type=dc.available, dc.issued] > date[TYPE=dateFrom]	codeBook > stdyDscr > citation > distStmt > distDate[DATE]
dates[TYPE=dc.dateSubmitted] > date[TYPE=dateFrom]	codeBook > stdyDscr > citation > distStmt > depDate[DATE]
location > address > electronic[TYPE=url] > value	codeBook > stdyDscr > citation > holdings[URI]
subject[TYPE=hasset]	codeBook > stdyDscr > stdyInfo > subject > keyword[VOCAB=S]
subject[TERMIDENTIFIER]	codeBook > stdyDscr > stdyInfo > subject > keyword[VOCAB=S; VOCABURI]
subject[TYPE=ukdasc]	codeBook > stdyDscr > stdyInfo > subject > topcClas
description[TYPE=full]	codeBook > stdyDscr > stdyInfo > abstract
coverage > temporal > date[TYPE=dateFrom]	codeBook > stdyDscr > stdyInfo > sumDscr > collDate[EVENT=start, single; DATE], timePrd[EVENT=start, single; DATE]
coverage > temporal > date[TYPE=dateTo]	codeBook > stdyDscr > stdyInfo > sumDscr > collDate[EVENT=end; DATE], timePrd[event=end; date]
coverage > spatial[TYPE=text]	codeBook > stdyDscr > stdyInfo > sumDscr > geogCover, geogUnit, nation; codeBook > stdyDscr > stdyInfo > subject > keyword[VOCAB=G]
relatedInfo[TYPE=metadata] > identifier[TYPE=uri]	'http://esds.ac.uk/DDI25/' + UKDA ID + '.xml'

RIF-CS 1.5 element	Source (using DDI 2.5 record)
relatedInfo[TYPE=metadata] > format > identifier[TYPE=uri]	'http://www.ddialliance.org/Specification/DDI-Codebook/2.5/XMLSchema/codebook.xsd'
relatedInfo[TYPE=publication] > identifier	Inferred from text processing codeBook > stdyDscr > othrStdyMat > relPubl if possible
relatedInfo[TYPE=publication] > relation[TYPE]	'isReferencedBy'
relatedInfo[TYPE=publication] > title	Inferred from text processing codeBook > stdyDscr > othrStdyMat > relPubl if possible
rights > rightsStatement	codeBook > stdyDscr > citation > prodStmt > copyright
rights > accessRights	codeBook > stdyDscr > dataAccs > useStmt > restrctn, conditions
citationInfo > citationMetadata > identifier[TYPE=doi]	codeBook > stdyDscr > citation > titlStmt > IDNo[AGENCY=datacite]
citationInfo > citationMetadata > contributor > namePart	codeBook > stdyDscr > citation > rspStmt > AuthEnty
citationInfo > citationMetadata > title	codeBook > stdyDscr > citation > titlStmt > titl
citationInfo > citationMetadata > version	codeBook > stdyDscr > citation > verStmt > version
citationInfo > citationMetadata > publisher	codeBook > stdyDscr > citation > distStmt > distrbtr
citationInfo > citationMetadata > date[type=publicationDate, available, issued]	codeBook > stdyDscr > citation > distStmt > distDate
citationInfo > citationMetadata > date[TYPE=dateSubmitted]	codeBook > stdyDscr > citation > distStmt > depDate
citationInfo > citationMetadata > url	codeBook > stdyDscr > citation > holdings[URI]

4.1 Related Objects

- A series is represented in the Registry as a Collection of type 'collection'. Each member of the series is listed in the series object as a Related Object with Relation type 'hasPart'. Within each member object, the series object is included as a Related Object with Relation type 'isPartOf'.
- The holding repository (i.e. UKDA) is represented in the Registry as a Collection of type 'repository'. Each dataset object includes the repository object as a Related Object with Relation type 'isLocatedIn'. The repository object lists each dataset object as a Related Object with Relation type 'isLocationFor'.
- Individuals and corporate entities contributing to the dataset are represented in the Registry as Parties. The dataset object lists each contributor as a Related Object with Relation type 'hasPrincipalInvestigator'. Each contributor object includes the dataset object as a Related Object with Relation type 'isPrincipalInvestigatorOf'.
- Funding information (recorded in the UKDA DDI records under codeBook > stdyDscr

> citation > prodStmt > fundAg) would most naturally be represented via an Activity object. The chain of relationships would be

Collection (dataset) *isOutputOf* Activity (study) *isFundedBy* Party (funder),

and conversely,

Party *isFunderOf* Activity *hasOutput* Collection.

It is, however, currently unclear how to populate the Activity record.

5

Mapping from UK GEMINI 2 to RIF-CS

The following table provides a mapping to populate a RIF-CS Collection record from one held by the NERC Data Catalogue Service. The mapping is based on UK GEMINI,[1] which is a profile of ISO 19115.

In considering whether this mapping might be suitable for other contexts, it may be helpful to consider MEDIN Discovery[2] to be a special case of NERC Discovery[3] – the profile originally used by the Data Catalogue Service – itself a special case of UK GEMINI, which is a special case of INSPIRE.[4] All of these are ISO 19115 profiles.

For brevity the element numbers and names from UK GEMINI are used, rather than the full XML paths from the ISO 19115 standard.

