

The FAIRness Project

Building Trust and FAIRness into 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



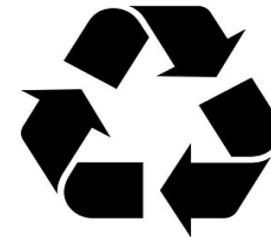
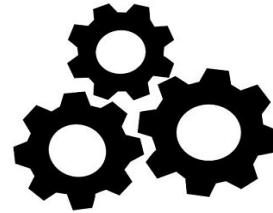
Federal Committee on
Statistical Methodology

FAIRness Project Purpose and Basic Information

- Improve users' experience (internal and external to the federal government) to find, access, and use federal data.
 - The Evidence Act requires a comprehensive federal data catalog that supports implementing the new data access authorities and responsibilities under Title III of the Evidence Act.
 - Decreases burden and increases interoperability across:
 - Data.gov
 - Geoplatform.gov,
 - The Statistical Access Portal,
 - Agencies Enterprise Data Inventories
 - Supports the Zero Trust Data Pillar and Open Science
 - Builds on FAIR principles (Findable, Accessible, Interoperable, and Reusable)

Q: What is “FAIRness”? A: Federal data to citizens are improved when they are consistently...

F_{indable} A_{ccessible} I_{nteroperable} R_{eusable}



FAIR enables Search, Discovery, Security and Sharing

How?

Consistently Document your Datasets!

FAIRness Project Status

- Project info: <https://github.com/DOI-DO/dcat-us/wiki>
- 118 Requirements submitted by interested stakeholders: <https://github.com/DOI-DO/dcat-us/labels/Requirements%20Request>
- Draft specification for review: <https://doi-do.github.io/dcat-us>
- Comment period ends **17 November 2023**



What does FAIRness solve?

- **User Centric:** FAIRness is all about making life easier for the user; search for what they need, find it, and use it!
- **Consistent Visual Brand and Identity:** There are 3 inconsistent federal data catalogs.
- **Authoritative, Easy-to-understand Content:** Content is often duplicated across the 3 sites. It's unclear which site authoritative for any subject.
- **Search Optimization:** There is not a unified effort across the 3 catalogs to optimize search. Standards and consistency are an essential building block for this.
- **Standardization:** Make it easier for data asset stewards to produce consistent metadata using COTS software

DCAT3 U.S.: What's important? Why?

Outcomes

- Improved and consistent user experience
- Increase automation
- Enhance data sharing, reuse and security
- Fill significant legal, regulatory, and policy, gaps for documenting data
- Decrease maintenance costs

How

- Optimize finding and sharing data by standardizing metadata production and use
- Improve FAIR maturity of data assets
- Support for Dataset Series simplifies metadata handling for large collections of datasets
 - Time series, map series
- Clearly mark restrictions: license, access, use, and CUI
- Leverage controlled vocabularies as reference data to eliminate ambiguity and standardize terminology and use
- Streamline and unify catalogs to reduce duplication
- Consistent, open, and trusted discovery,
- Data and service assets are documented consistently across the USG

Bottom Line

- DCAT3-US is not a “new” standard; it is a “profile” (a schema) of the [World Wide Web Consortium’s \(W3C\) DCAT3 standard](#) metadata vocabulary.
- DCAT3-US is compatible with existing DCAT v1.1 metadata.
 - Few new metadata elements are added while some elements are aligned or fixed;
 - No major changes to existing elements.
- To improve search, DCAT3-US supports new and updated controlled vocabularies, for consistently naming items like federal agencies, file formats, and units of measure.
- DCAT3-US overcomes the limitations of DCAT v1.1 when documenting geospatial data, eliminating the need for a separate federal standard.
 - e.g., CSDGM and ISO 19115.
- DCAT3-US follows a proven approach (e.g., European DCAT-AP); **RESULT**: vendor support is already in place, with additional support underway.

