Department for Business, Energy & Industrial Strategy

HDRUK Health Data Research UK

METADATAWORKS

Descriptive Metadata Specification

Prepared with

ipcc INTERGOVERNMENTAL PANEL ON climate change

Document Control Information						
Version	Status	Author(s)	Date Created	Reviewer(s)	Date Reviewed	Notes
V1.0.0	FINAL	Adam Milward Courtney Irwin	22.11.2021	Adam Milward	10.12.2021	
Minimum Retention Period			Permanent			
Disposal Required			Archive when new version created			
Classification			Public			
Next Review due:			10.6.2022			

Purpose

The following document has been developed to refine and iterate an alpha metadata schema / standard for metadata that will be used to drive greater data integration, discoverability, and federation for IPCC members and the wider climate research community.

The specification is initially focused on Descriptive metadata (rather than structural / administrative).

Contents

Introduction

As part of the continued development of the IPCC Data Distribution Centre, climate data is archived and published in accordance with FAIR data principles. Descriptive Metadata about datasets should be made human and machine-readable, allowing researchers to understand what data has been used in assessment reports, as well as increasing the utility and utilisation of climate data globally.

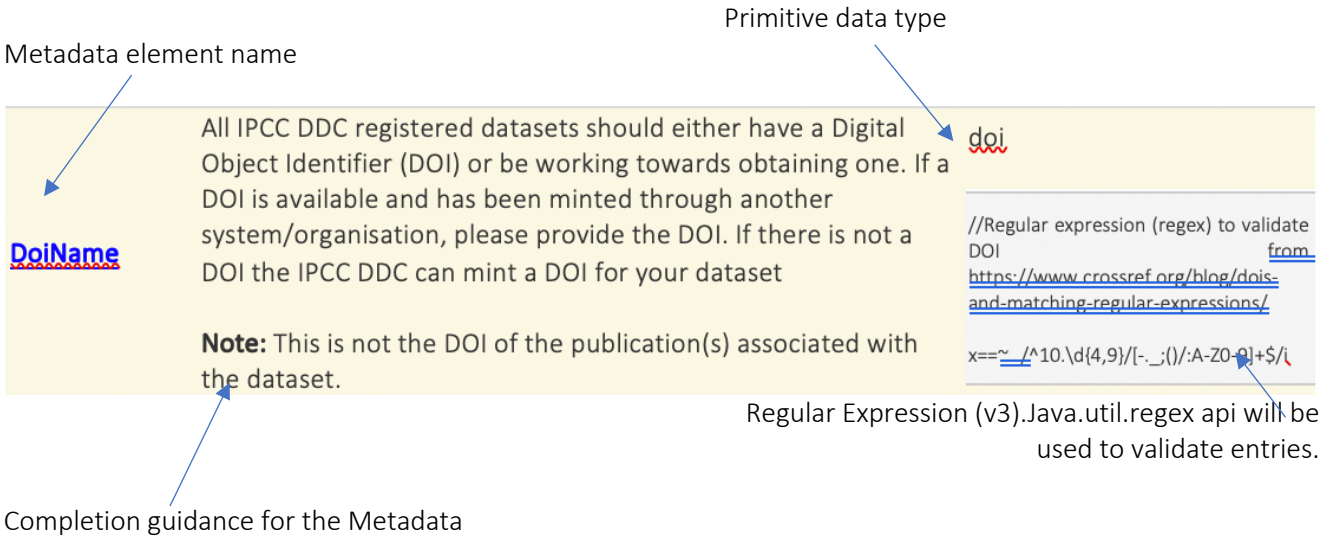
A metadata focus group has been established to agree and develop a standardised metadata schema that can be used by IPCC members to increase consistency and interoperability when archiving and cataloguing climate data. v0.1.0 of the Descriptive Metadata Schema is also being developed to support a minimum viable product for the DDC data catalogue. v0.1.0 will be developed as an 'alpha' with a view to collecting metadata and then reviewing submissions after 9 months. A wider consultation across users of the IPCC DDC can then be established to review submissions and make improvements to the schema that will allow users to better evaluate the utility ("fit for purpose x") of a dataset.

The following principles will govern the work and decisions made:

