The FAIRness Project

Building Trust and FAIRness into

the Process for Finding and Using Government Data

Chief Data Officers Council and Federal Committee on Statistical Methodology

Project Co-Chairs

Thomas Dabolt

Thomas_dabolt@ios.doi.gov

Michael Ratcliffe

michael.r.ratcliffe@census.gov

federal CDO council

FCSM Federal Committee on Statistical Methodology

FAIRness Project Goal

- Advance efforts to improve and enhance data users' (internal and external to the federal government) ability to find and assess the utility of federal data.
  - Aligned with the Evidence Act requirement for a comprehensive federal data catalog that supports implementing the new data access authorities and responsibilities under Title III of the Evidence Act.
  - Builds from FAIR (Findable, Accessible, Interoperable, and Reusable/Reproducible) principles.
- Project info: https://github.com/DOI-DO/dcat-us/wiki

DCAT-US 3 Aligning with Standards

Most changes from the current DCAT-US 1.1 "POD" schema are to align with the latest DCAT specifications developed by the W3C international standards body.

Highlighted Changes:

- Controlled vocabularies
- New and aligned classes and properties
- Addition of geospatial classes and properties

Some changes are specific to the draft DCAT3-US profile and driven by requirements and issues submitted by stakeholders and project contributors: Issues · DOI-DO/dcat-us · GitHub

- Current draft DCAT-US 3 specification: https://doi-do.github.io/dcat-us/
- How to contribute: https://github.com/DOI-DO/dcat-us/wiki#contribute

Changes from DCAT-US 1.1

Change Type	Description
New!	New DCAT-US 3.0 specific class that is not referred in DCAT specifications
Aligned	Class introduced in DCAT specifications that does not exist in current DCAT-US 1.1 ("POD") schema

Core Class- Changes from DCAT-US 1.1 (POD) to DCAT-US 3

Core Class name	Changes from DCAT-US 1.1	Rationale
Catalog	*Aligned*	Consistency in metadata representation
CatalogRecord	*Aligned*	Consistency in metadata representation
Dataset	*Aligned*	Align with international DCAT standards
Distribution	*Aligned*	Comprehensive data compatibility
DataService	*Aligned*	Improves discoverability and reduce confusion with dataset distributions
DatasetSeries	*Aligned*	Improves discoverability

Supporting Class- Changes from DCAT-US 1.1 (POD) to DCAT-US 3

Supporting Class name	Changes from DCAT-US 1.1	Rationale
AccessRestriction	*New!*	Ensure responsible access to records
Activity	*Aligned*	Simplify representation of activities; accommodate evolution of activities
Address (Location)	*New!*	Use "locn:Address" (from W3C Location ontology) as standardized, structured, and extensible format for physical addresses; improves “FAIRness” with consistent, interoperable, and precise sharing of location information.
Address (Contact Point)	*Aligned*	Integrate VCard contact point address ensures a standardized, interoperable format for presenting address data
Agent	*Aligned*	Improved automation, findability and interoperability (e.g., consistency for person and organization subclasses)
Attribution	*Aligned*	Data provenance best practices
Concept	*Aligned*	Aligns with international standards; recognizes controlled vocabularies
ConceptScheme	*Aligned*	Aligns with international standards; recognizes controlled vocabularies
Checksum	*Aligned*	Promotes data integrity
Contact	*Aligned*	Improve accessibility of data resources
CUIRestriction	*New!*	Compliance with NARA guidelines; interoperability
Document	*New!*	Aligns with international standards; interoperability
GeographicBoundingBox	*New!*	Remove ambiguity and improve findability and interoperability

Supporting Class- Changes from DCAT-US 1.1 (POD) to DCAT-US 3

Supporting Class name	Changes from DCAT-US 1.1	Rationale
Identifier	*New!*	Data inventory, governance, and FAIRness
LiabilityStatement	*New!*	Legal compliance
LicenseDocument	*Aligned*	Align with international standards; responsible data sharing
Location	*Aligned*	Align with geospatial standards
MediaType	*Aligned*	Data interoperability and data sharing
Metric	*Aligned*	Align to international data quality standards
Organization	*Aligned*	Improve data interoperability and discoverability
PeriodOfTime	*Aligned*	Remove inconsistencies and align with international standards
Person	*Aligned*	Aligns with best practices in data representation
Program	*New!*	Link datasets to agency programs for enriched data networks and enhanced "FAIRness"
ProvenanceStatement	*New!*	Align with international data quality and transparency standards
QualityMeasurement	*Aligned*	Improve data usability; aligns with international standards