RIF-CS 1.5 element	Source (using NERC record)
originatingSource	35 – Metadata point of contact > Organisation name, Email address (translated to identifier or URI)
identifier	36 – Unique resource identifier > Code
identifier[TYPE]	36 – Unique resource identifier > Code space
name[TYPE=primary]	1 – Title
name[TYPE=alternative]	2 – Alternative title
dates[TYPE=dc.created] > date[5]	8 – Dataset reference date > Date type = creation, Date
dates[TYPE=dc.issued] > date[5]	8 – Dataset reference date > Date type = publication, Date
location > address > electronic[TYPE=url] > value	36 – Unique resource identifier if resolvable, otherwise 19 – Resource locator > Resource locator URL[6]

1. AGI, *UK GEMINI*.
2. Becky Seeley et al., *Guidance notes for the production of discovery metadata for the Marine Environmental Data and Information Network (MEDIN)*, version 2.3.7 (Marine Environmental Data and Information Network, 14 Mar. 2013) ⟨http://bit.ly/1gHYldk⟩, accessed 20 Jan. 2014.
3. Natural Environment Research Council, *NERC Discovery Metadata Standard*, version 1.0 (1 Apr. 2011) ⟨http://data-search.nerc.ac.uk/documents/metadatastandard_v1.0.pdf⟩, accessed 20 Jan. 2014.
4. European Commission, 'Commission Regulation (EC) No 1205/2008 of 3 December 2008 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards metadata', *Official Journal of the European Union*, 2008/L326 (4 Dec. 2008), 12–30 ⟨http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32008R1205:en:NOT⟩, accessed 20 Jan. 2014.
5. A single date value in the GEMINI record would be represented by date[TYPE=dateFrom] in the RIF-CS record. A date range in the GEMINI record, such as '2004-03-02/2005-06-02', would be represented by date[TYPE=dateFrom] ('2004-03-02') and date[TYPE=dateTo] ('2005-06-02') in the RIF-CS record.
6. Where multiple Resource locator URLs are provided, one will be selected using the value of Resource locator function. The order of preference is download, order, offlineAccess, information, search, no specified function.

RIF-CS 1.5 element	Source (using NERC record)
subject[TYPE=iso19115topic]	5 – Topic category (ISO 19115 code list)
subject	6 – Keyword > Keyword value
subject[TYPE]	6 – Keyword > Originating controlled vocabulary
description[TYPE=full]	4 – Abstract
description[TYPE=lineage]	10 – Lineage
coverage > temporal > date[5]	7 – Temporal extent
coverage > spatial[TYPE=iso19139dcmiBox]	44 – Bounding box
coverage > spatial	5 – Extent > Extent name
coverage > spatial[TYPE]	5 – Extent > Originating controlled vocabulary > Thesaurus name
relatedInfo[type=publication, website] > identifier, identifier[TYPE], title	Inferred from text processing 27 – Additional information source and 19 – Resource locator (with function 'information') if possible
relatedInfo > relation[TYPE]	For publications, 'isDocumentedBy'; for websites, 'hasAssocationWith'
relatedInfo[TYPE=website] > relation[TYPE=hasAssocationWith] > description	'Has additional information'
relatedInfo[TYPE=metadata] > identifier[TYPE=uri]	URL of NERC DCS record[7]
relatedInfo[TYPE=metadata] > format > identifier[TYPE=uri]	'http://www.agi.org.uk/storage/standards/uk-gemini/'
rights > accessRights	25 – Limitations on Public Access, 26 – Use constraints
citationInfo > citationMetadata > identifier	36 – Unique resource identifier > Code
citationInfo > citationMetadata > identifier[TYPE]	36 – Unique resource identifier > Code space
citationInfo > citationMetadata > contributor > namePart	23 – Responsible organisation > Individual name[8]
citationInfo > citationMetadata > title	1 – Title
citationInfo > citationMetadata > publisher	23 – Responsible organisation > Organisation name[9]

7. This is an OGC CSW request with the following form: `http://⟨CSW host⟩/geonetwork/srv/en/csw?SERVICE=CSW&REQUEST=GetRecordById&ID=⟨file identifier⟩&elementSetName=full&OutputSchema=http://www.isotc211.org/2005/gmd`
8. Responsible party role is used to filter the responsible parties provided. Only one role is used, and they are checked in this order: author, originator, principalInvestigator, owner. If the author role is used, the order will be preserved using the RIF-CS SEQ attribute.
9. Responsible party role is used to filter the responsible parties provided. Only one role is used, and they are checked in this order: publisher, distributor, resource provider. Note that the distributor may be recorded in the ISO 19115 element distributionInfo > … > distributorContact instead of identificationInfo > … > pointOfContact.

RIF-CS 1.5 element	Source (using NERC record)
citationInfo > citationMetadata > date[TYPE=publicationDate, available, issued]	8 – Dataset reference date > Date type = publication, Date
citationInfo > citationMetadata > date[TYPE=created]	8 – Dataset reference date > Date type = creation, Date
citationInfo > citationMetadata > date[TYPE=modified]	8 – Dataset reference date > Date type = revision, Date
citationInfo > citationMetadata > url	36 – Unique resource identifier if resolvable, otherwise 19 – Resource locator > Resource locator URL[10]

5.1 Related Objects

- A series is represented in the Registry as a Collection of type 'collection'. Each member of the series is listed in the series object as a Related Object with Relation type 'hasPart'. Within each member object, the series object is included as a Related Object with Relation type 'isPartOf'.
- Individuals and corporate entities contributing to the dataset are represented in the Registry as Parties. The following table shows the mapping from UK GEMINI's controlled vocabulary of roles to those used by RIF-CS for types of relationship (note that 'hasPrincipalInvestigator' has a broader definition than the name suggests). Entities with other roles are not recorded in the Registry.

