EPA Non-Geospatial Metadata Style Guide

Version 1.0, March 17, 2017

Document Change Control

The following is the document control for revisions to this document.

Version Number	Date of Issue	Author(s)	Brief Description of Change
V1.0	March 17, 2017	Steven Jett, Catherine Harness, Suzanne Pierson, Torrin Hultgren, Innovate!, Inc.	Initial Version.

Contents

Introduction

This Style Guide provides instructions on metadata preparation for EPA non-geospatial data resources based on EPA's Metadata Technical Specification guidelines and Project Open Data (POD) requirements. Non-geospatial resources are without inherently spatial qualities such as a latitude and longitude. Compliant metadata records are published to EPA's Environmental Dataset Gateway (EDG) and Data.gov to aid in data search and discovery.

EDG's non-geospatial metadata template contains all required Project Open Data (POD) metadata fields, usage notes for each field, an indication of whether the field is required or optional, and for certain fields the ability to select from a dropdown box of valid values. Please pay careful attention to this guidance – the metadata validation requirements are very strict and compliance is critical to ensure successful publication to EDG and Data.gov. The following sections provide an overview of EDG non-geospatial metadata procedures and detailed guidance for the Excel template.

For additional information about geospatial metadata tools and associated EPA metadata style guides, please refer to EPA's Metadata Style Guide for Geospatial and Non-Geospatial Data.

EDG Non-Geospatial Metadata Procedures - Overview

To acquire EDG's non-geospatial metadata template and begin creating metadata, please follow the steps listed below.

1. Contact the EDG Team and request EDG's non-geospatial metadata template.
2. After receiving the metadata template, complete all the required fields and add Information in the optional fields when applicable as described in the **"Required?"** row of the template.
3. After you have completed the spreadsheet, please email it to the EDG Team. We will convert the Excel fields into POD metadata and validate for POD compliance. The EDG team will contact you via email if validation errors occur.
4. Once complete and validated, a member of the EDG Team will approve your metadata record.

5. After your metadata record has been approved, it will be published and discoverable in the EDG. The EDG team will then email you information about how to access your records within the EDG. It may take up to 48 hours for the records or any changes or updates to be visible at Data.gov.

EDG Non-Geospatial Metadata Template Layout - Details

The EDG non-geospatial template contains two tabs, "DCAT" and "Values". The "DCAT" tab is used for completing required metadata fields. The "Values" tab contains data used to generate the pre-populated dropdown boxes such as, "Publisher", found in the DCAT tab. These values can be edited if desired within the "Values" tab.

DCAT Tab

	A	B	C	D
1	Field	Title	Description	Tags
2	Definition	Human-readable name of the asset. Should be in plain English and include sufficient detail to facilitate search and discovery.	Human-readable description (e.g., an abstract) with sufficient detail to enable a user to quickly understand whether the asset is of interest.	Tags (or keywords) help users discover your dataset; please include terms that would be used by technical and non-technical users.
3	Required?	Yes, always	Yes, always	Yes, always
4	Accepted Values	String	String	Array of strings
5	Usage Notes	Acronyms should be avoided.	This should be human-readable and understandable to an average person.	Separate keywords with commas.
6	Example for Reference	Distributed Structure Searchable Toxicity	The Distributed Structure Searchable Toxicity (DSSTox) online resource provides high quality chemical structures and annotations in association with toxicity data. It helps to build a	chemical safety, chemical safety research, chemicals, chemical testing, innovative chemical testing, computational toxicology, chemical screening, ToxCast

The EDG non-geospatial metadata template "DCAT" tab contains six pre-populated rows of information provided as guidance for producing EPA and Project Open Data (POD)-compliant metadata. Each row is described in detail below:

Row 1 – Field: contains a column for each standard metadata element that needs to be defined to meet Project Open Data requirements.

List of Fields to be Defined for a Data Resource

Field	Title	Description	Tags	Last Update	Publisher
	Contact Name	Contact Email	Unique Identifier	Public Access Level	Bureau Code
	Program Code Name	Program Code	License	Rights	Spatial
	Bbox	Temporal	Access URL	Access URL Title	Access URL Description
	Download URL	Download Media Type	Download Format	Download Title	Download Description
	Frequency	Data Standard	Data Quality	Data Dictionary	Data Dictionary Type
	Collection	Release Date	Language	Homepage URL	Primary IT Investment UII
	Related Documents	System of Records	Category		

Row 2 – Definition: covers specific guidance for each field to be completed. This row provides an explanation for the information to be submitted for each field. For example:

Field	Title
Definition	Human-readable name of the asset. Should be in plain English and include sufficient detail to facilitate search and discovery.
Required?	Yes, always
Accepted Values	String
Usage Notes	Acronyms should be avoided.
Example for Reference	Distributed Structure Searchable Toxicity

In this example, guidance is provided in the definition row on how to complete the column for the field "Title".

Row 3 – Required?: includes information on whether or not a specific field is required, required if applicable, or not required. If the field is required if applicable, there is detailed information describing when the field would be required versus not required. For example:

Field	Rights
Definition	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.
Required?	Yes, if accessLevel is "restricted public" or "non-public"
Accepted Values	String
Usage Notes	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.

In this example, detailed information is provided when the field "Rights" is required to be completed.

