

Dublin Core / DataCite crosswalk

Following is a crosswalk of [Dublin Core Element Set](#) (the 'widely used' fields, a subset of what has become the [DCMI Metadata Terms](#)) with the [DataCite schema](#).

There are also three examples of SRD (I took the top three from Susan Makar's analysis: 13, JANAF, 69, Chemistry WebBook and 78, Atomic Spectroscopy Database) described using DataCite. I preferred DataCite to Dublin Core for this exercise since it was created specifically with data in mind, and not as a general purpose web-resource citation tool.

The only thing that seems a bit clunky to me with DataCite is the demand that the "Identifier" be a DOI. This makes perfect sense if you are DataCite, you get to make your rule, and DataCite is about connecting DOIs to resources. It makes less sense if you want to use the schema for your own purposes. I would recommend a redefinition of "Identifier" to include any reasonable identification scheme (including internal) and adding a property to keep track of what that scheme is (like you see on the DataCite element `AlternativeIdentifier`, for example).

I was initially bugged by not seeing an obvious place to put a URI for a resource. But it seems reasonable to use "AlternativeIdentifier" to serve this purpose (and stating "URI" as the value of "alternativeIdentifierType") – as long as the URI really points to the same exact resource that is being described in the record. I think this is true of the 3 SRD described in the spreadsheet, so I've gone ahead and done that. It would also be possible to put a URI in "Identifier" if we were to redefine that field as suggested in the paragraph above, with URI as a possible type of identifier.

The lack of an obvious place to put URI information is also found in the Dublin Core, but "identifier" could probably be used if necessary.

Characterizing coverage was something mentioned. DataCite has an element to describe geographic coverage, and Dublin core allows description of space or time in their "coverage" element. I would need to see some examples, but I think that DataCite would need to be expanded to allow this concept (we're potentially talking about wavelength ranges? what materials measurements may have been made on? And Dublin Core "coverage" would have to be redefined. And in both cases you would want the ability to specify the type of coverage.

These are my initial observations. Take a look at what follows and let me know if that brings up any questions/comments on your end. I'm happy to do additional work on this, but I'm not quite sure

Mandatory
Recommended
Optional

DC term*	DC Definition	DataCite term*	DataCite Definition	Properties
identifier	An unambiguous reference to the resource within a given context.	Identifier	The Identifier is a unique string that identifies a resource.	identifierType
creator	An entity primarily responsible for making the resource.	Creator	The main researchers involved in producing the data, or the authors of the publication, in priority order.	creatorName nameIdentifier nameIdentifierScheme schemeURI affiliation
title	A name given to the resource.	Title	A name or title by which a resource is known.	titleType
publisher	An entity responsible for making the resource available.	Publisher	The name of the entity that holds, archives, publishes prints, distributes, releases, issues, or produces the resource. This property will be used to formulate the citation, so consider the prominence of the role.	
		PublicationYear	The year when the data was or will be made publicly available.	
subject	The topic of the resource.	Subject	Subject, keyword, classification code, or key phrase describing the resource.	subjectScheme schemeURI
contributor	An entity responsible for making contributions to the resource.	Contributor	The institution or person responsible for collecting, managing, distributing, or otherwise contributing to the development of the resource.	contributorType contributorName nameIdentifier nameIdentifierScheme schemeURI affiliation
date	A point or period of time associated with an event in the lifecycle of the resource.	Date	Different dates relevant to the work.	dateType
language	A language of the resource.	<i>Language</i>	The primary language of the resource.	

type	The nature or genre of the resource. Recommended best practice is to use a controlled vocabulary such as the DCMI Type Vocabulary [DCMITYPE]. To describe the file format, physical medium, or dimensions of the resource, use the Format element.	ResourceType	A description of the resource. The format is open, but the preferred format is a single term of some detail so that a pair can be formed with the sub-property. Dataset/Census Data, where 'Dataset' is resourceTypeGeneral value and 'Census Data' is ResourceType value.	resourceTypeGeneral
		<i>Alternateldentifier</i>	An identifier or identifiers other than the primary Identifier applied to the resource being registered. This may be any alphanumeric string which is unique within its domain of issue. May be used for local identifiers. Alternateldentifier should be used for another identifier of the same instance (same location, same file).	alternateldentifierType
coverage	The spatial or temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which the resource is relevant.	GeoLocation	Spatial region or named place where the data was gathered or about which the data is focused	geoLocationPoint geoLocationBox geoLocationPlace
relation	Recommended best practice is to identify the related resource by means of a string conforming to a formal identification system.	RelatedIdentifier	Identifiers of related resources. These must be globally unique identifiers.	
		<i>Size</i>	Unstructured size information about the resource.	
format	The file format, physical medium, or dimensions of the resource.	<i>Format</i>	Technical format of the resource.	
		<i>Version</i>	The version number of the resource.	
rights	Information about rights held in and over the resource.	<i>Rights</i>	Any rights information for this resource.	rightsURI
description	An account of the resource.	Description	All additional information that does not fit in any of the other categories. May be used for technical information.	descriptionType

source	A related resource from which the described resource is derived.
--------	--