RIF-CS 1.5 values	UK GEMINI values
isManagedBy/isManagerOf	Custodian
isOwnedBy/isOwnerOf	Owner
hasPrincipalInvestigator/isPrincipalInvestigatorOf	Originator
hasPrincipalInvestigator/isPrincipalInvestigatorOf	Principal Investigator
isEnrichedBy/enriches	Processor
hasPrincipalInvestigator/isPrincipalInvestigatorOf	Author

10. Where multiple Resource locator URLs are provided, one will be selected using the value of Resource locator function. The order of preference is download, order, offlineAccess, information, search, no specified function.

6

Mapping from DataCite to RIF-CS

The following table provides a mapping to populate a RIF-CS Collection record from a DataCite record.[1]

RIF-CS 1.5 element	Source (using DataCite record)
collection[dateAccessioned]	Date[dateType=Accepted]
identifier[type=doi]	Identifier[identifierType=DOI]
identifier[2]	AlternateIdentifier
name[type=primary] > namePart	Title
name[type=alternative] > namePart	Title[titleType=AlternativeTitle]
dates[type=dc.available] > date[3]	Date[dateType=Available]
dates[type=dc.created] > date[3]	Date[dateType=Created]
dates[type=dc.dateAccepted] > date[3]	Date[dateType=Accepted]
dates[type=dc.dateSubmitted] > date[3]	Date[dateType=Submitted]
dates[type=dc.issued] > date[3]	Date[dateType=Issued]
dates[type=dc.valid] > date[3]	Date[dateType=Valid]
location > address > electronic[type=url] > value	'http://dx.doi.org/' + Identifier[identifierType=DOI]
subject[4]	Subject
description[type=full]	Description[descriptionType=Abstract]
description[type=lineage]	Description[descriptionType=Methods]
description[type=brief]	Description[descriptionType=Other]
coverage > spatial[type=iso19139dcmibox]	GeoLocation[geoLocationBox]
coverage > spatial[type=dcmiPoint]	GeoLocation[geoLocationPoint]

1. DataCite Metadata Working Group, *DataCite Metadata Schema for the Publication and Citation of Research Data*, version 3.0 (DataCite Consortium, 2013), DOI: 10.5438/0008.
2. DataCite does not use a controlled vocabulary for alternateIdentifierType, so some parsing would be needed to ensure that identifier types known to RIF-CS are recognised. The RIF-CS type 'local' is used as fallback for unrecognised types.
3. A single date value in the DataCite record would be represented by date[type=dateFrom] in the RIF-CS record. A date range in the DataCite record, such as '2004-03-02/2005-06-02', would be represented by date[type=dateFrom] ('2004-03-02') and date[type=dateTo] ('2005-06-02') in the RIF-CS record.
4. Some interpretation would be needed to translate values of DataCite's subjectScheme or schemeURI (e.g. 'http://id.loc.gov/authorities/subjects') into terms from the controlled vocabularies used by RIF-CS (e.g. 'lcsh'). The RIF-CS type 'local' is used as fallback for unrecognised schemes.

RIF-CS 1.5 element	Source (using DataCite record)
coverage > spatial[TYPE=text]	GeoLocation[GEOLOCATIONPLACE]
relatedInfo[TYPE] (see Section 6.2)	RelatedIdentifier[RELATIONTYPE]
relatedInfo > identifier	RelatedIdentifier
relatedInfo > identifier[TYPE] (see Section 6.2)	RelatedIdentifier[RELATEDIDENTIFIERTYPE]
relatedInfo > relation[TYPE] (see Section 6.2)	RelatedIdentifier[RELATIONTYPE]
relatedInfo > format > title	RelatedIdentifier[RELATEDMETADATASCHEME]
relatedInfo > format > identifier[TYPE=uri]	RelatedIdentifier[SCHEMEURI]
rights > rightsStatement	Rights
rights > rightsStatement[RIGHTSURI]	Rights[RIGHTSURI]
citationInfo > citationMetadata > identifier	Identifier[IDENTIFIERTYPE=DOI]
citationInfo > citationMetadata > contributor > namePart	Creator > creatorName
citationInfo > citationMetadata > title	Title
citationInfo > citationMetadata > version	Version
citationInfo > citationMetadata > publisher	Publisher
citationInfo > citationMetadata > date[TYPE=publicationDate]	PublicationYear
citationInfo > citationMetadata > date[TYPE=available]	Date[DATETYPE=Available]
citationInfo > citationMetadata > date[TYPE=created]	Date[DATETYPE=Created]
citationInfo > citationMetadata > date[TYPE=dateAccepted]	Date[DATETYPE=Accepted]
citationInfo > citationMetadata > date[TYPE=dateSubmitted]	Date[DATETYPE=Submitted]
citationInfo > citationMetadata > date[TYPE=issued]	Date[DATETYPE=Issued]
citationInfo > citationMetadata > date[TYPE=modified]	Date[DATETYPE=Updated]
citationInfo > citationMetadata > date[TYPE=valid]	Date[DATETYPE=Valid]
citationInfo > citationMetadata > url	'http://dx.doi.org/' + Identifier[IDENTIFIERTYPE=DOI]

