



EPA Metadata Crosswalk Implementation

Torrin Hultgren, Innovate!, Inc

hultgren.torrin@epa.gov



A compressed history

- Baseline Architecture
- The link between ISO and Project Open Data
- EPA Metadata Editors
- How we produce data.json
- Schema 1.0 compliance and pivot to Schema 1.1
- Path forward



Baseline Architecture

- Federated metadata management by a network of stewards
- Metadata produced using the EPA Metadata Editor
- Harvested into central metadata catalog (Esri GeoPortal Server) and from there to data.gov
- Geospatial community uses Esri stack and CSDGM
- Non-geo community used DMS and spreadsheets



DMS + POD = ISO?

- Previously implemented custom support for now-obsolete data.gov DMS format – strong motivation not to make the same mistake twice
- Initially GeoPortal could produce dcat.json, but not consume it
- CSDGM not a great fit for Project Open Data (POD) schema, ISO on our roadmap
- Conclusion – map POD to ISO, accelerate ISO adoption



EPA Metadata Editor 4.0

- EME 3.2 extremely successful metadata editor with both ArcCatalog integration and standalone functionality, but hardwired to Microsoft Access database and CSDGM schema
- ISO Support required near complete rewrite
- Development launched with goal of supporting fields necessary for data.gov compliance, full ISO editing implementation to follow

Description

[Title](#)

{***SUBJECT, GEOGRAPHIC EXTENT, RELEVANT TIME PERIOD, DATA OWNER/PROVIDER, OFFICE/REGION/RESEARCH LAB OF DATA OWNER***}

[Abstract](#)

This dataset contains the following layers: {***DEFINE LAYERS HERE***}. It is available as a {***Include any of the following as appropriate: webMapService, zipped download package, feature service, KML service, WMS service***}.

[Purpose](#)

{***TEXT DESCRIBING THE GENERAL PURPOSE OF THE DATASET***}

Keywords

ISO EPA User Place

- biota
- boundaries
- climatologyMeteorologyAtmosphere
- economy
- elevation
- environment**
- farming
- geoscientificInformation
- health
- imageryBaseMapsEarthCover
- inlandWaters

[Help](#)

D

Bounding Box

[Description](#) USA D

N 72 E -66

S 18 W -180

Time Periods

Temporal Extent of Resource

Time Period

Description

Begin

End

Interval

Dates for Resource

[Creation](#)

[Publication](#)

[Revision](#)

[Update Frequency](#)
asNeeded

Distribution 1 of 1 + x

- Contact (Required) 1 of 1 + x
- Digital Transfer Options < 1 of 2 > + x
- Standard Order Process 0 of 0 + x
- Distribution Format 0 of 0 + x

Resource Constraints

System of Records [dropdown]

Legal Constraints

Use Limitation Please check sources, scale, accuracy, currentness

Access Constraints [otherRestrictions]

Use Constraints [otherRestrictions]

Other Constraints None

Security Constraints

Use Limitation/ Access Level public

Classification unclassified

User Note / Comment [dropdown]

Classification System FIPS Pub 199

Handling Description Standard Technical Controls

Contacts

Resource Owner 1 of 1 + x

Role owner

Individual David Parrish

Organization U.S. Environmental Protection Agency

Position EDG Project Manager

Select Contact [dropdown] [D]

Program Code IT / Data Manager

Contact Details [input]

Resource Producer 1 of 1 + x

Metadata Author 1 of 1 + x

Metadata Information

Resource Type dataset

Last Modified [input] x !

Language eng; USA

Identifier [input] New Identifier

Metadata Standard

Standard Name ISO 19115 Geographic information - Metadata

Standard Version ISO 19115:2003(F)



Spreadsheet Editor

- Authored Excel spreadsheet with POD guidance, valid values as dropdowns and example records
- On receipt of spreadsheet, save as CSV, run Python script to map values to ISO xml template – harvest as ISO
- Great for batch editing, clunky in most other ways
- Map to ISO functional, but authored in isolation, and many fields required creative interpretation of ISO



Geospatial Reckoning

- Summer 2014: beta tests of EME 4.0 favorable, but geospatial community refused to convert records to ISO until Data Quality and Feature Catalog sections could be part of output (ISO 19115-3, Spring 2015)
- Meant November 30th 2014 deadline of full POD schema 1.0 compliance had to be met with CSDGM records
- Produced EME 3.2.1 for **minimal** POD compliance with CSDGM records