*<http://www.dublincore.org/documents/dces/>

* [DataCite-MetadataSchema_V31_Final_8-24-2015_0.pdf](#)

Note that some of the 'missing' DataCite terms can be found in the larger Dublin Core Metadata Terms set (<http://www.dublincore.org/documents/dcmi-terms/>). Here we are only looking at the smaller, more widely known, Dublin Core Elements set. (<http://www.dublincore.org/documents/dces/>)

DataCite term	SRD 13 (JANAF)	comments
Identifier		lacking - but required - looking for a DOI or similar
Creator	Thomas Allison	Took this from EDI "contactPoint"
Title	NIST-JANAF Thermochemical Tables - SRD 13	took name and appended SRD #, like EDI titles
Publisher	National Institute of Standards and Technology	
PublicationYear		2013
Subject	Thermochemical tables	
Contributor		
Date		Could give more specific date here if that makes sense
<i>Language</i>	en	
ResourceType	Dataset/Thermochemical Tables	"Dataset" comes from list, "Thermochemical Tables" is free text
<i>AlternateIdentifier</i>	http://kinetics.nist.gov/janaf/	Should add property alternateIdentifierType with value URI?
GeoLocation		
RelatedIdentifier		
<i>Size</i>		
<i>Format</i>		
<i>Version</i>	3rd edition	meant to be just numeric - does it matter that they refer to it as an 'edition' rather than a 'version'?
<i>Rights</i>	http://www.nist.gov/data/license.cfm	
Description	This database contains the data from the 3rd edition of the NIST-JANAF Thermochemical Tables.	

DataCite term	SRD 69 (Chemistry WebBook)	comments
Identifier		lacking - but required - looking for a DOI or similar
Creator	Peter Linstrom	
Title	NIST Chemistry WebBook - SRD 69	took name and appended SRD #, like EDI titles
Publisher	National Institute of Standards and Technology	
PublicationYear		2014
Subject	chemical data, thermochemical data, thermodynamic data, thermophysical data, enthalpy, entropy, heat capacity, heat of formation, chemical structure, ionization potential, thermochemistry, boiling point, vapor pressure, IR spectrum, mass spectrum, UV/Vis spectrum, retention index, InChI, InChIKey	
Contributor		
Date		
Language	en,es,fr,pt,cs	
ResourceType	Collection/Chemical data	
AlternateIdentifier	http://webbook.nist.gov/chemistry	Should add property alternateIdentifierType with value URI?
GeoLocation		
RelatedIdentifier		
Size		
Format		
Version		
Rights	http://www.nist.gov/data/license.cfm	

Description

The NIST Chemistry WebBook provides users with easy access to chemical and physical property data for chemical species through the internet. The data provided in the site are from collections maintained by the NIST Standard Reference Data Program and outside contributors. Data in the WebBook system are organized by chemical species. The WebBook system allows users to search for chemical species by various means. Once the desired species has been identified, the system will display data for the species.

DataCite term	SRD 78 (Atomic Spec)	comments
Identifier		lacking - but required - looking for a DOI or
Creator	Alexander Kramida, Yuri Ralchenko, and Joseph Reader, Edward B. Saloman	took the 'active' names from website - can also add affiliation and identifier for each name - DataCite allows more than one and recognizes ordering, DC wants "an entity"
Title	NIST Atomic Spectra Database - SRD 78	took name and appended SRD #, like EDI
Publisher	Atomic Spectroscopy Group, Physical Measurement Laboratory, National Institute of Standards and Technology	If it needs to be reduced - could take group and lab away and leave NIST
PublicationYear		2015
Subject	Atomic spectra, atomic ground state, atomic ionization energy, atomic transition probability, atomic energy levels	EDI entry lists all element names - might be overkill for this purpose
Contributor		Could list "past contributors" or "students contributing to data entry" - how far to go?
Date		Could give more specific date here if that makes sense
<i>Language</i>	en	
ResourceType	Dataset/Atomic Spectra	"Dataset" comes from list, "Atomic Spectra" is free text
<i>AlternateIdentifier</i>	http://www.nist.gov/pml/data/asd.cfm	Should add property alternateIdentifierType with value URI?
GeoLocation		
RelatedIdentifier	http://www.nist.gov/pml/pubs/atspec/index.cfm , http://www.nist.gov/pml/data/asd_contents.cfm , http://physics.nist.gov/PhysRefData/ASD/Html/help.html , http://www.nist.gov/pml/data/asbib/index.cfm	These are resources listed on the homepage (Intro to atomic spectroscopy, intro to ASD contents, help, bibliography)
<i>Size</i>		
<i>Format</i>		
<i>Version</i>	5	
<i>Rights</i>	https://www.nist.gov/data/license.cfm	

Description

This database provides access and search capability for NIST critically evaluated data on atomic energy levels, wavelengths, and transition probabilities that are reasonably up-to-date. The Atomic Spectroscopy Data Center has carried out these critical compilations. The Data Center is located in the Physical Measurement Laboratory at the National Institute of Standards and Technology (NIST).