6.1 Related Objects

- The holding repository (i.e. DataCite Publisher) is represented in the Registry as a Collection of type 'repository'. Each dataset object includes the repository object as a Related Object with Relation type 'isLocatedIn'. The repository object lists each dataset object as a Related Object with Relation type 'isLocationFor'.
- Individuals and corporate entities contributing to the dataset are represented in the

Registry as Parties. The dataset object lists each DataCite Creator and Contributor (with contributorTypes 'DataCollector', 'ProjectLeader', or 'WorkPackageLeader') as a Related Object with Relation type 'hasPrincipalInvestigator'. Each contributor object includes the dataset object as a Related Object with Relation type 'isPrincipalInvestigatorOf'. (Note that RIF-CS defines these relations as involving a researcher in the collection: it does not specify that the researcher has to be an official PI.)

- Funding information (recorded in DataCite records using Contributor elements with contributorType 'Funder') would most naturally be represented via an Activity object. The chain of relationships would be

  Collection (dataset) *isOutputOf* Activity (study) *isFundedBy* Party (funder),

  and conversely,

  Party *isFunderOf* Activity *hasOutput* Collection.

  It is, however, currently unclear how to populate the Activity record.

6.2 Related Information

The following table shows the mapping from DataCite's controlled vocabulary to that used by RIF-CS for related object identifiers.

RIF-CS 1.5	DataCite
ark	ARK
doi	DOI
ean13	EAN13
eissn	EISSN
handle	Handle

RIF-CS 1.5	DataCite
isbn	ISBN
issn	ISSN
istc	ISTC
lissn	LISSN
urn	LSID

RIF-CS 1.5	DataCite
local	PMID
purl	PURL
upc	UPC
uri	URL
urn	URN

The following table shows the mapping from DataCite's controlled vocabulary for types of relationship, to those used by RIF-CS for types of related object and types of relationship.

RIF-CS 1.5 values	DataCite values
publication / isCitedBy	IsCitedBy
publication / isSupplemetedBy	IsSupplementedBy
publication / isSupplementTo	IsSupplementTo
collection / isPartOf	IsPartOf
collection / hasPart	HasPart
publication / isReferencedBy	IsReferencedBy
publication / isDocumentedBy	IsDocumentedBy
collection / isDerivedFrom	isCompiledBy
collection / hasDerivedCollection	Compiles

All other DataCite relationships should be represented in RIF-CS as relations of type 'hasAssociationWith', with the text of the DataCite term put into normal case (e.g. 'Has metadata') and added to the RIF-CS relatedInfo > relation > description element.

The DataCite relationships 'isContinuedBy', 'Continues', 'isMetadataFor' 'isNewVer-

sionOf', 'isPreviousVersionOf', 'Documents', 'isVariantFormOf', 'isOriginalFormOf' and 'isIdenticalTo' imply a relatedInfo type of 'collection'. The relationships 'Cites' and 'References' imply a relatedInfo type of 'publication'.

Note that the DataCite relationship 'HasMetadata' could be with metadata (i.e. an alternative, full metadata record for the dataset in question), reuseInformation (e.g. data definitions, experimental parameters) or dataQualityInformation. The type of related object should therefore be left ambiguous by the automated mapping process.

7

Mapping from EPrints to RIF-CS

The following table provides a mapping to populate a RIF-CS Collection record from an EPrints record. This mapping takes account of variations to the standard EPrints metadata profile made by the ReCollect plugin (developed by the University of Essex) and the University of Glasgow.[1]

RIF-CS 1.5 element	Source (using EPrints record)
collection[DATEACCESSIONED]	datestamp
identifier[TYPE=local]	id_number
name[TYPE=primary] > namePart	title
name[TYPE=alternative] > namePart	alt_title
dates[TYPE=dc.available] > date[TYPE=dateFrom]	date_embargo
dates[TYPE=dc.created] > date	collection_date (ReCollect)
dates[TYPE=dc.dateSubmitted] > date	datestamp
dates[TYPE=dc.issued] > date[TYPE=dateFrom]	revision
location > address > electronic[TYPE=url] > value	Derived from id_number
subject[TYPE=local]	keywords
subject	subjects
description[TYPE=full]	abstract
description[TYPE=lineage]	provenance (ReCollect)
description[TYPE=note]	note
coverage > temporal > date	temporal_cover (ReCollect)
coverage > spatial[TYPE=text]	geographic_cover (ReCollect)
coverage > spatial[TYPE=iso19139dcmiBox]	bounding_box (ReCollect)
relatedInfo[TYPE=collection] > identifier	related_resources (ReCollect)
relatedInfo[TYPE=collection] > relation[TYPE]	related_resources > relationType (ReCollect)
rights > accessRights	security, restrictions (Glasgow), accessLimitations (Glasgow)
rights > licence	license
citationInfo > citationMetadata > identifier	id_number

