

# **Project Open Data Schema**

v1.1 Crosswalk with ISO 19115

# Crosswalk Developed

---

- Crosswalk is intended to be used as a resource in developing translation tools between the existing metadata standards
- Mapped POD v1.1 fields to ISO 19115
  - Separate effort worked on development of a crosswalk for CSDGM
- Focused on identifying universally applicable mappings
- Goal was to find a match for all mandatory fields and optional fields when available

# Catalog Fields

---

| Field       | Label               | Definition   | Required |
|-------------|---------------------|--|----------|
| @context    | Metadata Context    | URL or JSON object for the <a href="#">JSON-LD Context</a> that defines the schema used.                                 | No       |
| @id         | Metadata Catalog ID | IRI for the <a href="#">JSON-LD Node Identifier</a> of the Catalog. This should be the URL of the data.json file itself. | No       |
| @type       | Metadata Type       | IRI for the <a href="#">JSON-LD data type</a> . This should be <code>dcat:Catalog</code> for the Catalog.                | No       |
| conformsTo  | Schema Version      | URI that identifies the version of the Project Open Data schema being used.  | Always   |
| describedBy | Data Dictionary     | URL for the <a href="#">JSON Schema</a> file that defines the schema used.   | No       |
| dataset     | Dataset             | A container for the array of Dataset objects. See <a href="#">Dataset Fields</a> below for details.                      | Always   |

- This section describes the entire catalog file produced by an agency.
- Fields capture the version of the schema used and location of the agency catalog
- The dataset container field captures the records for each individual dataset in the catalog
- Anticipated that this section will be hard coded during the translation process

# Dataset Fields (Mandatory)

---

|             |             |  |        |
|-------------|-------------|--|--------|
| title       | Title       | Human-readable name of the asset. Should be in plain English and include sufficient detail to facilitate search and discovery.                 | Always |
| description | Description | Human-readable description (e.g., an abstract) with sufficient detail to enable a user to quickly understand whether the asset is of interest. | Always |
| keyword     | Tags        | Tags (or keywords) help users discover your dataset; please include terms that would be used by technical and non-technical users.             | Always |
| modified    | Last Update | Most recent date on which the dataset was changed, updated or modified.  | Always |

- 1:1 mapping between standards for these fields
- Given the inherent flexibility in the ISO 19115 standard to capture content the crosswalk captures all possible XPathS where the value can be stored.
  - /gmd:descriptiveKeywords/gmd:MD\_Keywords/gmd:keyword/gco:CharacterString | gmx:Anchor

# Dataset Fields (Mandatory)

---

When needed we have provided a series of choice options to populate fields in the POD json file -  
Example modified date

1. If the data are updated at a routine interval choose

```
//gmd:identificationInfo/gmd:MD_DataIdentification/gmd:resourceMaintenance/gmd:MD_MaintenanceInformation/gmd:maintenanceAndUpdateFrequency/gmd:userDefinedMaintenanceFrequency/gts:TM_PeriodDuration
```

2. Else Choose

```
//gmd:identificationInfo/gmd:MD_DataIdentification/gmd:citation/gmd:CI_Citation/gmd:date/gmd:CI_Date/gmd:date/gco:Date
```

```
WHEN gmd:CI_DateTypeCode == revision
```

3. ELSE pick first instance of:

```
/gmi:MI_Metadata/gmd:identificationInfo/gmd:MD_DataIdentification/gmd:citation/gmd:CI_Citation/gmd:date/gmd:CI_Date/gmd:date/gco:Date
```



# Dataset Fields (Mandatory)

---

|                              |                        |  |        |
|------------------------------|------------------------|--|--------|
| <a href="#">publisher</a>    | Publisher              | The publishing entity and optionally their parent organization(s).   | Always |
| <a href="#">contactPoint</a> | Contact Name and Email | Contact person's name and email for the asset.   | Always |
| <a href="#">identifier</a>   | Unique Identifier      | A unique identifier for the dataset or API as maintained within an Agency catalog or database.   | Always |
| <a href="#">accessLevel</a>  | Public Access Level    | The degree to which this dataset <b>could</b> be made publicly-available, <i>regardless of whether it has been made available</i> . Choices: public (Data asset is or could be made publicly available to all without restrictions), restricted public (Data asset is available under certain use restrictions), or non-public (Data asset is not available to members of the public). | Always |

- Publisher is a container for the publisher object
  - publisher - name
    - 1:1 mapping between standards  
CI\_ResponsibleParty when CI\_RoleCode="publisher"
    - Recommend using the Distributor if publisher is not populated in the ISO records
  - publisher - suborganizationOf
    - Optional field POD - No match in ISO 19115 as the standard does not support hierarchical organizational relationships



# Dataset Fields (Mandatory)

---

|              |                        |  |        |
|--------------|------------------------|--|--------|
| contactPoint | Contact Name and Email | Contact person's name and email for the asset. | Always |
|--------------|------------------------|--|--------|