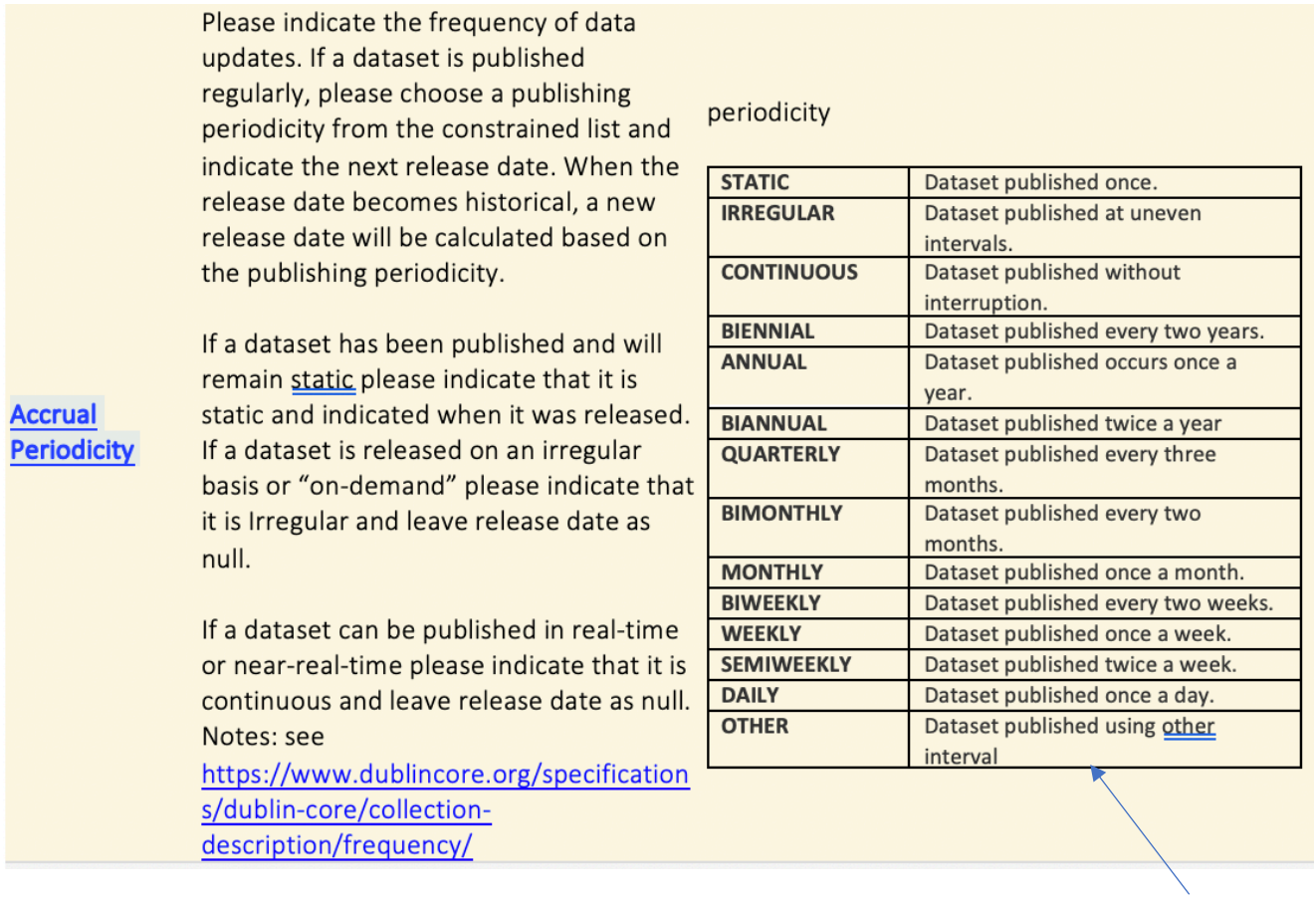
- Aim of v0.0.1 is to be "good enough" rather than perfect
- Metadata should be based on a user need and a user story
- Metadata should use standard terminologies whenever possible - DCAT is the de-Facto standard and links to Schema.org / Dublin Core provided where available
- If an identified use case does not have an equivalent DCAT entry, another standard should be selected
- Only if no standard can be found will a new metadata entry be created
- Naming convention: Metadata "titles" will be lowercase, underscore separated (note that in the human readable specification we are using Capitalised for ease of reading)
- Where standard attributes have been used, we will use the name from that standard. However, we may 'Label' the attributes in the user interface and provide guidance that makes the attribute more user friendly as many of our users will not be familiar with the technical terminology.
- Prioritisation of Metadata has been based on Impact vs Effort (effort needs to consider potential for automation in the longer-term vs manual in nearer term) however, this will be easier to quantify as the project progresses

Layout:

The document has been laid out in the following way:



Some metadata elements may have enumerated values rather than regular expressions to constrain the values provided:



Constrained list of values

Key terms:

Dataset: We are using the schema.org definition of a dataset as a "body of structured information describing some topic(s) of interest". Each data publisher is asked to reasonably evaluate what they consider a discrete dataset within the context of research use and discovery.

Label: In some cases, labels for attributes will be different from the underlying names used to store and interchange the information. Where possible, the name of the attribute is consistent with the underlying standard i.e. schema.org , whilst the label is representative of the user experience when they input the information.

User Story: Illustrates the purpose for collecting this metadata

As A [user]	Researcher
I Want To	Know the name of the dataset
So That	I can find, identify, and understand its relevance

Examples: Illustrative values that could be provided:

Examples	Biogeochemical Ocean Flux Study (BOFS) North Atlantic Data Set (1989-1991)

Priority: Indication of the priority of the metadata. Can be High, Medium or Low. This will be completed following consultation.

Priority	High

Completion Guidance: Metadata that has a min occurrence of >=1 or * is mandatory (* indicates many). If there is a min occurrence of 0 then the field is not mandatory. Some metadata may not be mandatory because it isn't applicable to an organisation, but completion is still be required, if it is relevant. Some metadata may not be mandatory, but completion is recommended.

Completion	Mandatory
Cardinality	Min Occurs: 1 Max Occurs: 1

Definition: Formal definition for the metadata, usually referencing a standard.

Definition	The name of the item. Canonical URL: http://schema.org/name Equivalent Property: dct:title

Schema.org alternative: If available, a schema.org alternative has been identified. This will allow us to provide structured information for search engines. The data catalogue will publish a JSON-LD profile for 'LIVE' models so it can be easily indexed by search engines e.g. Google.

Schema.org alternative	Schema.org Dataset/name

Survey metadata: A wider consultation after the alpha will include a survey for the next iteration of the schema.

Descriptive Metadata

Describes a resource for purposes of discovery and identification.

Note: the following section will be managed by the data catalogue internally however, the expectation is that a metadata record about a dataset should have the following attributes as a minimum.

Identifier	System dataset identifier. Resolvable, Persistent Dataset Identifier This is the persistent identifier that should resolve to the latest metadata revision.	uri or UUIDv4 // value is uri or uuid x ==~ /^((([A-Za-z]{3,9}:(?:\/\/)?)(?:[-;:&=\+\$,\w]+@)?[A-Za-z0-9.-]+(:[0-9]+)?|(?:www.|[-;:&=\+\$,\w]+@)[A-Za-z0-9.-]+)((?:\/[\+~%\/.\w-_]*)?\??(?:[-\+=&;%@.\w_]*)#?(?:[\w]*))?)|([a-fA-F0-9]{8}-[a-fA-F0-9]{4}-[a-fA-F0-9]{4}-[a-fA-F0-9]{4}-[a-fA-F0-9]{12})$/
Label	Resolvable Identifier	
Priority	High	
Completion	Mandatory	
Cardinality	Min Occurs: 1 Max Occurs: 1	
Note:	The identifier might be used as part of the URI of the item, but still having it represented explicitly is useful.	
Definition	A unique identifier of the item. RDF Property: dct:identifier Range: rdfs:Literal Source: dcat:resource_identifier Vocabulary: Data Catalog Vocabulary (DCAT)	
As A [user]	Researcher	
I Want To	Know the local identifier for the dataset	
So That	I can identify or refer to it	

Version	Dataset metadata version	semver / x.x.x x ==~ /^([0-9]+)\.([0-9]+)\.([0-9]+)$/
Example	1.1.0	
Cardinality	Min Occurs: 1 Max Occurs: 1	
Priority	High	
Title	Metadata Version	
Vocabulary	Dct:hasVersion	

Issued	Dataset metadata published date.	xs:dateTime

Title	Date Published
Cardinality	Min Occurs: 1 Max Occurs: 1
Vocabulary	Dcat:issued
Priority	High

Modified	Dataset metadata modification date	xs:dateTime
Title	Date Last Modified	
Cardinality	Min Occurs: 1 Max Occurs: 1	
Vocabulary	dcat:modified	
Priority	High	

Revisions

Revisions of Dataset metadata. This is a repeating section where all the revisions (versions) of the metadata can be captured and linked.