1. The metadata fields used by UK EPrints users to describe datasets is being tracked at `http://tinyurl.com/or34774`.

RIF-CS 1.5 element	Source (using EPrints record)
citationInfo > citationMetadata > contributor > namePart	creators
citationInfo > citationMetadata > title	title
citationInfo > citationMetadata > version	revision
citationInfo > citationMetadata > publisher	publisher
citationInfo > citationMetadata > date[type=publicationDate, issued]	revision
citationInfo > citationMetadata > date[TYPE=available]	date_embargo
citationInfo > citationMetadata > url	Derived from id_number

7.1 Related Objects

- The holding repository is represented in the Registry as a Collection of type 'repository'. Each dataset object includes the repository object as a Related Object with Relation type 'isLocatedIn'. The repository object lists each dataset object as a Related Object with Relation type 'isLocationFor'.
- Individuals and corporate entities contributing to the dataset are represented in the Registry as Parties. The dataset object lists each creator as a Related Object with Relation type 'hasPrincipalInvestigator'. Each creator object includes the dataset object as a Related Object with Relation type 'isPrincipalInvestigatorOf'.
- The project that produced the dataset (recorded in the ReCollect grant element or EPrints projects element) would be represented by a minimal Activity object:

Sample Record 1

registryObject
└ GROUP = …
· key = …
· originatingSource = …
· activity
└ TYPE = project
└ DATEMODIFIED = …
· · identifier = ⟨*grant number*⟩
· · relatedObject
· · · key = ⟨*Collection key*⟩
· · · relation
└ TYPE = hasOutput
· · · key = ⟨*Party key*⟩
· · · relation
└ TYPE = isFundedBy

The dataset object would have a reciprocal relation of type 'isOutputOf'.

- Funding information (recorded in ReCollect records using the funders element) would be represented by a Party object, related to the dataset object via the above project (Activity) object:

Sample Record 2

registryObject
└ GROUP = …
· key = …
· originatingSource = …
· party
└ TYPE = group
└ DATEMODIFIED = …
· · name
· · · namePart = ⟨*funder name*⟩
· · relatedObject
· · · key = ⟨*Activity key*⟩
· · · relation
└ TYPE = isFunderOf

The registry should already be populated with major funding bodies, so in most instances this will involve adding information to the existing record rather than creating a new one.

8

Mapping from MODS to RIF-CS

The following table provides a mapping to populate a RIF-CS Collection record from a MODS (Metadata Object Description Schema) record, version 3.5.[1]

RIF-CS 1.5 element	Source (using MODS record)
identifier	identifier
identifier[TYPE]	identifier[TYPE]
name[TYPE=primary] > namePart	titleInfo > title + ': ' + subTitle
dates[TYPE=dc.created] > date[TYPE=dateFrom, dateTo][2]	originInfo > dateCreated[POINT=start, end]
dates[TYPE=dc.issued] > date[TYPE=dateFrom, dateTo][2]	originInfo > dateIssued[POINT=start, end]
dates[TYPE=dc.valid] > date[TYPE=dateFrom, dateTo][2]	originInfo > dateValid[POINT=start, end]
location > address > electronic[TYPE=url] > value	location > url
subject	subject > topic
subject[TYPE][3]	subject > topic[AUTHORITY, AUTHORITYURI]
subject[TERMIDENTIFIER]	subject > topic[VALUEURI]
description[TYPE=full]	abstract[TYPE=content], abstract (no type)
description[TYPE=significanceStatement]	abstract[TYPE=review]
description[TYPE=lineage]	note[TYPE=conservation history]
coverage > temporal > date[TYPE=dateFrom, dateTo]	subject > temporal[POINT=start, end]
coverage > spatial[TYPE=iso31661, iso31662, iso31663]	subject > geographicCode[AUTHORITY=iso3166]
coverage > spatial[TYPE=dcmiPoint, kmlPolyCoords]	subject > cartographics > coordinates
coverage > spatial[TYPE=text]	subject > geographic

1. 'Outline of Elements and Attributes in MODS Version 3.5', Library of Congress ⟨http://www.loc.gov/standards/mods/mods-outline-3-5.html⟩, accessed 21 Mar. 2014.
2. If the MODS date does not specify a POINT, the RIF-CS 'dateFrom' is used.
3. RIF-CS TYPE uses the same controlled vocabulary as MODS AUTHORITY. Some interpretation would be needed, however, to translate values of MODS' AUTHORITYURI (e.g. 'http://id.loc.gov/authorities/subjects') into the controlled terms (e.g. 'lcsh'). The RIF-CS type 'local' is used as fallback for unrecognised schemes.