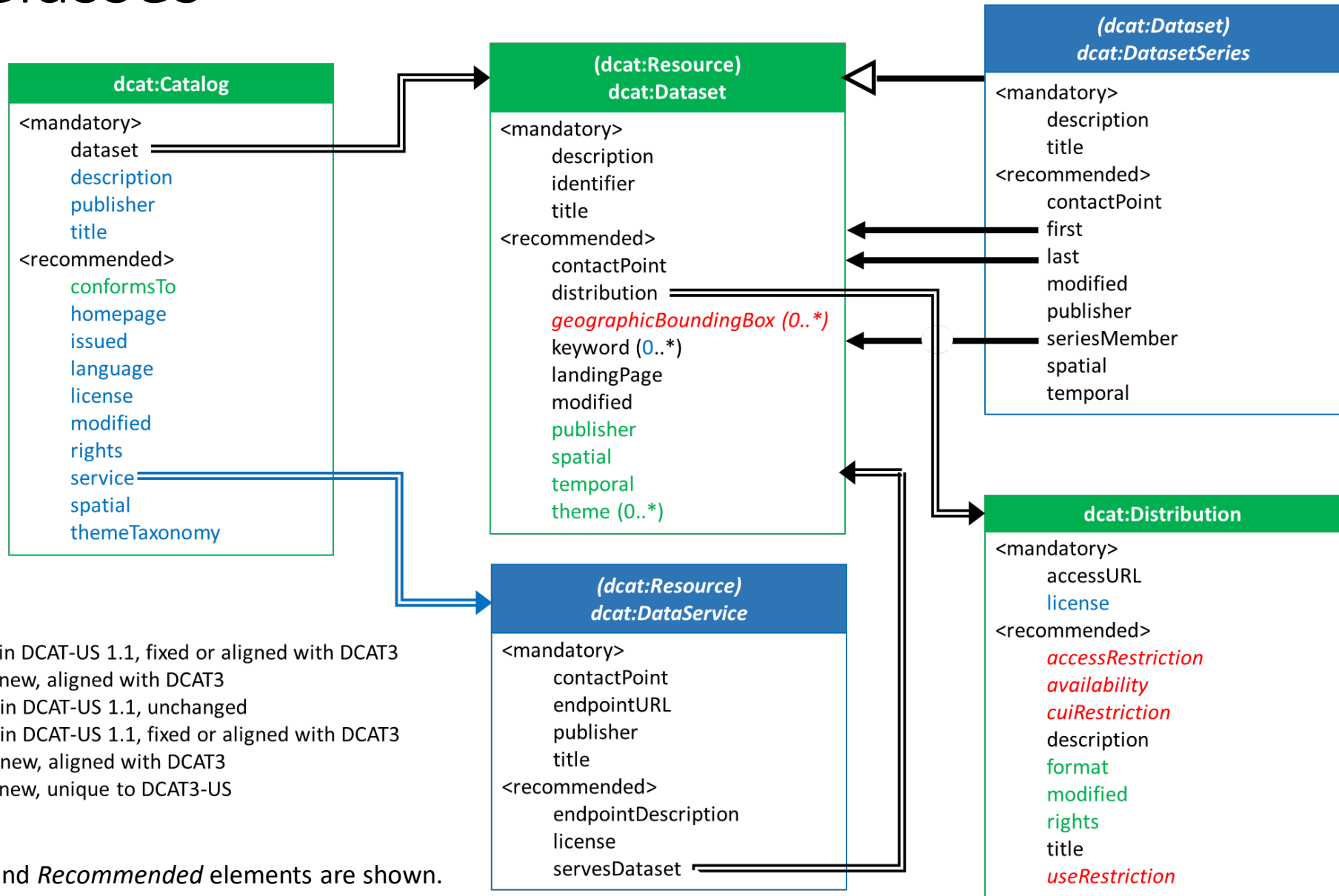
What's changed? What's the same?

Let's take a look under the hood.