Version	Dataset Metadata version	semver / x.x.x x ==~ /^([0-9]+)\.([0-9]+)\.([0-9]+)$/
Cardinality	Min Occurs: 1 Max Occurs: 1	
Priority	High	
Vocabulary	dct:hasVersion	

url	URL endpoint to obtain the version	url ///Regular expression (regex) to validate URL: x ==~ /((([A-Za-z]{3,9}:(?:\/\/)?)(?:[-;:&=\+\$,\w]+@)?[A-Za-z0-9.-]+(:[0-9]+)?|(?:www.|[-;:&=\+\$,\w]+@)[A-Za-z0-9.-]+)((?:\/[\+~%\/.\w-_]*)?\??(?:[-\+=&;%@.\w_]*)#?(?:[\w]*))?)/
Cardinality	Min Occurs: 1 Max Occurs: 1	
Vocabulary	dcat:distribution	

Summary

Summary metadata must be completed by Data Custodians onboarding metadata.

Title	Title of the dataset limited to 180 characters. It should provide a short description of the dataset and be unique across the IPCC DDC. If your title is not unique, please add a prefix with your organisation name or identifier to differentiate it from other datasets within the IPCC DDC. Please avoid acronyms wherever possible. Good titles should summarise the content of the dataset and if relevant, the region the dataset covers and ideally will be 80 characters or under.	characters180 //Regular expression (regex) to validate and constrain title between 2 and 180 characters x==~ /^.{2,180}$/
Examples	Biogeochemical Ocean Flux Study (BOFS) North Atlantic Data Set (1989-1991)	
Completion	Mandatory	
Cardinality	Min Occurs: 1 Max Occurs: 1	
Definition	A name given to the item. Vocabulary: Data Catalog Vocabulary (DCAT) RDF Property: dct:title Range: rdfs:Literal Source: dcat:resource_title	
As A [user]	Researcher	
I Want To	Know the name of the dataset	
So That	I can find, identify, and understand its relevance	
Schema.org alternative	Schema.org Dataset/name	

Abstract	Provide a clear and brief descriptive signpost for researchers who are searching for data that may be relevant to their research. The abstract should allow the reader to determine the scope of the data collection and accurately summarise its content. The optimal length is one paragraph (limited to 255 characters) and effective abstracts should avoid long sentences and abbreviations where possible. **Note**: Researchers will view Titles and Abstracts when searching for datasets in the Catalogue before choosing whether to explore their content further. Abstracts should be shorter (255 characters) and different from the full description for a dataset (see next section).	Characters255 //Regular expression (regex) to validate and constrain title between 5 and 255 characters x==~ /^.{5,255}$/
Example	This data set, produced by the Copernicus Emergency Management Service for the European Forest Fire Information System, provides complete historical reconstruction of meteorological conditions favourable to the start, spread and sustainability of fires.	
Completion	Mandatory	
Cardinality	Min Occurs: 1 Max Occurs: 1	
Definition	A summary of the resource. Vocabulary: DCMI Metadata Terms RDF Type: rdf:Property Refines: dc:description	

	Source: dct:abstract
As A [user]	Researcher
I Want To	Read a plain English summary of the dataset
So That	I can identify if it is of interest
Schema.org alternative	Schema.org Dataset/abstract

Contact Point	Please provide a valid email address that can be used for questions about the data, further information, and enquiries. **Notes:** This could be either the author of a figure/dataset or a generic email address for the data providers institution (IEA, EDGAR etc.). If an employee/author email address is provided an explicit consent must be obtained for this purpose.	emailAddress //Regular expression (regex) to validate email addresses x ==~/[^\s]+@[^\s]+\.[^\s]+/
Examples	ipcc.ddc.datasupport@metadata.atlassian.net	
Completion	Mandatory	
Cardinality	Min Occurs: 1 Max Occurs: 1	
Definition	Relevant contact information for the catalogued resource. Use of vCard is recommended [VCARD-RDF] . Vocabulary: Data Catalog Vocabulary (DCAT) RDF Property: dcat:contactPoint Range: vcard:Kind Source: dcat:resource_contact_point	
As A [user]	Researcher	
I Want To	Know the contact point	
So That	I can contact them about the dataset and request access	
Schema.org alternative	https://schema.org/email	

Keywords	Please provide relevant and specific keywords that can improve the SEO of your dataset as a comma separated list. **Notes**: IPCC could encourage users to at a minimum provide tags for: assessment report (AR[x]), Working Group, Chapter, Figure, Input, Final, Intermediate	keywordList x==~ /^.{2,80}$/
Default	NA	
Examples	["Land", "Precipitation", "Extremes", "Europe", "Flood", "Impacts", "AR5"]	
Completion	Required	
Cardinality	Min Occurs: 0 Max Occurs: *	
Definition	A keyword or tag describing the resource. Vocabulary: Data Catalog Vocabulary (DCAT) RDF Property: dcat:keyword Range: rdfs:Literal Source: dcat:resource_keyword	
As A [user]	Researcher	
I Want To	Identify the keywords associated with the data	

So That	I can index and search data, as well as look up similar datasets

doi	All IPCC DDC registered datasets should either have a Digital Object Identifier (DOI) or be working towards obtaining one. If a DOI is available and has been minted through another system/organisation, please provide the DOI. If there is not a DOI the IPCC DDC can mint a DOI for your dataset **Note:** This is not the DOI of the publication(s) associated with the dataset.	doi //Regular expression (regex) to validate DOI from https://www.crossref.org/blog/dois-and-matching-regular-expressions/ x==~ /^10.\d{4,9}/[-._;()/:A-Z0-9]+$/i

Completion	Optional
Cardinality	Min Occurs: 0 Max Occurs: 1
Definition	Digital Object Identifier. Provides an actionable, interoperable, persistent link Actionable – through use of identifier syntax and network resolution mechanism (Handle System®) Persistent – through combination of supporting improved handle infrastructure (registry database, proxy support, etc) and social infrastructure (obligations by Registration Agencies) Interoperable – through use of a data model providing semantic interoperability and grouping mechanisms. Vocabulary: DOI Data Dictionary Source: DOI
As A [user]	Data custodian / researcher
I Want To	Provide a persistent interoperable identifier for a dataset
So That	I can uniquely identify it, globally, and provide a HTTP URI to navigate to metadata that describes it.