RIF-CS 1.5 element	Source (using MODS record)
relatedInfo > identifier	relatedItem > identifier
relatedInfo > identifier[TYPE]	relatedItem > identifier[TYPE]
relatedInfo > relation[TYPE]	relatedItem[TYPE]
relatedInfo > format > identifier[TYPE=mediaType]	relatedItem > physicalDescription > internetMediaType
relatedInfo > title	relatedItem > titleInfo > title + ': ' + subTitle
rights > rightsStatement	accessCondition[TYPE=use and reproduction]
rights > accessRights	accessCondition[TYPE=restriction on access]
citationInfo > citationMetadata > identifier	identifier
citationInfo > citationMetadata > identifier[TYPE]	identifier[TYPE]
citationInfo > citationMetadata > contributor > namePart[TYPE=family, given]	name[TYPE=personal] > namePart[TYPE=family, given] where name > role > roleTerm = Author/author/aut
citationInfo > citationMetadata > contributor > namePart	name > namePart where name > role > roleTerm = Author/author/aut
citationInfo > citationMetadata > title	titleInfo > title + ': ' + subTitle
citationInfo > citationMetadata > version	originInfo > edition, note[type=version identification]
citationInfo > citationMetadata > publisher	originInfo > publisher, name > namePart where name > role > roleTerm = Publisher/publisher/pbl
citationInfo > citationMetadata > placePublished	originInfo > place > placeTerm[TYPE=text][4]
citationInfo > citationMetadata > date[TYPE=created]	originInfo > dateCreated
citationInfo > citationMetadata > date[TYPE=issued]	originInfo > dateIssued
citationInfo > citationMetadata > date[TYPE=modified]	originInfo > dateModified
citationInfo > citationMetadata > date[TYPE=publicationDate][5]	originInfo > dateIssued
citationInfo > citationMetadata > date[TYPE=valid]	originInfo > dateValid
citationInfo > citationMetadata > url	location > url[6]
citationInfo > citationMetadata > context	relatedItem[TYPE=series] > titleInfo > title + ': ' + subTitle

4. If originInfo[EVENTTYPE] is provided, only if the value is 'publication' should the element be used.
5. If there are a pair of MODS dateIssued elements with POINT attributes 'start' and 'end', this is represented in RIF-CS by dates of type 'startPublicationDate' and 'endPublicationDate'.
6. A url[USAGE=primary] takes priority.

8.1 Identifiers

The following table shows the mapping from the MODS suggested controlled vocabulary for identifier types to that used by RIF-CS.

RIF-CS 1.5	MODS
doi	doi
handle	hdl
isbn*	isbn
issn*	issn
istc*	istc
local	local
upc*	upc
uri	uri

* Not used in the context of identifying Registry Objects, but may be used for citations and relatedInformation.

Note that as the MODS vocabulary is a suggestion rather than an exhaustive list, other identifier schemes known to RIF-CS might be present in the MODS record.

8.2 Related Objects

- The holding repository is represented in the Registry as a Collection of type 'repository'. Each dataset object includes the repository object as a Related Object with Relation type 'isLocatedIn'. The repository object lists each dataset object as a Related Object with Relation type 'isLocationFor'.
- Individuals and corporate entities contributing to the dataset are represented in the Registry as Parties. The dataset object lists each MODS author as a Related Object with Relation type 'hasPrincipalInvestigator'. Each contributor object includes the dataset object as a Related Object with Relation type 'isPrincipalInvestigatorOf'. (Note that RIF-CS defines these relations as involving a researcher in the collection: it does not specify that the researcher has to be an official PI.)

8.3 Related Information

The following table shows the mapping from MODS' controlled vocabulary for types of related item, to those used by RIF-CS for types of related object and types of relationship.

RIF-CS 1.5 values	MODS values
collection / isDerivedFrom	original
collection / isPartOf	host
collection / hasPart	constituent
collection / hasDerivedCollection	otherVersion
publication / isReferencedBy	isReferencedBy

All other MODS relationships should be represented in RIF-CS as relations of type 'hasAssociationWith', with the text of the MODS term put into normal case (e.g. 'Other format') and added to the RIF-CS relatedInfo > relation > description element.

9

Mapping from OAI-PMH Dublin Core to RIF-CS

The following table provides a mapping to populate a RIF-CS Collection record from an OAI-PMH Dublin Core (`oai_dc`) record.[1]

RIF-CS 1.5 element	Source (using oai_dc record)
identifier[2]	dc:identifier
name[TYPE=primary] > namePart	dc:title
dates[TYPE=dc.issued] > date[TYPE=dateFrom]	dc:date
location > address > electronic[TYPE=url] > value	dc:identifier (if URL)
subject[TYPE=local]	dc:subject
description[TYPE=full]	dc:description
coverage > temporal > date[3]	dc:coverage (if parsed as date information)
coverage > spatial[TYPE=text]	dc:coverage (if not parsed as date information)
relatedInfo > identifier[2]	dc:relation
relatedInfo > relation[TYPE=hasAssociationWith] > description	'Unknown'
rights > rightsStatement	dc:rights
citationInfo > citationMetadata > identifier[2]	dc:identifier
citationInfo > citationMetadata > contributor > namePart	dc:creator, dc:contributor
citationInfo > citationMetadata > title	dc:title
citationInfo > citationMetadata > publisher	dc:publisher
citationInfo > citationMetadata > date[TYPE=publicationDate, available, issued]	dc:date
citationInfo > citationMetadata > url	dc:identifier (if URL)

1. *The Open Archives Initiative Protocol for Metadata Harvesting: Protocol Version 2.0*, ed. Carl Lagoze et al. (Open Archives Initiative, 7 Dec. 2008) ⟨`http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm`⟩, accessed 22 Jan. 2014.
2. The type of identifier would be inferred from pattern matching, falling back to 'local' if no match is found.
3. A single date value in the `oai_dc` record would be represented by date[TYPE=dateFrom] in the RIF-CS record. A date range in the `oai_dc` record, such as '2004-03-02/2005-06-02', would be represented by date[TYPE=dateFrom] ('2004-03-02') and date[TYPE=dateTo] ('2005-06-02') in the RIF-CS record.

