

Guide to the inventory schema

1. Inventory structure and terminology

An **inventory** is a catalogue of datasets held by an organisation. It can include both published and unpublished datasets. It contains references to:

datasets, which are sets of records however structured – not just datasets according to standard database terminology. Datasets contain:

data resources, which are files or feeds/streams of data records available in zero, one or more:

renditions, which are different formats of the data, such as: CSV, XML and PDF. Renditions may represent physical files or API calls.

document resources, which are files or web pages describing datasets. They may include an API page for a dataset and documents describing a dataset (eg rules for data inclusion, how data was redacted). Document resources are expressed as one or more:

renditions, which are different formats of the data, such as: ODF, PDF and HTML.

2. Inventory format

Each inventory is expressed in an XML format, with elements of the inventory structure nested within one another. The XML format is given by the [inventory.xsd schema](#).

2.1 Inventory metadata

Metadata, expressed as properties via XML elements for each level of the inventory is defined below. Properties will be defined using [Dublin Core Terms](#) wherever relevant terms exist.

The DataShare API provides an inventory feed in the XML structure.

Data.gov.uk (DGU) converts from terms used to equivalents in the CKAN structure. ([CKAN](#) is the open source data portal software used by DGU.) DGU does not distinguish resources and resource renditions, so these bottom two levels of the inventory structure are flattened into one in DGU. DGU sets its 'published/unpublished' flag for each dataset according to the presence/absence of a data resource rendition.

In each of the following sections metadata elements (i.e. 'terms') are assigned a cardinality as follows:

- 1 – mandatory that there is one
- 1..∞ - mandatory that there is at least one, but there may be more than one
- 0..1 – optional, but there can be no more than one

- 0.. ∞ - optional and there can be more than one

Elements not flagged as “YES” in the “Used by DGU” column will be carried as cargo. They may also be used to derive other elements.

2.2 Inventory metadata elements

Metadata describing the whole inventory:

Element / DC Term	Cardinality	Notes	Used by DGU
conformsTo (xmlns...)	1	URL of the XSD schema defining the structure of the inventory as a whole. This element is expressed by the standard xmlns notation in XML. The schema may change over time, so this element is mandatory so harvesters know the format to expect.	YES
modified	1	Date (or date and time) when the inventory was last modified. This complies with ISO 8601 . A date of format yyyy-mm-dd is normally sufficient	YES. Used by DGU to see if the inventory has changed since it was last read.
identifier	1	Identifier which the publisher can refer to. Must be unique within publisher. May contain publisher’s revision control.	
publisher	1	URI of the council or other body responsible for the inventory. Local authorities in England will use URLs from OpenDataCommunities (eg http://opendatacommunities.org/doc/unitary-authority/peterborough).	DGU looks up the organisation name and the default spatial coverage from the area it ‘governs GSS’ in OpenDataCommunities. DGU will cache such information and refresh the cache after changes to local government organisations and areas.
title	0..1	Title of the inventory	
description	0..1	Free text description of the inventory	
coverage.spatial	0..1	The geographical area described by the	YES

		<p>information. It defaults to the area of publisher if not specified.</p> <p>If it is specified, it should be a URI from statistics.data.gov.uk such as http://statistics.data.gov.uk/id/statistical-geography/E06000031 for the Peterborough area.</p> <p>This area will, by default, apply to datasets within the inventory, but it can be overridden for individual datasets.</p>	
language	0..1	Language code from ISO 639-1 . Assumed to be “en” if not specified. The code for Welsh is “cy”.	
creator	0..1	URI or text giving the organisation publishing the inventory. Assumed to be the creator if not specified. The publisher should normally take responsibility for creation so this element should be used with caution.	

2.3 Dataset metadata elements

Metadata describing each inventory item. Those items which have renditions will be considered to represent inventory items for *published* datasets.

DC Term	Cardinality	Notes	Used by DGU
identifier	1	Identifier unique within the inventory publisher organisation that uniquely identifies the dataset so a harvester knows whether to add it or update an existing dataset.	YES
title	1	Title of the dataset	YES
modified	0..1	<p>Date (or date and time) in ISO 8601 format when the dataset was last modified. This will not be later than the modified date of the whole inventory. It applies to both the data and its metadata, so changes to either will cause a change to the value of this element.</p> <p>Used to see if the dataset has changed since it was last read. If no value is present, the modified date/time of the inventory is assumed to apply. This cannot be later than the modified date of the inventory.</p>	YES
active	0..1	Defaults to ‘Yes’. If this is set to ‘No’ it denotes that the dataset is discontinued or withdrawn.	YES

description	0..1	Free text description of the dataset. This is not required but is strongly recommended.	YES
subject.service	0..∞	URI of from http://id.esd.org.uk/list/services of the service to which the data relates. This is strongly recommended where a dataset pertains to a service.	YES
subject.function	0..∞	URI from http://id.esd.org.uk/list/functions of the function to which the data relates. This is strongly recommended where a dataset pertains to a function of (local) government.	YES
subject.scheme	0..∞	<p>URI of descriptor (eg demographic category) from another commonly used controlled list that covers topics other than local government services and functions.</p> <p><i>Scheme</i> is an attribute of the subject XML element (which may have other attributes). Schemes of 'service' and 'function' must not be used for URIs from lists other than the above two lists.</p> <p>The circumstance list may be relevant to some datasets.</p> <p>The ICO model publication scheme class, where the dataset represents a publication scheme item from this ICO document may be used where the <i>scheme</i> is publicationScheme.</p> <p>Where there is no scheme, pure text can be used (eg for the locally defined categories within each DataShare instance)</p>	DGU adds scheme and the value as key / value pairs available for API queries.
protected This is not a DC Term and there is no obvious equivalent	0..1	<p>Present if it is the publisher's view that the dataset is subject to protection, in which case one of:</p> <ul style="list-style-type: none"> • protected personal • protected Fol exempt • protected intellectual property • charge applies • other <p>The above list is subject to consultation.</p> <p>Where the dataset represents records described by ESD's service records retention URI, the 'impact level' given by that URI might denote the type of protection applicable.</p>	YES
protectedText	0..1	Text explanation of the 'protected' category	YES