Alternate Identifiers	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.	characters180 //Regular expression (regex) to validate and constrain title between 2 and 80 characters x==~ /^.{2,180}$/

Completion	Optional
Cardinality	Min Occurs: 0 Max Occurs: *
Definition	See Section 11, page 18 of http://schema.datacite.org/meta/kernel-4.2/doc/DataCite-MetadataKernel_v4.2.pdf
As A [user]	Data custodian / researcher
I Want To	Provide a persistent interoperable identifier for a dataset
So That	I can uniquely identify it if it does not have a DOI or has several identifiers

Publication Date	The date this version of the dataset was published.	date xs:dateTime

Completion	Optional
Cardinality	Min Occurs: 0 Max Occurs: *
Definition	See http://schema.datacite.org/meta/kernel-4.2/doc/DataCite-MetadataKernel_v4.2.pdf
As A [user]	Data custodian / researcher

I Want To	Know the currency of the data
So That	I can understand its relevancy in relation to other data sources

Publisher

Completion Notes: This is the organisation responsible for making the data available, as well as publishing and maintaining the metadata.

Identifier	Please provide an identifier for your organisation
	
	Note: common institutions/affiliations can use a ROR (https://ror.org/) identifier and or re3data PIDs (https://www.re3data.org/). ROR replaces the older GRID (https://www.grid.ac).
		url
		
		//Regular expression (regex) to validate URL:
		x ==~ /((([A-Za-z]{3,9}:(?:\/\/)?)(?:[-;:&=\+\$,\w]+@)?[A-Za-z0-9.-]+(:[0-9]+)?|(?:www.|[-;:&=\+\$,\w]+@)[A-Za-z0-9.-]+)((?:\/[\+~%\/.\w-_]*)?\??(?:[-\+=&;%@.\w_]*)#?(?:[\w]*))?)/

Completion	Optional	
Cardinality	Min Occurs: 0 Max Occurs: 1	
Definition:	Global Research Identifier Database URI	
As A [user]	Data custodian / researcher / coordinator	
I Want To	Access standardised metadata about the organisation	
So That	I can understand the context of the organisation	

Name	Please provide the name of the publisher.	characters80 //Regular expression to validate min, max character limit x==~ /^.{2,80}$/
As A [user]	Data custodian / researcher / coordinator	
I Want To	Know the organisation responsible for submitting and managing the metadata and maintaining the data	
So That	I can identify who is responsible for the metadata	
Completion	Mandatory	
Cardinality	Min Occurs: 1 Max Occurs: 1	
Definition	The name of the item. Canonical URL: http://schema.org/name Equivalent Property: dct:title	

Logo	Please provide a logo associated with the Organisation using a valid URL. The following formats will be accepted .jpg, .png or .svg.	url //Regular expression (regex) to validate URL: x==~ /https?:\/\/(www\.)?[-a-zA-Z0-9@:%._\+~#=]{1,256}\.[a-zA-Z0-9()]{1,6}\b([-a-zA-Z0-9()@:%_\+.~#?&//=]*)/

Example	https://www.ipcc-data.org/img/DKRZ_Logo_281x127_2014.png
Completion	Optional
Cardinality	Min Occurs: 0 Max Occurs: 1
Definition	An associated logo. Canonical URL: http://schema.org/logo
As A [user]	Data custodian / researcher / coordinator
I Want To	Know the organisation responsible for submitting and managing the metadata
So That	I can identify who is responsible for the metadata

Description	Please provide a URL that describes the organisation.	url
		
		//Regular expression (regex) to validate URL:
		x ==~ /((([A-Za-z]{3,9}:(?:\/\/)?)(?:[-;:&=\+\$,\w]+@)?[A-Za-z0-9.-]+(:[0-9]+)?|(?:www.|[-;:&=\+\$,\w]+@)[A-Za-z0-9.-]+)((?:\/[\+~%\/.\w-_]*)?\??(?:[-\+=&;%@.\w_]*)#?(?:[\w]*))?)/

Examples	http://www.ciesin.columbia.edu/aboutus.html	
Completion	Optional	
Cardinality	Min Occurs: 0 Max Occurs: 1	
Definition	A description of the organisation Canonical URL: http://schema.org/description	
As A [user]	Data custodian / researcher / coordinator	
I Want To	Know more about the organisation responsible for submitting and managing the metadata	
So That	I can understand the context of the data and the submission	

Contact PointOrg	Please provide a valid email address that can be used to coordinate data enquiries with the publisher. Organisations are expected to provide a dedicated email address associated with the data access request process for their organisation. This will then be used as a default contact point for all datasets submitted by the organisation. However, there will be the opportunity to overwrite his value for each dataset.	emailAddress //Regular expression (regex) to validate email addresses x ==~/[^\s]+@[^\s]+\.[^\s]+/
Examples	ciesin.info@ciesin.columbia.edu	
Completion	Optional	
Cardinality	Min Occurs: 0 Max Occurs: 1	
Definition	Relevant contact information for the catalogued resource. Use of vCard is recommended [VCARD-RDF] . Vocabulary: Data Catalog Vocabulary (DCAT) RDF Property: dcat:contactPoint	

	Range: vcard:Kind Source: dcat:resource_contact_point
As A [user]	Researcher
I Want To	Know the contact point
So That	I can contact them about the dataset and request access
Schema.org alternative	https://schema.org/email

Documentation

Dataset documentation can include a rich text description of the dataset or links to media such as documents, images, presentations, videos or links to data dictionaries, profiles or dashboards. Organisations are required to confirm that they have permission to distribute any additional media.

Description	An HTML account of the data that provides context and scope of the data, limited to 10000 characters, and/or a resolvable URL that describes the dataset.
	
	Additional information can be recorded and included using media.
		charaters10000
		
		//Regular expression to validate min, max character limit : min 2, max 10000
		X==~ /.{2,10000}/

Completion	Optional	
Cardinality	Min Occurs: 0 Max Occurs: 1	
Definition	A free-text account of the record. Vocabulary: Data Catalog Vocabulary (DCAT) RDF Property: dc:description Range: rdfs:Literal Source: dcat:record_description See Schema.org Dataset:description	
As A [user]	Researcher	
I Want To	Know the description	
So That	I can understand information about the dataset	

Associated Media	Please provide any media associated with the dataset using a valid URI for the content. This is an opportunity to provide additional context that could be useful for researchers wanting to understand more about the dataset and its relevance to their research question.
	
