A Collection of Crosswalks from Fifteen Research Data Schemas to Schema.org



RDA Supporting Output

DOI: 10.15497/RDA00069

Authors: Mingfang Wu, Penelope Hagan, Baptiste Cecconi, Stephen M. Richard, Chantelle Verhay,

RDA Research Metadata Schemas WG

Published: 9 March 2022

Version: 1.0

Abstract: The RDA Research Metadata Schemas Working Group has collected and aligned crosswalks from 15 source research metadata schemas to Schema.org. The source schemas include discipline agnostic schemas Dublin Core, Data Catalogue Vocabulary(DCAT), Data Catalogue Vocabulary - Application Profile (DCAT-AP), Registry Interchange Format - Collections and Services (RIF-CS), DataCite Schema, Dataverse; and discipline schemas ISO19115-1, EOSC/EDMI, Data Tag Suite (DATS), Bioschemas, B2FIND, Data Documentation Initiative (DDI), European Clinical Research Infrastructure Network (ECRIN), Space Physics Archive Search and Extract (SPASE); as well as CodeMeta for software.

The collection can serve as a reference for data repositories when they develop their crosswalks, as well as an indication of semantic interoperability among the schemas. The visualisation tool (developed by the World Data System - International Technology Office) provides a user-friendly interface to inspect the crosswalks, by querying either an individual property or a schema name.

The dataset:

The collection of the crosswalks is available in the following formats:

- The xlsx file has all information about the revision history and the same crosswalks in two sheets with different classifications. The first classification (labelled "crosswalks") includes categories: properties as recommended by the Google dataset search guidelines (A), DCAT properties that can be mapped to Schema.org (B) and DCAT properties that are unable to be mapped to Schema.org. The second classification (labelled "NISO classification") organises the mapped Schema.org properties with the categories from the NISO metadata type, including Descriptive Metadata, Technical Metadata, Preservation Metadata, Rights Metadata and Structural Metadata.
- The "crosswalks" sheet is also available in the csv file and the pdf file. The CSV file is accessible via the group's git repository, here.

Impact: The collection of crosswalks provides an important insight into semantic interoperability of metadata schemas commonly used for research data. The collection can serve as a reference for data repositories when they develop their own crosswalks or for projects that develop common metadata frameworks.

Language: English

License: Attribution 4.0 International (CC BY 4.0)

RDA webpage: https://www.rd-alliance.org/group/research-metadata-schemas-

 $\underline{wg/outcomes/collection\text{-}crosswalks\text{-}fifteen\text{-}research\text{-}data\text{-}schemas}$

Citation and Download: Wu, M., Hagan, P., Cecconi, B., Richard, S. M., Verhey, C., & RDA Research Metadata Schemas WG. (2021). A Collection of Crosswalks from Fifteen Research Data Schemas to

Schema.org. Research Data Alliance. https://doi.org/10.15497/RDA00069

This is a collection of crosswalks from 15 metadata schemas to Schema.org

Revision history

2019-11-20

ARDC intern Penelope Hagan (CSU student) aligned the 13 crosswalks against mapped Schema.org properties, in the sheet labelled "crosswalks".

Rows in green color are covered by the Google dataset search guide - they are 20 in total. The rows in dark green are "Required", the other rows in light green are "Recommended".

2020-09-04

- * Mingfang Wu (ARDC) re-adjusted the schema.org columns (B-E), the version before this date is named: V0.2 before-alighment 20200903. The Keep proprties from the v0.2 are largely from the CodeMeta Properties.
- * After the adjustment, keep the 20 properties from google, added properties from DCAT & DCAT-AP (dataset, datacatalogy and distribution) the properties that can be mapped to schema.org are in orange color, the properties can't find proper mapping to shema.org are in light organse color.
- * Baptiste Cecconi (Observatoire de Paris PSL) added the SPASE column. SPASE data model is available at: https://spase-group.org/data/model/index.html Baptiste Cecconi (LESIA, Observatoire de Paris, CNRS, PSL Research University) contributed the mapping here.