- ContactPoint
  - Select
    - gmd:pointOfContact  
  
ELSE if not populated select
      - //gmd:MD\_DataIdentification/gmd:citation/gmd:CI\_Citation/gmd:citedResponsibleParty when CI\_RoleCode = pointofContact
  - ContactPoint - defined as the full name of the contact person
    - Recommend concatenating organisationName and individualName
    - As these fields are optional in ISO 19115 concatenating the two fields will ensure that some contact information is captured in the translation process
  - hasEmail
    - 1:1 mapping between standards



# Dataset Fields (Mandatory)

---

|                              |                        |  |        |
|------------------------------|------------------------|--|--------|
| <a href="#">publisher</a>    | Publisher              | The publishing entity and optionally their parent organization(s).   | Always |
| <a href="#">contactPoint</a> | Contact Name and Email | Contact person's name and email for the asset.   | Always |
| <a href="#">identifier</a>   | Unique Identifier      | A unique identifier for the dataset or API as maintained within an Agency catalog or database.   | Always |
| <a href="#">accessLevel</a>  | Public Access Level    | The degree to which this dataset <b>could</b> be made publicly-available, <i>regardless of whether it has been made available</i> . Choices: public (Data asset is or could be made publicly available to all without restrictions), restricted public (Data asset is available under certain use restrictions), or non-public (Data asset is not available to members of the public). | Always |

- identifier
  - MD\_Identifier - When it contains “DOI” in the text or associated attribute
  - gmd:fileIdentifier
  - Title when neither field is populated
    - //gmd:MD\_DataIdentification/gmd:citation/gmd:CI\_Citation/gmd:title/gco:CharacterString





# Dataset Fields (Mandatory)

---

|             |                     |  |        |
|-------------|---------------------|--|--------|
| accessLevel | Public Access Level | The degree to which this dataset <b>could</b> be made publicly-available, <i>regardless of whether it has been made available</i> . Choices: public (Data asset is or could be made publicly available to all without restrictions), restricted public (Data asset is available under certain use restrictions), or non-public (Data asset is not available to members of the public). | Always |
|-------------|---------------------|--|--------|

## access Level

1. Set POD json field value to "non-public"

```
//gmd:MD_LegalConstraints/gmd:accessConstraints/gmd:MD_RestrictionCode ==  
restricted OR
```

```
//gmd:MD_SecurityConstraints/gmd:classification/gmd:MD_ClassificationCode ==  
restricted, confidential, secret, or topSecret
```

1. ELSE set POD json field value to "restricted public"

```
//gmd:MD_LegalConstraints/gmd:accessConstraints/gmd:MD_RestrictionCode == copyright,  
patent, patentPending, trademark, license, or intellectualPropertyRights
```

1. ELSE set value of POD json field to "public" if these fields do not exist in the metadata

# Dataset Fields (Mandatory)

---

|  |              |   |        |
|--|--------------|---|--------|
| <code>bureauCode<sup>USG</sup></code>  | Bureau Code  | Federal agencies, combined agency and bureau code from OMB Circular A-11, Appendix C ( <a href="#">PDF</a> , <a href="#">CSV</a> ) in the format of <code>015:11</code> . | Always |
| <code>programCode<sup>USG</sup></code> | Program Code | Federal agencies, list the primary program related to this data asset, from the <a href="#">Federal Program Inventory</a> . Use the format of <code>015:001</code> .      | Always |

- Bureau and Program Codes are listed in OMB Circular A-11 and the Federal Program Inventory

Department of Commerce/National Oceanic and Atmospheric Administration/National Environmental Satellite, Data and Information Service

bureauCode - 006:48

programCode - 006:059

# Dataset Fields (Mandatory)

---

|                            |              |  |        |
|----------------------------|--------------|--|--------|
| bureauCode <sup>USG</sup>  | Bureau Code  | Federal agencies, combined agency and bureau code from OMB Circular A-11, Appendix C ( <a href="#">PDF</a> , <a href="#">CSV</a> ) in the format of 015:11 . | Always |
| programCode <sup>USG</sup> | Program Code | Federal agencies, list the primary program related to this data asset, from the <a href="#">Federal Program Inventory</a> . Use the format of 015:001 .      | Always |

- No direct mapping exists between the ISO 19115 standards to populate these values in the POD inventory file
- Recommendation that agencies insert these values into the POD json file as part of the translation/harvest process
- Alternate option is to include these codes as keywords pointing to the appropriate OMB thesaurus.
  - Given the volume of existing metadata records there were questions as to whether this was a viable cross-agency solution therefore not included in the crosswalk

# Dataset Fields (Conditional)

---

|                          |          |   |               |
|--------------------------|----------|---|---------------|
| <a href="#">license</a>  | License  | The license or non-license (i.e. Public Domain) status with which the dataset or API has been published. See <a href="#">Open Licenses</a> for more information.  | If-Applicable |
| <a href="#">rights</a>   | Rights   | This may include information regarding access or restrictions based on privacy, security, or other policies. This should also serve as an explanation for the selected "accessLevel" including instructions for how to access a restricted file, if applicable, or explanation for why a "non-public" or "restricted public" data asset is not "public," if applicable. Text, 255 characters. | If-Applicable |
| <a href="#">spatial</a>  | Spatial  | The range of spatial applicability of a dataset. Could include a spatial region like a bounding box or a named place.   | If-Applicable |
| <a href="#">temporal</a> | Temporal | The range of temporal applicability of a dataset (i.e., a start and end date of applicability for the data).  | If-Applicable |