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

Core Class Name	Changes from DCAT-US 1.1	Rationale
<u>Catalog</u>	<i>Aligned to DCAT3</i>	Consistency in metadata representation
<u>CatalogRecord</u>	<i>New, Aligned to DCAT3</i>	Consistency in metadata representation
<u>Dataset</u>	<i>Aligned to DCAT3</i>	Align with international DCAT standards
<u>Distribution</u>	<i>Aligned to DCAT3</i>	Comprehensive data compatibility
<u>DataService</u>	<i>New, Aligned to DCAT3</i>	Improves discoverability and reduce confusion with dataset distributions
<u>DatasetSeries</u>	<i>New, Aligned to DCAT3</i>	Improves discoverability

Core Classes

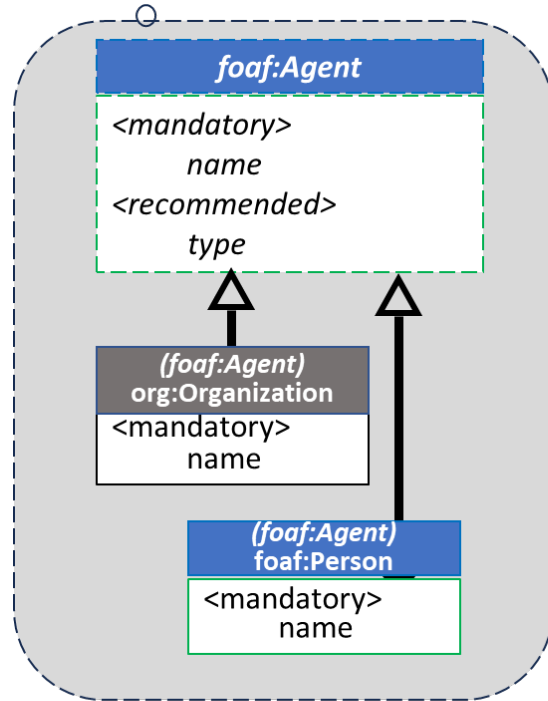


Note: Only *Mandatory* and *Recommended* elements are shown.
 No *Optional* or *Deprecated*. More details on the [DCAT3-US Spec](#)

Supporting Classes – Mandatory and Recommended

Supporting Class Name	Used by DCAT3 Property	Changes from DCAT-US 1.1	Rationale
dcat-us:AccessRestriction	accessRestriction	New	Align with standards
dcat-us:CuiRestriction	cuiRestriction	New	Compliance with NARA guidelines; interoperability
dcat-us:GeographicBoundingBox	geographicBoundingBox	New	Remove ambiguity and improve findability and interoperability
dcat-us:UseRestriction	useRestriction	New	Align with NARA guidelines; improve compliance and interoperability
dct:LicenseDocument	license	Aligned	Aligns with international standards; interoperability
dct:Location	spatial	Aligned	Remove ambiguity and improve findability and interoperability
dct:MediaType	format	Aligned	Align with standards
dct:PeriodOfTime	temporal	Aligned	Remove ambiguity and improve findability and interoperability
dct:RightsStatement	rights	Aligned	Enhance dataset transparency
dct:Standard	conformsTo	Aligned	Align with standards
foaf:Agent	publisher	Aligned	Improved automation, findability and interoperability (e.g., consistency for person and organization subclasses)
foaf:Document	homepage	Aligned	Aligns with international standards; interoperability
skos:Concept	availability theme	Aligned	Aligns with international standards; recognizes controlled vocabularies
vcard:Contact	contactPoint	Aligned	Improve accessibility of data resources

Supporting Classes - Mandatory and Recommended



foaf:Document

<mandatory>
 creator
 issued
 publisher
 publisherName
 title

<recommended>
 bibliographicCitation
 description
 Identifier
 type

dct:Location

<mandatory>
 geographicName

<recommended>
 bbox (GML or WKT string)

dcat-us:GeographicBoundingBox

<mandatory>
 westBoundingLongitude
 eastBoundingLongitude
 southBoundingLatitude
 northBoundingLatitude

dct:MediaType

<recommended>
 label

dcat-us:AccessRestriction

<recommended>
 restrictionStatus
 specificRestriction

dct:PeriodOfTime

<recommended>
 endDate
 startDate

dct:RightsStatement

<recommended>
 label
 attributionText

dcat-us:UseRestriction

<recommended>
 restrictionStatus
 specificRestriction

dct:Standard

<recommended>
 description
 identifier
 issued
 title
 type
 version

skos:Concept

<mandatory>
 inScheme
 prefLabel

<recommended>
 definition

vcard:Contact

<mandatory>
 fn
 hasEmail

dcat-us:CuiRestriction

<mandatory>
 cuiBannerMarking
 designationIndicator

Key

- class = in DCAT-US 1.1, unchanged
- class = in DCAT-US 1.1, fixed or aligned with DCAT3
- class = new, aligned with DCAT3
- class = new, unique to DCAT3-US