2021-10-4

- * Mingfang Wu checked latest version of Google dataset guidelines, DCATV3, BioSchema, DataCite, adding more property mappings, more property mappings are also added for the dc&dc terms, RIF-CS, Dataverse ECRIN.
- * Mingfang Wu created the Tab NISO classification for classifing mapped Schema.org properties into the NISO categories: Descriptive Metadata (rows in green), Administrative Metadata (rows in yellow) and Structure Metadata (rows in pink). NISO classification reference is available: http://www.niso.org/press-releases/2017/01/niso-publishes-understanding-metadata-primer

2021-11-13

* Stephen M. Richard (US Geoscience Information Network, USGIN) added the mapping from ISO19115-1

For ongoing updated version, please visit: https://github.com/rd-alliance/Research-Metadata-Schemas-WG

Visualisation tool (https://rd-alliance.github.io/Research-Metadata-Schemas-WG)

The World Data System (International Technology Office) has developed a visulisation tool, which provides a user-friendly interface to inspect the crosswalks, by querying either an individual property or a schema name.

Parent Type	Property R: Required	Туре	Description (text in bracket is from the	DCAT-AP	DCATv3	<u>Datacite</u>	ISO19115-1	EOSC/EDMI	<u>Dataverse</u>	<u>DATS</u>	RIF-CS	DC and DC Term	<u>BioSchema</u>	<u>B2FIND</u>	<u>DDI</u>	ECRIN Clinical Research	CodeMeta	SPASE
	n. nequired		Google dataset search guide)			M: Mandatory R: Recommended O: Optional		M: Minimum R: Recommended O: Optional								Research Metadata Schema		
schema:Thing	description (R)	Text	A description of the item.	dct:description(M)	dcterms:description	Description (R)	MD_Identification/abstract (M)	description (M)	Description; DescriptionValu e (M)		collection/description[@type='brief'] OR collection/description[@type='full']	Description - dcterms: description	description (M)	Description (R)	abstract	Description (R)	description	Reso r/Des
schema:Thing	name (R)	Text	A descriptive name of the item (e.g. dataset, software, Organization).	dct:title(M)	dcterms:title	Title (M)	MD_Identification/citation//title(M)	name (M)	Title (M)	title (M)	registryObject:collection:name	Title - dcterms:title	name (M)	Title (M)	titl	Object Title (M)	name	Res r/Re e
Schema:DataDownload	distribution. contentUrl	URL	(The link for the download.)	dct:accessURL (M) dct:downloadURL (O)	dcat:downloadURL		MD_Metadata/distributionIn fo//online//linkage [function = download]	accessURL			location/address/electronic @type="url" @ target="directDownload"		distribution					Acce on/A URL
schema:Thing	identifier	Text	The identifier property represents any kind of identifier for any kind of Thing, such as ISBNs, GTIN codes, UUIDs etc. Schema.org provides dedicated properties for representing many of these, either as textual strings or as URL (URI) links. See background notes for more details.		dcterms:identifier	Identifier (M)	MD_Identification/citation//identifier//code	identifier (M)	Dataset Persistent ID	identifier (R)	collection/citationInfo/citationMetadata/identifier AND/OR collection/identifier		identifier (M)	Identifier (M)	IDNo	DOI (M)	identifier	Reso r/DO Res
schema:Thing	alternateName	Text	An alias for the item (Alternative names that have been used to refer to this dataset, such as aliases or abbreviations.)		rdfs:label skos:altLabel	titleType	MD_Identification/citation//a IternateTitle(O)		Alternative title			dcterms:alternative	alternateName		altTitl			Reser/Alte
schema:Thing	sameAs	URL	URL of a reference Web page that unambiguously indicates the item's identity. E.g. the URL of the item's Wikipedia page, Wikidata entry, or official website.		owl:sameAs skos:exactMatch	RelatedIdentifer(R) AlernateIdentifier (O)*	MD_Identification/citation//i dentifier//code (use authority or codespace in MD_Identifier to indicate context for this alternate identifier)	sameAs (O)	Alternative URI	alternateIdentifi er (O)		occinis and market				Related resource identifiers, with relation type indicator (R)	sameAs	Reser/Info