	The following formats will be previewed .jpg, .png or .svg, .pdf, .xslx or .docx. Other formats can be used as url links.
	
	Note: media asset can be hosted by the organisation or uploaded
		urlList
		
		//Regular expression (regex) to validate URL:
		x==~ /https?:\/\/(www\.)?[-a-zA-Z0-9@:%._\+~#=]{1,256}\.[a-zA-Z0-9()]{1,6}\b([-a-zA-Z0-9()@:%_\+.~#?&//=]*)/

Notes	This could be a map, figure or further detail about the datasets, or a graph or charts that provides further context about the dataset.	
Completion	Optional	
Cardinality	Min Occurs: 0 Max Occurs: *	
Definition	A media object that encodes this CreativeWork. This property is a synonym for encoding. Vocabulary: https://schema.org/associatedMedia	
As A [user]	Researcher	
I Want To	Know the description	
So That	I can understand information about the dataset	
HOP Notes	Users must confirm that they have permission to publish any media	

Is Part Of	Please complete only if the dataset is part of a group or collection of datasets	charaters80

	//Regular expression to validate min, max character limit x==~ /^.{2,80}$/
Examples	AR6
Completion	Optional
Cardinality	Min Occurs: 0 Max Occurs: *
Definition	Indicates an item or Creative Work that this item, or Creative Work (in some sense), is part of. Inverse-property: hasPart.
As A [user]	Researcher
I Want To	Know whether the dataset is part of a wider group
So That	I can find other datasets of the group

Coverage

This information includes attributes for geographical and temporal coverage, cohort details etc. to enable a deeper understanding of the dataset content so that researchers can make decisions about the relevance of the underlying data.

Spatial Coverage	A spatial region, area or named place covered by the dataset	charaters80 //Regular expression to validate min, max character limit x==~ /^.{2,80}$/
Example	Europe	
Label	Spatial Coverage	
Completion	Optional	
Cardinality	Min Occurs: 0 Max Occurs: 1	
Definition	Usage Note: The spatial coverage of a dataset may be encoded as an instance of dct:Location or may be indicated using a URI reference (link) to a resource describing a location. Vocabulary: Data Catalog Vocabulary (DCAT) RDF Property: dct:spatial Range: dct:Location (A spatial region or named place) Source: dcat:dataset_spatial See also: Schema.org Dataset/spatialCoverage	
As A [user]	Researcher	
I Want To	Know the geographic are covered by the dataset	
So That	I can judge whether it covers the region I am interested in	

Spatial Aggregation	A statement describing the type of spatial aggregation used in the dataset (eg. region, area, grid, etc.)	charaters80 //Regular expression to validate min, max character limit x==~ /^.{2,80}$/
Example	Global, Continent, Country, Other, Gridded	
Completion	Optional	
Cardinality	Min Occurs: 0 Max Occurs: 1	
Definition		
As A [user]	Researcher	
I Want To	Know the spatial aggregation of the dataset	

So That	I can judge whether it is suitable for my research (e.g., is not aggregated globally if I am interested in a specific region).

Spatial Resolution	A statement of the amount of detail provided to sufficiently assess fitness for purpose. Precision can be expressed as a distance or expressed without unit of measure specifying spatial resolution by means of an "equivalent scale" with a fraction (e.g., 1:1,000, 1:1,000,000) or alternatively specifying the angular distance. Where the spatial resolution of the dataset is not the original resolution of the data (e.g., the resolution at which a climate model was run), you can note that in the description.	charaters80 //Regular expression to validate min, max character limit x==~ /^.{2,80}$/
Example	Reanalysis: 0.25° x 0.25° Mean, spread and members: 0.5° x 0.5°	
Completion	Optional	
Cardinality	Min Occurs: 0 Max Occurs: 1	
Definition	dqv:isMeasurementOf :spatialResolutionAsDistance ; dqv:isMeasurementOf :spatialResolutionAsEquivalentScale ; dqv:isMeasurementOf :spatialResolutionAsAngularDistance ; see https://www.w3.org/TR/vocab-dqv/#ExpressDatasetAccuracyPrecision for difference between precision and accuracy.	
As A [user]	Researcher	
I Want To	Know the spatial resolution of the dataset	
So That	I can judge whether it is suitable for my research (e.g., sufficiently highly resolved for my purpose but not unnecessarily high given data download, storage, and computing efficiency)	

Start Date	The start of the time period that the dataset provides coverage for. If there are multiple cohorts in the dataset with varying start dates, please provide the earliest date and use the description or the media attribute to provide more information.	date xs:dateTime
Completion	Mandatory	
Cardinality	Min Occurs: 1 Max Occurs: 1	
Definition	The start of the period. Vocabulary: Data Catalog Vocabulary (DCAT) RDF Property: dcat:startDate Domain: dcat:PeriodOfTime Range: rdfs:Literal encoded using the relevant ISO 8601 Date and Time compliant string [DATETIME] and typed using the appropriate XML Schema datatype [XMLSCHEMA11-2] (xsd:gYear , xsd:gYearMonth , xsd:date , or xsd:dateTime). Source: dcat:period_start_date See also: Schema.org Dataset/temporalCoverage	
As A [user]	Researcher	
I Want To	Know the time period from when data is available	
So That	I can use data that covers the time period I am interested and understand if there is enough history	

End Date	The end of the time period that the dataset provides coverage for. If the dataset is “Continuous” and has no known end date, please leave blank. If there are multiple cohorts in the dataset with varying end dates, please provide the latest date and use the description or the media attribute to provide more information.	date xs:dateTime
Completion	Optional	
Cardinality	Min Occurs: 0 Max Occurs: 1	
Definition	The end of the period. Vocabulary: Data Catalog Vocabulary (DCAT) RDF Property: dcat:endDate Domain: dcat:PeriodOfTime Source: dcat:period_end_date Range: rdfs:Literal encoded using the relevant ISO 8601 Date and Time compliant string [DATETIME] and typed using the appropriate XML Schema datatype [XMLSCHEMA11-2] (xsd:gYear , xsd:gYearMonth , xsd:date , or xsd:dateTime). Source : Schema.org Dataset/temporalCoverage	
As A [user]	Researcher	
I Want To	Know the time period until when data is available	
So That	I can use data that covers the time period I am interested in and understand if there is enough history.	