Supporting Class- Changes from DCAT-US 1.1 (POD) to DCAT-US 3

Supporting Class name	Changes from DCAT-US 1.1	Rationale
Relationship	*Aligned*	Improves flexibility in documenting dataset relationships and discovery
RightsStatement	*Aligned*	Enhance dataset transparency
Role	*Aligned*	Enriches data networks with precise, navigable, and semantically transparent relationships among datasets, enhancing "FAIRness"
Standard	*Aligned*	Align with standards
UseRestriction	*New!*	Align with NARA guidelines; improve compliance and interoperability

Requirement Levels for DCAT-US 3 Classes

DCAT-US defines requirement levels for data receivers and senders:

Mandatory property: a receiver **MUST** be able to process the information for that property; a sender **MUST** provide the information for that property.

Recommended property: a receiver ***MUST*** be able to process the information for that property; a sender ***SHOULD*** provide the information for that property if it is available (Conditionally Mandatory).

Optional property: a receiver **MUST** be able to process the information for that property; a sender **MAY** provide the information for that property but is not obliged to do so.

Deprecated property: a receiver **SHOULD** be able to process information about instances of that property; a sender **SHOULD NOT** provide the information about instances of that property.

**More details on the DCAT-US 3 Spec*

Properties with Controlled Vocabularies

Property URI	Used for Class	Vocabulary URI	Usage note
adms:representationTechnique	Distribution	http://resources.data.gov/vocab/SpatialRepresentationType**	DCAT-US extension.
adms:status	Distribution	http://purl.org/adms/status/	The list of terms in the ADMS status vocabulary is included in the [VOCAB-ADMS] specification
+dcat:hadRole	Attribution	http://resources.data.gov/vocab/ResponsiblePartyRole**	DCAT-US extension.
dcat:mediaType	Distribution	http://www.iana.org/assignments/media-types	
dcat:theme	Dataset Catalog Data Service	https://resources.data.gov/vocab/data-theme** https://resources.data.gov/vocab/ngda-data-theme**	The values to be used for this property are the URIs of the concepts in the vocabulary.
dcat:themeTaxonomy	Catalog	https://resources.data.gov/vocab/data-theme**	The value to be used for this property is the URI of the vocabulary itself, i.e. the concept scheme, not the URIs of the concepts in the vocabulary.
dcat-us:availability	Distribution	https://resources.data.gov/vocab/availability/**	The list of terms for the avalability levels of a dataset distribution in the DCAT-US specification.
dct:accessRights	Data Service Dataset Catalog Distribution	https://resources.data.gov/vocab/access-right**	
dct:accrualPeriodicity	Dataset Dataset Series	http://purl.org/cld/freq/	

**Vocabularies to be formalized in next phase of profile

Properties with Controlled Vocabularies

Property URI	Used for Class	Vocabulary URI	Usage note
dct:conformsTo	Data Service Dataset Distribution Data Service	http://www.opengis.net/def/crs/EPSG/0/ https://resources.data.gov/vocab/ProtocolValue**	
dct:format	Distribution	https://resources.data.gov/vocab/file-type**	
dct:language	Catalog Catalog Record Dataset Distribution	https://id.loc.gov/vocabulary/iso639-1/	
dct:publisher	Catalog Dataset	https://resources.data.gov/vocab/corporate-body	The Corporate bodies must be used for US institutions and a small set of international organisations. In case of other types of organisations, national, regional or local vocabularies should be used.
dct:spatial	Catalog Dataset Data Service	https://resources.data.gov/resource/continent** https://resources.data.gov/resource/country** https://resources.data.gov/resource/place ** http://sws.geonames.org/	The US Vocabularies Name Authority Lists must be used for continents, countries and places that are in those lists; if a particular location is not in one of the mentioned Named Authority Lists, [GEONAMES] URIs must be used.
sdmx-attribute:unitMeasure	Quality Measurement	http://www.qudt.org/vocab/unit	DCAT-US extension.