Row 4 – Accepted Values: offers detailed information on acceptable values each field. For example:

Field	Last Update
Definition	Most recent date on which the dataset was changed, updated or modified.
Required?	Yes, always
Accepted Values	ISO 8601 Date
Usage Notes	Dates should be ISO 8601 of least resolution. In other words, as much of YYYY-MM-DDThh:mm:ss.sTZD as is relevant to this dataset. If this file is brand-new, enter the issued date here as well.
Example for Reference	2013-12-17

In this example, the only accepted value for the field "Last Update" is an ISO 8601 Date.

Note: *pay very close attention to the Accepted values. If you input your field information in a format that is not accepted, your record will fail validation. For example, if you input your date information as 2/7/2017, your record will fail validation.*

Row 5 – Usage Notes: provides additional remarks on the information to be provided for each field, in many cases, elaborating on the information provided in the definition and accepted values rows often including tips for filling out the field correctly.

Field	Title
Definition	Human-readable name of the asset. Should be in plain English and include sufficient detail to facilitate search and discovery.
Required?	Yes, always
Accepted Values	String
Usage Notes	Acronyms should be avoided.
Example for Reference	Distributed Structure Searchable Toxicity

In this example, the Usage Notes provides additional information for the "Title" field explaining that Acronyms should be avoided when generating a "Title" for a data resource.

Row 6 – Example for Reference – provides a model to follow as an example for each field. The row in its entirety is a complete valid Project Open Data metadata record to be used as an example when completing the template for submission.

	A	B	C	D
1	Field	Title	Description	Tags
2	Definition	Human-readable name of the asset. Should be in plain English and include sufficient detail to facilitate search and discovery.	Human-readable description (e.g., an abstract) with sufficient detail to enable a user to quickly understand whether the asset is of interest.	Tags (or keywords) help users discover your dataset; please include terms that would be used by technical and non-technical users.
3	Required?	Yes, always	Yes, always	Yes, always
4	Accepted Values	String	String	Array of strings
5	Usage Notes	Acronyms should be avoided.	This should be human-readable and understandable to an average person.	Separate keywords with commas.
6	Example for Reference	Distributed Structure Searchable Toxicity	The Distributed Structure Searchable Toxicity (DSSTox) online resource provides high quality chemical structures and annotations in association with toxicity data. It helps to build a	chemical safety, chemical safety research, chemicals, chemical testing, innovative chemical testing, computational toxicology, chemical screening, ToxCast

Values Tab

	E	F	G	H	I
1					
2	**Program Code**	**Language**	**Category**	**Format**	**Access Level Comment**
3	020:001	en-US	farming	text/plain	EPA Category: Mission Sensitive, NARA (
4	020:002	en-CA	biota	text/csv	EPA Category: Drinking Water Vulnerabilit
5	020:003	es-MX	boundaries	text/html	EPA Category: Sensitive Drinking Water F
6	020:004	es-PR	climatology	application/	EPA Category: IT Security, NARA Catego
7	020:005	ch	economy	application/	EPA Category: Law Enforcement Sensitive
8	020:006	sm	elevation		EPA Category: Attorney Client Privilege, N
9	020:007		environment		EPA Category: Attorney Work Product, N/
10	020:008		geoscientificInformation		EPA Category: Deliberative Process Privile
11	020:009		health		EPA Category: Personally Identifiable Infor
12	020:010		imageryBaseMapsEart		EPA Category: Proprietary, NARA Catego
13	020:011		intelligenceMilitary		EPA Category: Confidential Business Infor
14	020:012		inlandWaters		EPA Category: Source Selection Informati
15	020:013		location		
16	020:014		oceans		
17	020:015		planningCadastre		
18	020:016		society		
19	020:017		structure		
20	020:018		transportation		
21	020:019		utilitiesCommunication		

The EDG non-geospatial metadata template "Values" tab contains prepopulated values used to produce the dropdown boxes found in the "DCAT" tab.

DCAT Tab:

Publisher	Co
The publishing entity.	Co
	na
Yes, always	Ye
String	Str
The plaintext name of the entity publishing this dataset.	-
U.S. Environmental Protection Agency	An

U.S. EPA Office of Water (OW)
U.S. EPA Office of Solid Waste and Emergency Response (OSWER) - Office o
U.S. EPA Office of the Administrator (OA) - Office of Policy (OP)
U.S. EPA Office of the Chief Financial Officer (OCFO)
U.S. EPA Office of Water (OW)
U.S. EPA Office of Water (OW) - Office of Ground Water and Drinking Wate
U.S. EPA Office of Water (OW) - Office of Science and Technology (OST)
U.S. EPA Office of Water (OW) - Office of Wastewater Management (OWM)
U.S. EPA Office of Water (OW) - Office of Wetlands Oceans & Watersheds (

Values Tab:

Publisher
Chesapeake Bay Program (CBP)
Great Lakes National Program Office (GLNPO)
U.S. Environmental Protection Agency
U.S. Environmental Protection Agency, Headquarters
U.S. Environmental Protection Agency, Region 1
U.S. Environmental Protection Agency, Region 10
U.S. Environmental Protection Agency, Region 2
U.S. Environmental Protection Agency, Region 3
U.S. Environmental Protection Agency, Region 4
U.S. Environmental Protection Agency, Region 5
U.S. Environmental Protection Agency, Region 6
U.S. Environmental Protection Agency, Region 7
U.S. Environmental Protection Agency, Region 8

The Publisher column in the "Values" tab populates the "Publisher" field dropdown box in the "DCAT" tab.