Temporal Resolution	Minimum time period resolvable in the dataset distribution.	temporalResolutionEnum STATIC: Static dataset ANNUAL: Annual interval BIANNUAL: Biannual interval QUARTERLY: Quarterly interval SEASONALLY: By climatological season (DJF, MAM, JJA, SON) BIMONTHLY: Bimonthly interval MONTHLY: Monthly interval BIWEEKLY: Biweekly interval WEEKLY: Weekly interval SEMIWEEKLY: Semiweekly interval DAILY: Daily interval 6HOURLY: Interval of 6 hours HOURLY: Hourly interval OTHER: Dataset resolution using other interval
Example	Daily, Monthly, Annually	
Completion	Mandatory	
Cardinality	Min Occurs: 0 Max Occurs: 1	
Definition	RDF Property: dcat:temporalResolution Definition: Minimum time period resolvable in the dataset distribution. Range: xsd:duration	

	Usage note: If the dataset is a time-series this should correspond to the spacing of items in the series. For other kinds of dataset, this property will usually indicate the smallest time difference between items in the dataset. Usage note: Alternative temporal resolutions might be provided in different dataset distributions
As A [user]	Researcher
I Want To	Know the temporal resolution of the dataset
So That	I can judge whether it is suitable for my research (e.g., sufficiently highly resolved for my purpose but not unnecessarily high given data download, storage, and computing efficiency)

Geographic Bounding Box

The geographic bounding box describes the spatial coverage in terms of a square as specified by the longitude and latitude values of the lower left and upper right corners.

Please enter the four values in terms of degrees North or East, with longitudes given relative to the prime meridian (yielding values between -180E and 180E).

Note that a "north up" area is described so that generally the upper right longitude will be greater than the lower left longitude.

lowerLeftLatitude	The geographic bounding box describes a "north up" area. The lower left latitude will be smaller than the upper right latitude.	string //Regular expression to validate latitude ^[+-]?(([1-8]?[0-9])(\.[0-9]{1,6})?|90(\.0{1,6})?)$
Examples	-90.0000	
Completion	Optional	
Cardinality	Min Occurs: 0 Max Occurs: 1	

lowerLeftLongitude	The geographic bounding box describes a "north up" area. Generally, the lower left longitude will be smaller than the upper right longitude. However when the area crosses the 180° meridian, the value of the lower left longitude will be greater than the value of the upper right longitude.	string //Regular expression to validate longitude ^[+-]?((([1-9]?[0-9]|1[0-7][0-9])(\.[0-9]{1,6})?)|180(\.0{1,6})?)$
Examples	-180.0000	
Completion	Optional	
Cardinality	Min Occurs: 0 Max Occurs: 1	

upperRightLatitude	The geographic bounding box describes a "north up" area. Upper right latitude will be greater than the lower left latitude.	string //Regular expression to validate latitude ^[+-]?(([1-8]?[0-9])(\.[0-9]{1,6})?|90(\.0{1,6})?)$

Examples	90.0000
Completion	Optional
Cardinality	Min Occurs: 0 Max Occurs: 1

upperRightLongitude	The geographic bounding box describes a "north up" area. Generally the upper right longitude will be greater than the lower left longitude. However when the area crosses the 180° meridian, the value of the lower left longitude will be greater than the value of the upper right longitude.	string //Regular expression to validate longitude ^[+-]?((([1-9]?[0-9]|1[0-7][0-9])(\.[0-9]{1,6})?)|180(\.0{1,6})?)$
Examples	180.000	
Completion	Optional	
Cardinality	Min Occurs: 0 Max Occurs: 1	

Provenance

Provenance information allows researchers to understand data within the context of its origins and can be an indicator of quality, authenticity, and timeliness.

Purpose	Please indicate the purpose(s) that the dataset was collected.	characters3000
		//Regular expression to validate min, max character limit
		x==~ /^.{2,3000}$/

Examples	Project Description, Long Term Observations X, Study Y	
Completion	Optional	
Cardinality	Min Occurs: 0 Max Occurs: 1	
As A [user]	Researcher	
I Want To	Know the why the data was collected	
So That	I can understand if it is relevant to my use cases	
Definition	The purpose of the ResourcePackage, why the ResourcePackage took place. This should include detailed information on the investigator's primary ResourcePackage questions or hypotheses as well as information on any legal basis for the data collection, such as laws requiring the collection of census data for apportionment purposes. Legal or other authorization should be provided in detail within AuthorizationSource. Supports multiple language versions of the same content as well as optional formatting of the content. Vocabulary: DDI: https://ddialliance.org/Specification/DDI-Lifecycle/3.3/XMLSchema/FieldLevelDocumentation/	

Source	Please indicate the source of the data extraction.	characters3000 //Regular expression to validate min, max character limit x==~ /^.{2,3000}$/
Examples	Raw Data, Secondary Data	
Completion	Optional	
Cardinality	Min Occurs: 0 Max Occurs: 1	
Definition	A related resource from which the described resource is derived. From: https://dublincore.org/specifications/dublin-core/dcmi-terms/#source	
As A [user]	Researcher	
I Want To	Know the source of the data	
So That	I can understand resulting strengths and weaknesses e.g. biases and other properties of the data	

Accessibility

Accessibility information allows researchers to understand access, usage, limitations, formats, standards and linkage or interoperability with toolsets. This section includes information about how the data can be used and how it is currently being used.

Usage

This section includes information about how the data can be used and how it is currently being used.

License	Legal document under which the distribution of data is made available.
	
	Usage note: Information about licenses and rights SHOULD be provided on the level of Distribution. Information about licenses and rights MAY be provided for a Dataset in addition to but not instead of the information provided for the Distributions of that Dataset. Providing license or rights information for a Dataset that is different from information provided for a Distribution of that Dataset SHOULD be avoided as this can create legal conflicts.
		url
		
		//Regular expression (regex) to validate URL:
		x==~ /https?:\/\/(www\.)?[-a-zA-Z0-9@:%._\+~#=]{1,256}\.[a-zA-Z0-9()]{1,6}\b([-a-zA-Z0-9()@:%_\+.~#?&//=]*)/

Example	https://creativecommons.org/licenses/by/3.0/legalcode	
Completion	Mandatory	
Cardinality	Min Occurs: 1 Max Occurs: 1	
As A [user]	Researcher	
I Want To	Know the license that applies to the data	
So That	I Know the legal information that describes how the distribution is made available	
Definition	Vocabulary: Data Catalog Vocabulary (DCAT) RDF Property: dct:license Range: dct:LicenseDocument Source: dcat:distribution_license Source: Dataset/license	