Note: Only *Mandatory* and *Recommended* elements are shown.
 No *Optional* or *Deprecated*. More details on the [DCAT3-US Spec](#)



Evidence Act Requirements

- Comply with the Evidence Act - Data Inventory and Federal Data Catalog requirements.// 44 USC 3511(a)(2) - Guidance.--The Director shall establish guidance for agencies to develop and maintain comprehensive data inventories under paragraph (1). Such guidance shall include the following: (A) A requirement for the head of an agency to include in the comprehensive data inventory metadata on each data asset of the agency, including, to the maximum extent practicable, the following...
 - <https://www.law.cornell.edu/uscode/text/44/3511>
- Three DCAT-US 3.0 requirements submitted
 - [Guidance on rights and access to data #40](#)
 - [Comply with the Evidence Act Data Inventory and Federal Catalog requirements. #118](#)
 - [Indicate when an agency has not made an open government data asset determination #119](#)

DCAT-US 3.0 for Evidence Act

Evidence Act Requirement	DCAT-US 3.0 Elements
Description	Dataset.description (description), Dataset.describedBy (data dictionary), Distribution.describedBy (data dictionary)
Name/ title	Dataset.title (title)
Indication of Open Govt data asset determination	Distribution.accessRestriction.restrictionStatus (dcat-us:AccessRestriction) * (range defined in controlled vocabulary; see NARA Access Restriction Status)
Indication of disclosure (full, partial, exempt)	Distribution.accessRestriction.restrictionStatus (dcat-us:AccessRestriction) * (range defined in controlled vocabulary; see NARA Access Restriction Status)
Indication of open data determination	Distribution.license (license), Distribution.mediaType (media type), Distribution.format (format), Distribution.conformsTo (conforms to), Distribution.accessRestriction.restrictionStatus (dcat-us:AccessRestriction)
Indication whether data asset determination has been made	Distribution.accessRestriction.restrictionStatus (dcat-us:AccessRestriction) * (range defined in controlled vocabulary; see NARA Access Restriction Status)
Determination under Section 3582 (CIPSEA)	Distribution.accessRestriction.restrictionStatus (dcat-us:AccessRestriction) * (range defined in controlled vocabulary; see NARA Access Restriction Status) Distribution.accessURL (access URL), Distribution.page (foaf:Document)
Description of the method for public access	Dataset.distribution (dataset distribution)
Date of most recent update	Dataset.issued (release date), Distribution.issued (release date), Dataset.modified ()
Agencies responsible for <i>maintaining</i> the data asset	Dataset.publisher (publisher) ** (by convention `publisher` is used as the entity responsible for a data asset; assumes publisher URI is unique/ unambiguous so POD `bureauCode` is deprecated)
Data asset owner	Dataset.creator (creator) *** (unclear what "owner" means; 'creator' is not necessarily the "owner"; DCAT uses controlled vocabulary for more precise definition of roles; see ISO CI RoleCode including "custodian", "owner", "originator", "sponsor", etc) Dataset.qualifiedAttribution (qualified attribution)
Use restrictions	Distribution.license (license), Distribution.useRestriction (usage restriction), Distribution.cuiRestriction ()
Location of the data asset	Distribution.accessURL (access URL), Distribution.downloadURL (download URL)

Example

- **Source DCAT-US 1.1 (POD) JSON:** <https://catalog.data.gov/harvest/object/1eb394e5-e255-4048-8e62-bde0ae074e01>
- **DCAT-US 3.0 (JSON-LD):** <https://raw.githubusercontent.com/DOI-DO/dcat-us/main/docs/examples/example1-dcat-us-3.0.json>
- **DCAT-US 3.0 (TTL):** <https://raw.githubusercontent.com/DOI-DO/dcat-us/main/docs/examples/example1-dcat-us-3.0.ttl>