**Vocabularies to be formalized in next phase of profile

Geospatial Metadata

DCAT-US 3 specification provides a standardized way to represent metadata about datasets and services, including information about their spatial properties not represented in DCAT-US 1.1.

New in DCAT-US 3

Class:

Geographic Bounding Box

Properties on dcat:Dataset:

Geographic Bounding Box

Spatial

Spatial Resolution in meters

Geographic Bounding Box

Class dcat-us:GeographicBoundingBox

Property	Range	ReqLevel
west bounding longitude	xsd:decimal	**M**
east bounding longitude	xsd:decimal	**M**
south bouding latitude	xsd:decimal	**M**
north bounding latitude	xsd:decimal	**M**

RDF Class: dcat-us:GeographicBoundingBox

Definition: GeographicBoundingBox describes the spatial extent of domain of application of an resource and is standardized in WGS 84 Lat/Long coordinate system.

Rationale Geographic Bounding box is an important construct enabling efficient indexing and search. There is no consensus and common vocabulary to describe spatial bounding box in the community. GML Envelope was proposed but it is too cumbersome to process. For simplicity and interoperability, use four separate fields for each bound (west, east, north and south) that removes any ambiguity and improves findability.

Geographic Bounding Box

Property of dcat:Dataset

Property	URI	Range	ReqLevel	Card	Changes from DCAT-US 1.1
geographic bounding box	dcat-us:geographicBoundingBox	dcat-us:GeographicBoundingBox	R	0..n	New!

Property geographic bounding box

Requirement level Recommended

Cardinality 0..n

URI dcat-us:geographicBoundingBox

Range dcat-us:GeographicBoundingBox

Spatial

Property of dcat:Dataset

Property	Range	ReqLevel	Changes from DCAT-US 1.1
spatial/geographical coverage	dct:Location	R	Fixed

Property	spatial/geographical coverage
Requirement level	Recommended
Cardinality	0..n
URI	dct:spatial
Range	dct:Location

Spatial Resolution in Meters

Property of dcat:Dataset

Property	Range	ReqLevel	Changes from DCAT-US 1.1
spatial resolution in meters	rdfs:Literal (typed as xsd:decimal)	O	Aligned

Property	spatial resolution in meters
Requirement level	Optional
Cardinality	0..n
URI	dcat:spatialResolutionInMeters
Range	rdfs:Literal (typed as xsd:decimal)
Usage Note	•If the dataset is an image or grid this should correspond to the spacing of "pixels" or grid cells. For other kinds of spatial datasets, this property will usually indicate the smallest distance between items in the dataset. •The range of this property is a decimal number representing a length in meters. Provides a summary indication of the spatial resolution of the data as a single number. More complex descriptions of various aspects of spatial precision, accuracy, resolution and other statistics can be provided using the Data Quality Vocabulary [VOCAB-DQV].

Core Classes with Mandatory Properties

Catalog
CatalogRecord
Dataset
Distribution
DatasetSeries*
DataService*

Note: These classes have additional properties that are recommended/ conditionally mandatory or optional but whose use is encouraged for improved “FAIRness”.

**Aligned for better representation of datasets, services and groupings of related datasets*

Supporting Classes with Mandatory Properties

AccessRestriction*
Agent
Attribution*
CheckSum
Concept
ConceptScheme
Contact
CUIRestriction*
Document*
GeographicBoundingBox*
Location*
Metric
Organization
Person
QualityMeasurement
Relationship
UseRestriction*

**New to DCAT-US 3*

Note: These classes have additional properties that are recommended/ conditionally mandatory or optional but whose use is encouraged for improved "FAIRness".

Additional Supporting Classes

Supporting classes with recommended/conditionally mandatory or optional properties

Identifier*
LiabilityStatement*
LicenseDocument
MediaType
PeriodOfTime
ProvenanceStatement*
RightsStatement
Standard

Activity
Address (Contact Point)
Address (Location)*
Investment*
Program*
Role

**New to DCAT-US 3*

Note: These classes have additional properties that are recommended/ conditionally mandatory or optional but whose use is encouraged for improved "FAIRness".

Comments and feedback on the DCAT-US 3 profile are welcome and appreciated!

Current draft DCAT-US 3 specification:
https://doi-do.github.io/dcat-us/

How to contribute:
https://github.com/DOI-DO/dcat-us/wiki#contribute