Resource Creator	The main researchers involved in producing the data, or the authors of the publication, in priority order. To supply multiple creators, repeat this property. **Note**: May be a corporate/institutional or personal name and more than one author be provided.	charaters1000List // value is between 2 and 1000 characters x ==~ /^.{2,1000}$/
Example	van Donkelaar, A.	
Completion	Mandatory	
Cardinality	Min Occurs: 1 Max Occurs: *	
Definition	The entity responsible for producing the resource. Usage Note: Resources of type foaf:Agent are recommended as values for this property. See also: § 6.11 Class: Organization/Person Vocabulary: Data Catalog Vocabulary (DCAT) RDF Property: dct:creator	

	Range: foaf:term_Agent Source: dcat:resource_creator
As A [user]	Researcher
I want to	Know the creator
So That	I can understand the context and means for attribution

Investigations	Please provide a link to any active projects that are using the dataset.	urlList //Regular expression (regex) to validate URL: x==~ /https?:\/\/(www\.)?[-a-zA-Z0-9@:%._\+~#=]{1,256}\.[a-zA-Z0-9()]{1,6}\b([-a-zA-Z0-9()@:%_\+.~#?&//=]*)/
Completion	Mandatory	
Cardinality	MinOccurs: 0 Max Occurs: *	

isReferencedBy	Please provide the keystone paper associated with the dataset. Also include a list of known citations, if available, and should be links to existing resources where the dataset has been used or referenced.	urlList //Regular expression (regex) to validate URL: x==~ /https?:\/\/(www\.)?[-a-zA-Z0-9@:%._\+~#=]{1,256}\.[a-zA-Z0-9()]{1,6}\b([-a-zA-Z0-9()@:%_\+.~#?&//=]*)/
Label		
Example	https://www.ipcc.ch/report/ar6/wg1/	
Priority		
Completion	Optional	
Cardinality	Min Occurs: 0 Max Occurs: *	
Definition	A related resource, such as a publication, that references, cites, or otherwise points to the catalogued resource. Usage Note: In relation to the use case of data citation, when the catalogued resource is a dataset, the dct:isReferencedBy property allows to relate the dataset to the resources (such as scholarly publications) that cite or point to the dataset. Multiple dct:isReferencedBy properties can be used to indicate the dataset has been referenced by multiple publications, or other resources. Usage Note: This property is used to associate a resource with the resource (of type dcat:Resource) in question. For other relations to resources not covered with this property, the more generic property dcat:qualifiedRelation can be used. Vocabulary: Data Catalog Vocabulary (DCAT) RDF Property: dct:isReferencedBy Source: dcat:resource_is_referenced_by	
As A [user]	Researcher	

I Want To	Know how to cite the asset
So That	I can appropriately cite the resource in my research publications

References	A related resource that is referenced, cited, or otherwise pointed to by the described resource.	urlList //Regular expression (regex) to validate URL: x==~ /https?:\/\/(www\.)?[-a-zA-Z0-9@:%._\+~#=]{1,256}\.[a-zA-Z0-9()]{1,6}\b([-a-zA-Z0-9()@:%_\+.~#?&//=]*)/
Example	https://doi.org/10.1016/j.gloenvcha.2016.05.009	
Completion	Optional	
Cardinality	Min Occurs: 0 Max Occurs: *	
Definition	URI http://purl.org/dc/terms/references Label References Definition A related resource that is referenced, cited, or otherwise pointed to by the described resource. Comment This property is intended to be used with non-literal values. This property is an inverse property of Is Referenced By. Type of Term Property Subproperty of Relation (http://purl.org/dc/elements/1.1/relation) Relation (http://purl.org/dc/terms/relation)	
As A [user]	Researcher	
I Want To	Cite the resource used in my dataset	
So That	I can provide appropriate recognition	

Access

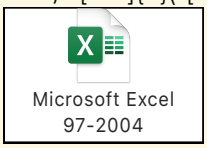
This section includes information about data access.

Access URL	A URL of the resource that gives access to a distribution of the dataset. E.g. landing page, download link, service desk request.	url //Regular expression (regex) to validate URL: x==~ /https?:\/\/(www\.)?[-a-zA-Z0-9@:%._\+~#=]{1,256}\.[a-zA-Z0-9()]{1,6}\b([-a-zA-Z0-9()@:%_\+.~#?&//=]*)/
Default	NA	

Completion	Optional
Cardinality	Min Occurs: 0 Max Occurs: 1
Definition	A URL of the resource that gives access to a distribution of the dataset. E.g. landing page, feed, SPARQL endpoint.
As A [user]	Researcher
I Want To	Know where I can download the data
So That	I can use it for my analysis
Schema.org alternative	https://schema.org/downloadUrl

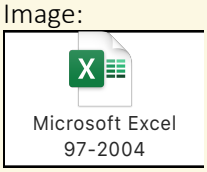
Access Service	Please provide a brief description of the environment where data can be accessed by researchers.
	If no environment is currently available, please indicate the current plans and timelines when the data will be made available through an access environment or process.
		charaters5000
		
		//Regular expression to validate min, max character limit, min 2, max 5000
		X==~ /.{2,5000}/

Completion	Optional	
Cardinality	Min Occurs: 0 Max Occurs: 1	
Definition	A data service that gives access to the distribution of the dataset Vocabulary: Data Catalog Vocabulary (DCAT RDF Property: dcat:accessService Usage Notes: cat:accessService SHOULD be used to link to a description of a dcat:DataService that can provide access to this distribution.	
As A [user]	Researcher	
I Want To	Know how I can access the data	
So That	I can plan and decide accordingly	

Jurisdiction	Please use country code from ISO 3166-1 country codes and the associated ISO 3166-2 for regions, cities, states etc. for the country/state under whose laws the data subjects' data is collected, processed and stored. Multiple Jurisdictions may be provided (if applicable).	isoCountryCodeEnum // value is decimal number x ==~ /^[A-Z]{2}(-[A-Z]{2,3})?$/ 
Completion	Required (If applicable)	
Cardinality	Min Occurs: 0 Max Occurs: *	
Definition	A named and identified geospatial area with defined borders which is used for exercising the action of the Rule. An IRI MUST be used to represent this value. Note: A code value for the area and source of the code must be presented in the Right Operand. Example: the [iso3166] Country Codes or the Getty Thesaurus of Geographic Names. Narrower terms: spatialCoordinates. Vocabulary: Dublin Core Source: http://purl.org/dc/terms/Jurisdiction	
Notes for HOP	Multi-select needed	
As A [user]	Researcher	