		defined above	
context This is not a DC Term and there is no obvious equivalent	0..1	<p>One of:</p> <ul style="list-style-type: none"> ● reference ● operational ● aggregated ● strategic <p>This indicates the level to which data has been aggregated and analysed before publishing of the dataset.</p> <p>Headings have these meanings:</p> <ul style="list-style-type: none"> ● Reference – data that describes the existence of things, and what we know about them ● Operational – data that describes cases, activities, transactions and so on ● Aggregated – data that has been derived such as statistics and analytics ● Strategic – data that describes decisions, allocation of resources, targets, and so on <p>Aggregated and analytical categorisations are important because they represent conclusions made, which might be open to challenge.</p>	
rights	0..1	<p>URI of licence or URL to a web page describing the licence. (DGU uses the web page URL of the Open Government Licence where that licence applies).</p> <p>This is required if the dataset published.</p>	YES
rightsHolder	0..1	Text name of an organisation (if any) that holds the rights.	
chargeURL	0..1	URL of page describing any charges applicable to accessing the dataset.	
relation	0..∞	Identifier of another dataset from to which this dataset is related in a way given by the relation.type element.	YES
relation.type	0..∞	<p>One of these are DGU package relationships:</p> <ul style="list-style-type: none"> ● depends on ● derived from ● links to <p>Pairs of relation and relation.type are used for each related dataset.</p>	

commitmentDate	0..1	For unpublished datasets, this element shows any date by which there is a public commitment to release the dataset in ISO 8601 format.	YES
----------------	------	--	-----

2.47.3 Data resource and document resource metadata elements

Metadata describing each resource is given below. The same elements apply to both data resources and document resources, although typically fewer elements will be populated for a document resource. For example, rules for redaction are likely to be independent of temporal or spatial coverage.

Those items which have renditions will be considered to represent inventory items for *published* datasets.

DC Term	Cardinality	Notes	Used by DGU
identifier	1	Identifier unique within the inventory publisher organisation that uniquely identifies the dataset so a harvester knows whether to add it or update an existing dataset.	YES
type	1	Flag to denote if the resource represents: <ul style="list-style-type: none"> • data OR • documentation on the dataset 	YES
title	1	Title of the resource	YES
modified	0..1	Date (or date and time) in ISO 8601 format when the resource was last modified. This will not be later than the modified date of the whole inventory. It applies to both the data and its metadata, so changes to either will cause a change to the value of this element. Used to see if the resource has changed since it was last read. If no value is present, the modified date/time of the dataset is assumed to apply. This cannot be later than the modified date of the dataset.	YES
active	0..1	Defaults to 'Yes'. If this is set to 'No' it denotes that resource is discontinued or withdrawn.	YES
description	0..1	Free text description of the data resource. This is not required but is strongly recommended.	YES
coverage.spatial	0..∞	URIs of geographical areas (normally from statistics.data.gov.uk or neighbourhoods.esd.org.uk/ to which the data pertains. Assumed to be the spatial coverage of the inventory as a whole if not	YES, DGU will use the inventory's spatial coverage where one is not

		specified.	specified for the dataset. DGU will look up the geographical boundaries from the URIs
coverage.temporal.from	0.1	Starting date/time of the dataset rendition in ISO 8601 format. This 'from' date is used alone if the dataset pertains to a single point in time	YES
coverage.temporal.to	0.1	End date/time of the dataset rendition in ISO 8601 format.	YES
language	0..1	Language code from ISO 639-1 . Assumed to be "en" if not specified. The code for Welsh is "cy"	

2.5 Data and document resource rendition metadata elements

Unless otherwise stated, the elements given below apply to renditions of both datasets and documents.

DC Term or other property name	Cardinality	Notes	DGU
identifier	1	URL of the rendition	YES. DGU calls this 'URL'
contentType	0..1	A standard MIME type (text/csv, text/xml, application/json, etc.) for the format. See the template column here for a list of types. This corresponds to the DC term format.	YES
modified	0..1	Date (or date and time) in ISO 8601 format when the resource was last modified. This will not be later than the modified date of the whole inventory. It applies to both the data and its metadata, so changes to either will cause a change to the value of this element. Used to see if the rendition has changed since it was last read. If no value is present, the modified date/time of the resource is assumed to apply. This cannot be later than the modified date of the resource.	YES
active	0..1	Defaults to 'Yes'. If this is set to 'No' it denotes that the rendition is discontinued or withdrawn.	YES
title	0..1	Recommended file name for the resource	YES. This is used to generate the

			DGU name for the resource, which may be modified to make it unique across DGU
description	0..1	Description of the resource. DGU has confirmed that both the dataset and the resource can have descriptions that are displayed online.	YES
conformsTo	0..∞	URL for the machine readable schema / definition that gives the format of the dataset rendition. Normally this element only applies to dataset resource renditions.	YES
availability This is not a DC Term and there is no obvious equivalent	0..1	One of <ul style="list-style-type: none"> ● Download ● Service ● On application Normally this element only applies to dataset resource renditions.	YES