- license - Not in ISO metadata
- rights -
  - `//gmd:MD_DataIdentification/gmd:resourceConstraints/gmd:MD_LegalConstraints/gmd:accessConstraints/gmd:MD_RestrictionCode`
  - `//gmd:MD_DataIdentification/gmd:resourceConstraints/gmd:MD_SecurityConstraints/gmd:classification/gmd:MD_ClassificationCode`

# Dataset Fields (Conditional)

---

|                          |          |   |               |
|--------------------------|----------|---|---------------|
| <a href="#">license</a>  | License  | The license or non-license (i.e. Public Domain) status with which the dataset or API has been published. See <a href="#">Open Licenses</a> for more information.  | If-Applicable |
| <a href="#">rights</a>   | Rights   | This may include information regarding access or restrictions based on privacy, security, or other policies. This should also serve as an explanation for the selected "accessLevel" including instructions for how to access a restricted file, if applicable, or explanation for why a "non-public" or "restricted public" data asset is not "public," if applicable. Text, 255 characters. | If-Applicable |
| <a href="#">spatial</a>  | Spatial  | The range of spatial applicability of a dataset. Could include a spatial region like a bounding box or a named place.   | If-Applicable |
| <a href="#">temporal</a> | Temporal | The range of temporal applicability of a dataset (i.e., a start and end date of applicability for the data).  | If-Applicable |

- spatial - Maps 1:1
- Temporal
  - Range of dates - maps 1:1
  - Single date - recommendation to use the one date as both start and end date
  - Multiple dates - choose earliest and latest date values
  - Open issues - No mapping for ongoing data collected at irregular intervals

# Dataset Fields (Distribution - Conditional)

---

|                                 |                      |  |               |
|---------------------------------|----------------------|--|---------------|
| <a href="#">accessURL</a>       | Access URL           | URL providing indirect access to a dataset, for example via API or a graphical interface.  | If-Applicable |
| <a href="#">conformsTo</a>      | Data Standard        | URI used to identify a standardized specification the distribution conforms to.  | No            |
| <a href="#">describedBy</a>     | Data Dictionary      | URL to the data dictionary for the distribution found at the <a href="#">downloadURL</a> . Note that documentation other than a data dictionary can be referenced using Related Documents as shown in the expanded fields. | No            |
| <a href="#">describedByType</a> | Data Dictionary Type | The machine-readable file format ( <a href="#">IANA Media Type</a> or <a href="#">MIME Type</a> ) of the distribution's <a href="#">describedBy</a> URL.   | No            |
| <a href="#">description</a>     | Description          | Human-readable description of the distribution.  | No            |
| <a href="#">downloadURL</a>     | Download URL         | URL providing direct access to a downloadable file of a dataset.   | If-Applicable |
| <a href="#">format</a>          | Format               | A human-readable description of the file format of a distribution.   | No            |
| <a href="#">mediaType</a>       | Media Type           | The machine-readable file format ( <a href="#">IANA Media Type</a> or <a href="#">MIME Type</a> ) of the distribution's <a href="#">downloadURL</a> .  | If-Applicable |
| <a href="#">title</a>           | Title                | Human-readable name of the distribution.   | No            |

- [downloadURL](#) - distribution URL when `CI_OnLineFunctionCode="download"`
- [mediaType](#) - No mapping to an ISO field given POD domain constraints

# Things to be aware of...

---

- If your agency has previously developed custom translation tools the definition of some fields has changed between the POD v1.0 and POD v1.1 schemas.
- View the POD changelog for more information - <https://project-open-data.cio.gov/metadata-changelog/>

# Crosswalk - More Information

---

- Project Open Data Schema version 1.1
  - <https://project-open-data.cio.gov/v1.1/schema/>
    - Page includes field definitions and domain restrictions
- Bureau and Program Codes
  - Bureau Codes -  
[http://www.whitehouse.gov/sites/default/files/omb/assets/a11\\_current\\_year/app\\_c.pdf](http://www.whitehouse.gov/sites/default/files/omb/assets/a11_current_year/app_c.pdf)
  - Program Codes -  
[http://goals.performance.gov/sites/default/files/images/FederalProgramInventory\\_FY13\\_MachineReadable\\_091613.xls](http://goals.performance.gov/sites/default/files/images/FederalProgramInventory_FY13_MachineReadable_091613.xls)
- Complete ISO 19115 to POD crosswalk is available through Google Drive
  - [https://docs.google.com/a/noaa.gov/spreadsheet/ccc?key=0AnITNk7dilPidEpXT2szS2RRSEdwR25GMm13SUdFdmc&usp=drive\\_web#gid=1](https://docs.google.com/a/noaa.gov/spreadsheet/ccc?key=0AnITNk7dilPidEpXT2szS2RRSEdwR25GMm13SUdFdmc&usp=drive_web#gid=1)
  - To request access follow instructions that pop-up when accessing the above link



**Questions?**