9.1 Related Objects

- Individuals and corporate entities contributing to the dataset are represented in the Registry as Parties. The dataset object lists each creator as a Related Object with Relation type ‘hasPrincipalInvestigator’. Each creator object includes the dataset object as a Related Object with Relation type ‘isPrincipalInvestigatorOf’.
- If the value of dc:source matches an object already in the Registry, a Relation of type ‘isDerivedFrom’ is added to the current record (indicating the referenced object) and a Relation of type ‘hasDerivedCollection’ is added to the referenced record (indicating the current object).

10 Tips for contributors

10.1 High priority metadata from harvested records

While RIF-CS has little in the way of mandatory metadata elements, a certain level of information will be needed if a harvested record is going to make a useful contribution to the Registry. The two highest priorities are

1. ensuring enough information can be displayed about the dataset so that users of the Registry can determine if a dataset might be useful for them;
2. gathering enough information to allow the dataset to show up in search results and the browsing interfaces.

The following items of information are listed in approximate order of importance, with the most important first. Shown under each heading are the relevant RIF-CS element(s), an indication of how the information is used by the Registry, and whence the information is harvested.

10.1.1 Dataset name

RIF-CS	name > namePart[TYPE=primary], citationInfo > citationMetadata > title
Usage	Used as the title of the record, in lists of records, and in the sample citation.
DDI	codeBook > stdyDscr > citation > titlStmt > titl
NERC	1 – Title
DataCite	Title
EPrints	title
MODS	titleInfo > title, subTitle
oai_dc	dc:title

10.1.2 Full description

RIF-CS	description[TYPE=full]
Usage	Used prominently in the displayed record, and in free text searches.
DDI	codeBook > stdyDscr > stdyInfo > abstract
NERC	4 – Abstract
DataCite	Description[DESCRIPTIONTYPE=Abstract]
EPrints	abstract

MODS abstract (no type or TYPE = content)

oai_dc dc:description

10.1.3 Identifier

RIF-CS identifer,
citationInfo > citationMetadata > identifier

Usage Used prominently in the displayed record, in free text searches, and in the sample citation.

DDI '10.5255/UKDA-SN-' + UKDA ID + '-1'

NERC 36 – Unique resource identifier > Code

DataCite Identifier[IDENTIFIERTYPE=DOI]

EPrints id_number

MODS identifier

oai_dc dc:identifier

10.1.4 Subject

RIF-CS subject

Usage Used prominently in the displayed record, in free text searches, and as a browsing hierarchy.

DDI codeBook > stdyDscr > stdyInfo > subject > keyword[VOCAB=S],
codeBook > stdyDscr > stdyInfo > subject > topClas

NERC 5 – Topic category,
6 – Keyword > Keyword value

DataCite Subject

EPrints keywords,
subjects

MODS subject > topic

oai_dc dc:subject

10.1.5 URL

RIF-CS location > address > electronic[TYPE=url] > value,
citationInfo > citationMetadata > url

Usage Used in the displayed record to provide access, and in the sample citation.

DDI 'http://dx.doi.org/10.5255/UKDA-SN-' + UKDA ID + '-1'

NERC 36 – Unique resource identifier if resolvable, otherwise 19 – Resource locator > Resource locator URL

DataCite 'http://dx.doi.org/' + Identifier[IDENTIFIERTYPE=DOI]

EPrints derived from id_number

MODS location > url

oai_dc dc:identifier (if URL)

10.1.6 Date published, issued, made available

RIF-CS dates[TYPE=dc.issued] > date[TYPE=dateFrom],
citationInfo > citationMetadata > date[TYPE=publicationDate, available, issued]

Usage Used as a search refinement parameter, and in the sample citation.

DDI codeBook > stdyDscr > citation > distStmt > distDate

NERC 8 – Dataset reference date > date type = publication, date

DataCite PublicationYear,
Date[DATETYPE=Available, Issued]

EPrints *published, issued:* revision,
available: date_embargo

MODS originInfo > dateIssued

oai_dc dc:date

10.1.7 Creator

RIF-CS citationInfo > citationMetadata > contributor > namePart

Usage Used in the sample citation, and to connect to Party records.

DDI codeBook > stdyDscr > citation > rspStmt > AuthEnty

NERC 23 – Responsible organisation > Individual name

DataCite Creator[CREATORNAME]

EPrints creators

MODS name > namePart (where name > role > roleTerm is 'Author'/'author'/'aut')

oai_dc dc:creator,
dc:contributor

10.1.8 Rights information

RIF-CS rights > …

Usage Used in the displayed record.

DDI *rightsStatement:* codeBook > stdyDscr > citation > prodStmt > copyright,
accessRights: codeBook > stdyDscr > dataAccs > useStmt > restrctn, conditions

NERC *accessRights:* 25 – Limitations on Public Access
26 – Use Constraints

DataCite *rightsStatement:* Rights

EPrints *accessRights:* security, restrictions (Glasgow), accessLimitations (Glasgow),
licence: license

MODS accessCondition

oai_dc *rightsStatement:* dc:rights

10.1.9 Spatial Coverage

RIF-CS coverage > spatial

Usage Used in the displayed record, and as a browsing tool.