schema:Thing	url	URL	URL of the item. (Location of a page describing the dataset.)	foaf:homepage	dcat:endPointURL	valueURL	distributionInfo//transferOptions//onLine//linkage [function = information]	url (M)		identifier (R)	RDA key= URL of record. "https://researchdata.ands.org.au/view?key= <insert key="">"</insert>	dc:identifier (if URL)	url (M)		accsPlac		url	Accon/a
schema:CreativeWork	citation	CreativeWork or URL	A citation or reference to another creative work, such as another publication, web page, scholarly article, etc. (Identifies academic articles that		dcterms: bibliographicCitatio n	isReferencedBy isDocumentedBy	MD_Identification/associate dResource//CI_Citation (suggest using associationType from dataCite vocabulary)		Related Publication		collection/relatedInfo [@type='publication'] OR relatedObject:collection where related collection [@type='publication']	dcterms:isReferencedBy dc:relation	citation (R)		relPubl(biblCit)		citation	Res r/Pu
			are recommended by the data provider be cited in addition to the dataset itself. Provide the citation for the dataset itself with other properties, such as name, identifier, creator, and publisher property can uniquely identify a related academic publication such as a data descriptor, data paper, or an article for which this dataset is supplementary material for.)															
schema:CreativeWork	keywords	Text	Keywords or tags used to describe this content. Multiple entries in a keywords list are typically delimited by commas.		dcat:keyword	Subject(R)*	MD_Identification/descriptiv eKeywords//keyword	keywords (R)	Subject (M); Topic Classification Term; Keywords	keywords (O)	collection/subject	Subject - dcterms:subject	keywords (M)		keywords		keywords	ke
schema:CreativeWork	license	CreativeWork or URL	A license document that applies to this content, typically indicated by URL. (A license under which the dataset is distributed.)		dct:license	Rights	MD_Identification/resource Constraints//reference/CI_C itation, or text in MD_LegalConstraints/useLi mitation [restrictionCode = license]			licenses (R)	collection/rights/licence[@rightsURI] AND/OR collection/rights/licence[@type] AND collection/rights/licence	Rights	license (R)	Rights (O)	rights	Rights (O)	license	
schema:CreativeWork	creator	Organization or Person	The creator/author of this CreativeWork (dataset). This is the same as the Author property for CreativeWork. (To uniquely identify individuals, use ORCID ID as the value of the sameAs property of the Person type. To uniquely identify institutions and organizations, use ROR ID.)	dct:creator	dcterms:creator	Creator (M)	MD_Identification/citation//c tedResponsibleParty//name [role = one of {author, coAuthor, originator, editor}]		Author; authorName (M)	creator (M)	collection/citationInfo/citationMetadata/contributor OR relatedObject relatedInfo party/name where relation=IsPrincipalInvestigatorO OR relatedObject relatedInfo party/name where relation=author OR relatedObject relatedInfo party/name where relation=colnvestigator OR relatedObject relatedInfo party/name where relation=colnvestigator OR relatedObject relatedInfo party/name where relation=hasCollector		creator (R)	Creator (R)	AuthEnty*	Creators (M)	author	Res r/Cc [@r Ilnv Res r/Cc [@r duc
schema:CreativeWork	isPartOf	CreativeWork	Indicates a CreativeWork that this CreativeWork is (in some sense) part of. Reverse property hasPart. If the dataset is a collection of smaller datasets, use the hasPart property to denote such relationship. Conversly, if the dataset is part of a larger dataset, use isPartOf.		dcterms:isPartOf	isPartOf	MD_Identifcation/associate dResource/name/CI_Citatio n [associationType = 'largerWorkCitation']	includedIn(Dataset)			relatedObject relatedInfo collection where relation[@type="isPartOf"]	dcterms:isPartOf			isPartOf		isPartOf	Pa for res