Dataset

DCAT-US 1.1 (“POD”)

```
{  
  "@type": "dcat:Dataset",
```

DCAT-US 3.0 (draft)

```
{  
  "@context": "https://raw.githubusercontent.com/DOI-  
DO/dcat-us/main/context/dcat-us-3.0.jsonld",  
  "@id": "https://inventory.data.gov/dataset/9a9e946e-124e-  
46f5-a934-e458a6c1c2b2",  
  "@type": "dcat:Dataset",
```


Identifier

DCAT-US 1.1 (“POD”)

```
"identifier": "GSA-4495",
```

DCAT-US 3.0 (draft)

```
"identifier": "GSA-4495",  
"otherIdentifier": [  
  {  
    "@type": "adms:Identifier",  
    "notation": "GSA-4495",  
    "schemaAgency": "https://www.gsa.gov"  
  }  
],
```

Publisher

DCAT-US 1.1 (“POD”)

```
"publisher": {  
  "@type": "org:Organization",  
  "name": "General Services Administration"  
},
```

DCAT-US 3.0 (draft)

```
"publisher": {  
  "@id": "https://www.gsa.gov",  
  "@type": "org:Organization",  
  "name": "General Services Administration",  
  "notation": [  
    "023:0",  
    "47"  
  ]  
},
```

Distribution

DCAT-US 1.1 (“POD”)

```
"distribution": [  
  {  
    "@type": "dcat:Distribution",  
    "conformsTo": "http://www.isotc211.org/2005/gmd",  
    "downloadURL": "https://inventory.data.gov/dataset/9a9e946e-124e-46f5-a934-e458a6c1c2b2/resource/bfbd1f68-b7d8-4822-98bb-4e4ceabd2c3a7/download/iolp19139_080123.xml",  
    "mediaType": "application/xml",  
    "title": "IOLP19139_080123.xml"  
  },  
]
```

```
"distribution": [  
  {  
    "@id": "https://inventory.data.gov/dataset/9a9e946e-124e-46f5-a934-e458a6c1c2b2/resource/bfbd1f68-b7d8-4822-98bb-4e4ceabd2c3a7",  
    "@type": "dcat:Distribution",  
    "conformsTo": "http://www.isotc211.org/2005/gmd",  
    "downloadURL": "https://inventory.data.gov/dataset/9a9e946e-124e-46f5-a934-e458a6c1c2b2/resource/bfbd1f68-b7d8-4822-98bb-4e4ceabd2c3a7/download/iolp19139_080123.xml",  
    "mediaType": "application/xml",  
    "title": "IOLP19139_080123.xml",  
    "license": "https://creativecommons.org/publicdomain/zero/1.0/",  
    "accessRestriction": {  
      "@type": "dcat-us:AccessRestriction",  
      "restrictionStatus":  
        "https://data.resources.gov/vocab/nara/AccessRestrictionStatusAuthorityList/Unrestricted"  
    },  
    "useRestriction": {  
      "@type": "dcat-us:UseRestriction",  
      "restrictionStatus":  
        "https://data.resources.gov/vocab/nara/UseRestrictionStatusAuthorityList/Unrestricted"  
    }  
  },  
]
```

Theme

DCAT-US 1.1 (“POD”)

```
"theme": [  
  "geospatial"  
],
```

DCAT-US 3.0 (draft)

```
"theme": [  
  {  
    "@id": "https://data.gov/vocab/themes/geospatial",  
    "@type": "skos:Concept",  
    "prefLabel": "Geospatial",  
    "definition": "Geospatial data theme",  
    "inScheme": "https://data.gov/vocab/themes"  
  }  
],
```

Spatial

DCAT-US 1.1 (“POD”)

```
"spatial": "United States",
```

DCAT-US 3.0 (draft)

```
"spatial": [  
  {  
    "@id": "http://id.loc.gov/authorities/names/n78095330",  
    "@type": "dcterms:Location",  
    "prefLabel": "United States",  
    "notation": [  
      "US",  
      "USA"  
    ]  
  }  
],
```

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>