I Want To	Know about potential constraints on the data due to jurisdiction
So That	I can tailor my request and research accordingly (or realise the data isn't suitable)

Language	This should list all the languages in which the dataset metadata and underlying data is made available. This refers to the natural language used for textual metadata (i.e. titles, descriptions, etc) of a catalogued resource (i.e. dataset or service) or the textual values of a dataset distribution.	languageEnum 
Examples	en, es, zh, ru, ar, fr	
Completion	Mandatory	
Cardinality	Min Occurs: 1 Max Occurs: *	
Definition	A language of the item. This refers to the natural language used for textual metadata (i.e. titles, descriptions, etc) of a catalogued resource (i.e. dataset or service) or the textual values of a dataset distribution. Usage Note: Resources defined by the Library of Congress (ISO 639-1, ISO 639-2) SHOULD be used. If an ISO 639-1 (two-letter) code is defined for language, then its corresponding IRI SHOULD be used; if no ISO 639-1 code is defined, then IRI corresponding to the ISO 639-2 (three-letter) code SHOULD be used. Repeat this property if the resource is available in multiple languages. If representations of a dataset are available for each language separately, define an instance of dcat:Distribution for each language and describe the specific language of each distribution using dcat:language (i.e. the dataset will have multiple dcat:language values and each distribution will have just one as the value of its dcat:language property). Vocabulary: Data Catalog Vocabulary (DCAT) RDF Property: dct:language Range: dct:LinguisticSystem Source: dcat:language	
As A [user]	Researcher	
I Want To	Know the language of the dataset	
So That	I understand the contents using that language	

Format	The file format of the distribution. If multiple formats are available, please specify. See application, audio, image, message, model, multipart, text, video, https://www.iana.org/assignments/media-types/media-types.xhtml **Note**: If your file format is not included in the current list of formats, you will be directed to a service desk page where you can request your additional format. If not, please go to: https://metadata.atlassian.net/servicedesk/customer/portal/6 to request your format.	charaters80 Application:  Text:  Image:  Video:

	 Audio: 
Example	application/x-netcdf
Completion	Mandatory
Cardinality	Min Occurs: 1 Max Occurs: *
Definition	The file format of the distribution. Usage Note: dcat:mediaType SHOULD be used if the type of the distribution is defined by IANA [IANA-MEDIA-TYPES]. RDF Property: http://purl.org/dc/terms/format Range: https://dublincore.org/specifications/dublin-core/dcmi-terms/#MediaTypeOrExtent Source: https://www.w3.org/TR/vocab-dcat-2/#Property:distribution_format
As A [user]	Researcher
I Want To	Know the format
So That	I Know how to process it

Enrichment & Linkage

This section includes information about related datasets that may have previously been linked, as well as indicating if there is the opportunity to link to other datasets in the future. If a dataset has been enriched and/or derivations, scores and existing tools are available this section allows providers to indicate this to researchers.

Qualifiedrelations	If applicable, please provide the DOI of other datasets that have previously been linked to this dataset and their availability. If no DOI is available, please provide the title of the datasets that can be linked, where possible using the same title of a dataset previously onboarded. The other datasets linked to here might also be referencing or be referenced by the dataset in question in cases where specifying these relations under the respective, separate attributes is not technically possible.	urlList //Regular expression (regex) to validate URL: x ==~ /((([A-Za-z]{3,9}:(?:\/\/)?)(?:[-;:&=\+\$,\w]+@)?[A-Za-z0-9.-]+(:[0-9]+)?|(?:www.|[-;:&=\+\$,\w]+@)[A-Za-z0-9.-]+)((?:\/[\+~%\/.\w-_]*)?\??(?:[-\+=&;%@.\w_]*)#?(?:[\w]*))?)|ALL|NA/
Completion	Optional	
Cardinality	Min Occurs: 0 Max Occurs: *	
Definition	Link to a description of a relationship with another resource. Usage Note: Used to link to another resource where the nature of the relationship is Known but does not match one of the standard [DCTERMS] properties (dct:hasPart, dct:isPartOf, dct:conformsTo, dct:isFormatOf, dct:hasFormat, dct:isVersionOf, dct:hasVersion, dct:replaces, dct:isReplacedBy, dct:references, dct:isReferencedBy, dct:requires,	

	dct:isRequiredBy) or [PROV-O] properties (prov:wasDerivedFrom, prov:wasInfluencedBy, prov:wasQuotedFrom, prov:wasRevisionOf, prov:hadPrimarySource, prov:alternateOf, prov:specializationOf). Vocabulary: Data Catalog Vocabulary (DCAT) RDF Property: dcat:qualifiedRelation Sub-property of: prov:qualifiedInfluence Domain: dcat:Resource Range: dcat:Class:Relationship Source: dcat:resource_qualified_relation
As A [user]	Researcher
I Want To	Know if there are any datasets that have already been linked
So That	I can use more pre-processed data in my research

Tools	Please provide the URL of any analysis tools or models that have been created for this dataset and are available for further use. Multiple tools may be provided.	urlList //Regular expression (regex) to validate URL: x==~ /https?:\/\/(www\.)?[-a-zA-Z0-9@:%._\+~#=]{1,256}\.[a-zA-Z0-9()]{1,6}\b([-a-zA-Z0-9()@:%_\+.~#?&//=]*)/
Example	https://www.esmvaltool.org/	
Completion	Optional	
Cardinality	Min Occurs: 0 Max Occurs: *	
As A [user]	Researcher	
I Want To	Know whether there are any available tools created for this dataset	
So That	I can consider using it for my own analysis	

Structural Metadata

Data Class	Name of the table in the dataset. Use a fully qualified name if appropriate.	xs:string	Min Occurs: 1 Max Occurs: 1
Data Class Description	Description of the table in the dataset. Use a fully qualified name if appropriate.	xs:string	Min Occurs: 1 Max Occurs: 1
Data Element	Name of the column/variable in the table.	xs:string	Min Occurs: 1 Max Occurs: 1
Data Element Description	Description of the column/variable in the table.	xs:string	Min Occurs: 1 Max Occurs: 1
Data Type	Type of data contained in the column/variable.	xs:string	Min Occurs: 1 Max Occurs: 1
Measurement Unit	Unit of measure	xs:string	Min Occurs: 0 Max Occurs: 1