Parent Type	Property R: Required	Туре	Description (text in bracket is from the	DCAT-AP	DCATv3	<u>Datacite</u>	ISO19115-1	EOSC/EDMI	<u>Dataverse</u>	DATS	RIF-CS	DC and DC Term	BioSchema	B2FIND	DDI	ECRIN Clinical Research	CodeMeta	SPASE
			Google dataset search guide)			M: Mandatory R: Recommended O: Optional		M: Minimum R: Recommended O: Optional								Metadata Schema		
schema:CreativeWork	version	Number or Text	The version of the CreativeWork embodied by a specified resource.	owl:versionInfo	owl:versionInfo	Version (O)	MD_Identification/citation//e dition	version (O)	Version	version (R)	registryObject:collection:citationInfo: citationMetadata:version		version (R)		version	version (O)	version	Provider
schema:CreativeWork	temporalCovera ge	a Text		dct:temporal	dcterms:temporal	Date	MD_Identification/extent//te mporalElement/extent/TM_ Primitive		Time Period Covered		collection/coverage/temporal	dcterms:temporal (start); dcterms:temporal (end)		Temporal Coverage (O)	temporal			Tempora ption/Tin
schema:CreativeWork	spatialCoverag	e Text, Place	dataset has a temporal dimension.) The spatialCoverage of a CreativeWork indicates the place (s) which are the focus of the content. (You can provide a single point that describes the spatial aspect of the dataset. Only include this property if the dataset has a spatial dimension.)	dct:spatial	dcterms:spatial	GeoLocation(R)	MD_Identification/extent//ge ographicElement . EX_GeographicBounding box or EX BoundingPolygon with GM_Object is Point for centroid (yukkk!:(). Some profiles recommend putting place names in keywordS, keywordTypeCode = 'place'. Many possibilities	spatialCoverage (O)			collection/coverage/spatial	dcterms:coverage (spatial)		Spatial Coverage (O)	geogCover geoBndBox coverage			Observe n
schema:Dataset	measurementTo	a Toyt LIDI	A technique er technology used		0000		will have to configure for particular ISO profile MD Metadata/resourceLine	magauramentTeehn	:				measurementT		instrumentDeve	1		Measur
	chnique	e Text, OILL	A technique or technology used in a Dataset (or DataDownload, DataCatalog), corresponding to the method used for measuring the corresponding variable(s)		sosa: usedProcedure		age//statement	ques (R)*					echnique (R)		opment			pe
schema:Dataset	variableMeasur ed	PropertyValue , Text	The variableMeasured property can indicate (repeated as necessary) the variables that are measured in some dataset, either described as text or as pairs of identifier and description using PropertyValue.		sosa: observedProperty		MD_Metadata/contentInfo//f eatureCatalogue [requires import of ISO 19110]			dimensions (C			variablesMeas red (R)	1	measure			Parame
			(The variable that this dataset measures. For example, temperature or pressure.)															
schema:Dataset	includedInData Catalog	DataCatalog	A data catalog which contains this dataset. (The catalog to which the dataset belongs.)		dcterms:isPartOf	publisher	MD_Identification/associate dResource/name/CI_Citatio n [associationType = 'collectiveTitle']			distribution. storedIn (O)		dcterms:isPartOf	includedInData Catalog (M)					
Schema:Dataset	distribution	DataDownloa d		dct:distribution (R)	dcat:distribution	distribution	MD_Metadata/distributionIn fo/MD_Distribution/transfer Options	distribution (R)		distribution (R	location/address/electronic @type="url" @ target="directDownload"		distribution		distrbtr (URI)			
Schema:DataDownload	distribution. encodingForma	at	(The file format of the distribution.)	dcterms:format	dcat:mediaType dcterms:format	format	MD_Metadata/distributionIn fo/MD_Distribution/transfer Options//distributionFormat						encordingForm at					Access on/For
							115 11 05 0 0					dcterms:format						4
schema:CreativeWork	funder	Organization or Person	A person or organization that supports (sponsors) something through some kind of financial contribution.		dcterms:contributor	0	MD_Identification/credit or MD_Identification/citation//citedResponsibleParty//name [role = funder]		Grant information	(0)	relatedObject relatedInfo party/name where relation=isFundedBy	dcterms:contributor			grantNo (agency)		funder	r/Fundi
schema:CreativeWork	publisher	Organization or Person	The publisher of the creative work	dct:publisher (R)	dcterms:publisher	Publisher (M)	MD_Identification/citation//citedResponsibleParty//name [role = publisher]		Dataset Publisher		registryObject:collection:citationInfo: citationMetadata:publisher OR registryObject@Group	Publisher - dcterms:publisher		Publisher (R)	publisher	Publisher/ Provider (M)	publisher	Resou r/Conta [@role er]
schema:Thing	contactPoint			dct:contactPoint (R)	dct:contactPoint (R)		MD_Identification/pointOfC ontact//name (for the resource);					dcterms:mediator	dct:contactPoir	t dct:contactPoint	contact	Access Contact (M)	t .	Resou r/Cont
schema:CreativeWork or schema:Thing	about or subjectOf	CreativeWork, Event, Communication		dcat:theme (R) dct:subject(O)	dcat:theme (R)	Subject*	MD_Metadata/contact//nam MD_Identification/topicCate gory; additional categories in MD_Identification/descriptiv eKeywords//keyword [type = subTopicCategory]				collection/subject							
schema:Thing	inLanguage	Language or Text	The language of the content or performance or used in an action. Please use one of the language codes from the IETF BCP 47 standard.	dct:language (R)	dcterms:language		MD_Metadata/language (for metadata record content language); MD_Dataldentification/lang uage (for resource content language)		Language		(default to English) "en"	Language - dcterms:language		Language (O)	language	Language (R)		
schema:CreativeWork	dateCreated	Date or DateTime	The date on which the CreativeWork was created or the item was added to a DataFeed.		dcterms:created(O)	Date (R)/Created		dateCreated (M)	Date of collection		collection/citationInfo/citationMetadata/cate[@type='created'] OR collection/dates[@type='created']	dcterms:created	dateCreated		created	Dates (date type) (O)	datedCreated	