DDI codeBook > stdyDscr > stdyInfo > sumDscr > geogCover, geogUnit, nation,
codeBook > stdyDscr > stdyInfo > subject > keyword[VOCAB=G]

NERC 44 – Bounding box,
5 – Extent > Extent name

DataCite GeoLocation

EPrints bounding_box,
geographic_cover

MODS subject > geographicCode[AUTHORITY=iso3166], subject > cartographics > coordinates, subject > geographic

oai_dc dc:coverage (if not parsed as date information)

10.1.10 Publisher

RIF-CS citationInfo > citationMetadata > publisher

Usage Used in the sample citation.

DDI codeBook > stdyDscr > citation > distStmt > distrbtr

NERC 23 – Responsible organisation > Organisation name

DataCite Publisher

EPrints publisher

MODS name > namePart (where name > role > roleTerm is 'Publisher'/'publisher'/'pbl'),
originInfo > publisher

oai_dc dc:publisher

10.2 Registry quality levels

For Collection objects, the registry software defines three quality levels.

10.2.1 Level 1

The derived RIF-CS record should be valid, and it should have at least the following form:

Sample Record 3

registryObjects
· registryObject
└ GROUP
· · key
· · collection
└ TYPE

10.2.2 Level 2

- The Collection should have at least one name[TYPE=primary] element.

- The Collection should be related to at least one Party.
- The Collection should have at least one description[TYPE=full] or description[TYPE=brief] element.
- The Collection should have at least one rightsStatement, licence or accessRights element.
- The Collection should have at least one location > address element.

10.2.3 Level 3

- The Collection should have at least one identifier element.
- The Collection should be related to at least one Activity.
- The Collection should have at least one subject element.
- The Collection should have at least one coverage > spatial element.
- The Collection should have at least one coverage > temporal element.
- The Collection should have at least one citationInfo element.
- The Collection should have at least one dates element.

Bibliography

Association for Geographic Information, *UK GEMINI: Specification for discovery metadata for geospatial data resources*, version 2.2 (Dec. 2012) ⟨http://www.agi.org.uk/storage/standards/uk-gemini/GEMINI2.2.pdf⟩, accessed 20 Jan. 2014.

DataCite Metadata Working Group, *DataCite Metadata Schema for the Publication and Citation of Research Data*, version 3.0 (DataCite Consortium, 2013), DOI: 10.5438/0008.

European Commission, 'Commission Regulation (EC) No 1205/2008 of 3 December 2008 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards metadata', *Official Journal of the European Union*, 2008/L326 (4 Dec. 2008), 12–30 ⟨http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32008R1205:en:NOT⟩, accessed 20 Jan. 2014.

'GEMET Thesaurus', European Environment Information and Observation Network ⟨http://www.eionet.europa.eu/gemet/⟩, accessed 31 Jan. 2014.

ISO 19115, *Geographic Information – Metadata* (International Organization for Standardization, 2003).

ISO 2146, *Information and Documentation – Registry Services for Libraries and Related Organizations* (International Organization for Standardization, 2010).

'Joint Academic Coding System (JACS) Version 3.0', Higher Education Standards Authority ⟨http://www.hesa.ac.uk/content/view/1776/649/⟩, accessed 31 Jan. 2014.

The Open Archives Initiative Protocol for Metadata Harvesting: Protocol Version 2.0, ed. Carl Lagoze et al. (Open Archives Initiative, 7 Dec. 2008) ⟨http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm⟩, accessed 22 Jan. 2014.

Natural Environment Research Council, *NERC Discovery Metadata Standard*, version 1.0 (1 Apr. 2011) ⟨http://data-search.nerc.ac.uk/documents/metadatastandard_v1.0.pdf⟩, accessed 20 Jan. 2014.

'Our HASSET Thesaurus', UK Data Archive ⟨http://data-archive.ac.uk/find/hasset-thesaurus⟩, accessed 31 Jan. 2014.

'Outline of Elements and Attributes in MODS Version 3.5', Library of Congress ⟨http://www.loc.gov/standards/mods/mods-outline-3-5.html⟩, accessed 21 Mar. 2014.

'Registry Interchange Format – Collections and Services (RIF-CS) v1.5: Schema Guidelines', Australian National Data Service ⟨http://services.ands.org.au/documentation/rifcs/1.5/guidelines/rif-cs.html⟩, accessed 22 Nov. 2013.

Seeley, Becky, et al., *Guidance notes for the production of discovery metadata for the Marine Environmental Data and Information Network (MEDIN)*, version 2.3.7 (Marine Environmental Data and Information Network, 14 Mar. 2013) ⟨http://bit.ly/1gHYldk⟩, accessed 20 Jan. 2014.

‘Summary of Peer Review harmonisation activities’, Research Councils UK ⟨http://www.rcuk.ac.uk/research/efficiency/researchadmin/harmonisation/⟩, accessed 31 Jan. 2014.

UK GOVERNMENT, E-GOVERNMENT UNIT, IPSV BOARD, *Integrated Public Sector Vocabulary*, version 2.0 (London: Porism, 2006) ⟨http://id.esd.org.uk/IPSV/2.00⟩, accessed 31 Jan. 2014.

‘Vocabularies for Registry Schema’, Australian National Data Service ⟨http://services.ands.org.au/documentation/rifcs/1.5/vocabs/vocabularies.html⟩, accessed 22 Nov. 2013.