	Property R: Required	Туре	Description (text in bracket is from the	DCAT-AP	DCATv3	<u>Datacite</u>	ISO19115-1	EOSC/EDMI	<u>Dataverse</u>	DATS	RIF-CS	DC and DC Term	BioSchema B2	FIND DDI		<u>CodeMeta</u>	SPASE
	R: Required		(text in bracket is from the Google dataset search guide)			M: Mandatory R: Recommended O: Optional		M: Minimum R: Recommended O: Optional							Research Metadata Schema		
schema:CreativeWork	datePublished	Date	Date of first broadcast/publication.	dct:issued (R)	dcterms: dateAccepted	PublicationYear (M)	same as above, dateType = publication	datePublished (O)	Publication Date		collection/citationInfo/citationMetadata/date[@type='publicationDate'] OR collection/citationInfo/citationMetadata/date[@type='issued'] OR collection/dates[@type='dc.issued'] OR collection/dates[@type='dc.available'] OR date the record was ingested into RDA		Pul (R)	olicationYear distD		ear datePublished	Release
schema:CreativeWork	dateModified	Date or DateTime	The date on which the CreativeWork was most recently modified or when the item's entry was modified within a DataFeed.		dcterms:modified	Date (R)/Updated *	same as above, dateType = lastUpdate	dateModified (M)	Version Date		date the record was ingested into NDA	dcterms:modified	dateModified	versi	Dates (date) by Dates (date) (O)	dateModified	Revision Revision Release
schema:CreativeWork	copyrightHolder		The party holding the legal copyright to the CreativeWork.	dct:rightsHolder (O)	dcterms: rightsHolder		MD_Identification/resource Constaints//responsiblePart y//name [restrictionCode = copyright]					dcterms:rightsHolder				copyrightHold	er
schema:CreativeWork	mentions	Thing	Indicates that the CreativeWork contains a reference to, but is not necessarily about a concept.	dct:relation	dct:relation		MD_Identifcation/associate dResource/name/CI_Citatio n [associationType = 'crossReference']					dcterms:relation					
schema:CreativeWork	isBasedOn		A resource from which this work is derived or from which it is a modification or adaption	dct:source(O)	dct:source(O)	isContinuedBy isCompiledBy isOriginalFromOf isSourceOf isDerivedFrom	MD_Metadata/resourceLine age//source//sourceCitation					dc:source		deriv	ation		
schema:Thing	isRelatedTo	Product, Service	A pointer to another, somehow related product (or multiple products).	dct:relation	dct:relation	relationType	MD_Identifcation/associate dResource/name/CI_Citatio n [associationType = 'crossReference']					dct:relation					
schema:CreativeWork	encodingForma	Text, URL	Media type typically expressed using a MIME format	dct:format (R) dct:mediaType(O)	dct:format (R) dct:m	Format(O)*	MD_Metadata/distributionIn fo//transferOptions//distribut ionFormat//medium//name	format (R)		distribution. format	collection/address/electronic [@target='directDownload']/mediaType	dct:format (R) dct:mediaType(C))	forma medi		encordingForr at	n Acces
schema:MediaObject	contentURL	URL	Actual bytes of the media object, for example the image file or video file.	dcat:accessURL (M) dcat:downloadURL(O)	dcat:downloadURL		MD_Metadata/distributionIn fo//online//linkage [function = download]			access (M)							
schema:MediaObject	contentSize	Text, URL	File size in (mega/kilo) bytes.	dct:byteSize(O)	dct:byteSize	Size(O)*	MD_Metadata/distributionIn fo//transferOptions//transfer Size (and unitsOfDistribution)	size(R)				collSize					
schema:Thing:Schedule	repeatFrequenc y	Duration, text	This term is proposed for full integration into Schema.org* Defines the frequency at which Events will occur according to a schedule Schedule. The intervals between events should be defined as a Duration of time.		dcterms: accrualPeriodicity		MD_Metadata/distributionIn fo//transferOptions//transfer Frequency					dcterms:accrualPeriodicity					
schema:CreativeWork	funder	Organization or Person	A person or organization that supports (sponsors) something through some kind of financial contribution.	dcterms:contributor	dcterms:contributor		MD_Identification/citation//citedResponsibleParty//name [role = funder]		Grant information	acknowledges (O)	relatedObject relatedInfo party/name where relation=isFundedBy	dcterms:contributor				funder	Resou r/Fund cy
schema:CreativeWork	contributor	Organization or Person	A secondary contributor to the CreativeWork or Event.	dcterms:contributor	dcterms:contributor	Contributor (R)	MD_Identification/citation//citedResponsibleParty//name [role = contributor]		Contributor			Contributor - dcterms: contributor (Depositor)	Col	ntributor (O) contr	butor Contributo	rs (O) Contributor	Resou r/Cont [@role
schema:CreativeWork	producer	Organization or Person	The person or organization who produced the work (e.g. music album, movie, tv/radio series etc.).	dcterms:contributor	dcterms:contributor	contributor with contributortype=" Producer"	MD_Identification/citation//citedResponsibleParty//name [role = https://schema.org/producer]		Producer			dcterms:contributor		produ	cer	producer	
schema:CreativeWork	editor	Person	Specifies the Person who edited the CreativeWork.				MD_Identification/citation//citedResponsibleParty//name [role = editor]									editor	
schema:CreativeWork	copyrightYear	Number	The year during which the claimed copyright for the CreativeWork was first asserted.		dcterms: dateCopyrighted	Copyrighted	MD_Identification/citation//d ate [dateType = https://schema.org/copyrightYear]					dcterms:dateCopyrighted				copyrightYear	
	isAccessibleFor Free	Boolean	A flag to signal that the publication is accessible for free.											aviSt data/		isAccessibleF Free	or
			rights	dct:rights(O)	dct:rights	Rights(O)*	MD_Identification/resource Constraints/MD_LegalCons traints				collection/rights	rights		rights discla	imer		
	in proposing: https://github. com/schemaorg	ı	access rights conformance result / conformity (datta quality)	dct:accessRights(O) dct:conformsTo	dct:accessRights dct:conformsTo		MD_Identification/resource Constraints/MD_LegalCons traints MD_Metdata/dataQualityInf o/DQ_DataQuality (import					dct:accessRights		acce	sRights		Acces
			has version	dct:hasVersion		hasVersion	ISO19157) MD_Identification/associate dResource							hasV	ersion		
			is version of	dct:isVersionOf		isVersionOf	MD_Identification/associate dResource							isVer	sionOf		
			provenance	dct:provenance	prov: qualifiedAttribution		MD_Metadata/resourceLine age							meth relati			Revisi

RDA Crosswalks from schemas to schema.org (V1.0, 2021-12-16)

Parent Type	Property R: Required	Туре	Description (text in bracket is from the Google dataset search guide)	DCAT-AP	DCATV3	M: Mandatory R: Recommended O: Optional	ISO19115-1	EOSC/EDMI M: Minimum R: Recommended O: Optional	<u>Dataverse</u>	DATS	<u>RIF-CS</u>	DC and DC Term	<u>BioSchema</u>	<u>B2FIND</u>	DDI	ECRIN Clinical Research Metadata Schema	<u>CodeMeta</u>	<u>SPASE</u>
													also referenced: https: //bioschemas. org/profiles/					