File	Edit	Tools	Help
Basic Data Set Information Quality, Coordinate System, and Attribute Information Distribution & Metadata Information			
Citation			
* <u>Origin:</u>	US Environmental Protection Agency	<input type="button" value="D"/>	
* <u>Title:</u>	USEPA Geospatial Metadata Sample Title	<input type="button" value="D"/>	
Publisher			
* <u>Published by:</u>	U.S. EPA Office of Environmental Information (OEI)	<input type="button" value="D"/>	
* <u>Published at:</u>	Washington, DC	* <u>Date:</u>	2007 <input type="button" value="today"/>
Online Linkage <input type="button" value="+"/>			
** <u>Primary Linkage:</u>	https://edg.epa.gov	<input type="button" value="D"/>	<input checked="" type="checkbox"/>
** <u>Secondary Linkage:</u>	http://www.epa.gov/geospatial/	<input type="button" value="D"/>	<input checked="" type="checkbox"/>
Description			
* <u>Abstract:</u>	<input type="text"/>		
* <u>Purpose:</u>	<input type="text"/>		
<u>Supplemental Info:</u> <input type="text"/>			
Time Period <input type="button" value="D"/>			
* Date of Data Set	<input type="button" value="D"/>		
Single Date OR Multi Dates: Date1, Date2,	* <u>Progress of data:</u>	Planned	<input type="button" value="D"/>
OR Range of dates: Date1 - Date2	* <u>Data currency:</u>	Publication date	<input type="button" value="D"/>
<input type="text"/>	* <u>Update frequency:</u>	monthly	<input type="button" value="D"/>
		<input type="button" value="today"/>	
Bounding Box			
USA <input type="button" value="D"/>			
* <u>N:</u>	61.7110157	* <u>E:</u>	-12.68151645
* <u>S:</u>	6.65223303	* <u>W:</u>	-138.21454852
Keywords			
ISO	EPA	User	Place
biota boundaries climatologyMeteorologyAtmosphere economy elevation environment farming geoscientificInformation health imageryBaseMapsEarthCover inlandWaters intelligenceMilitary			
			<input type="button" value="Help"/>
			<input type="button" value="D"/>
Data Set Constraints			
* <u>Access:</u>	None	<input type="button" value="D"/>	
* <u>Use:</u>	None. Please check sources, scale, acc.	<input type="button" value="D"/>	
* <u>Security Classification:</u>	public	<input type="button" value="D"/>	
Contact			
<input type="radio"/> Primary Person		<input checked="" type="radio"/> Primary Organization	
U.S. Environmental Protection Agency, Region 2, G			<input type="button" value="D"/>
YELLOW * mandatory GREEN ** mandatory if applicable BLUE optional Click on text to link to element description			
<input type="button" value="Save"/>		<input type="button" value="Save & Close"/>	<input type="button" value="Cancel"/>



Esri GeoPortal and POD

- Permits flexible mapping of elements from standard metadata formats (CSDGM, ISO, Dublin Core) to DCAT output fields using Xpath
- Full output generated and cached per schedule, also available via API for custom queries
- Elements can be hardcoded or given default values
- No ability to translate between different domains, cascade through multiple elements, or suppress invalid values

- After frenzy of activity, the deadline was met, but the landscape was already different
- Schema 1.1 released November 1st, requiring a new approach
- GeoPortal now supports harvesting of POD records
- Mid-November summit of FGDC and GSA to begin ISO crosswalk standardization revealed many different approaches and perspectives worth consideration



Present Day

- Participation in FGDC ISO Metadata Crosswalk development – clear need for consistency
- Developed new spreadsheet for POD 1.1 – but strongly considering handling non-geo records as dcat.json rather than ISO xml
- Received and are busy staging update to GeoPortal server to support POD 1.1
- No new mandatory fields means we will meet Feb. 1st deadline with existing CSDGM records



Harvesting Aside

- We currently harvest all our records (aside from two NGDA records) to data.gov via data.json
- CSW harvest fails because CKAN can only handle one metadata format per CSW endpoint
- WAF harvest fails because our WAF is dynamic and URLs contain “?”
- We’ve recently addressed these issues and are considering switching to harvesting geo records in native formats, but wary of validation issues



Path forward for ISO

- EPA remains committed to ISO implementation
- Strong user support for continued EME development, but significant work remains
- Must seriously consider adopting ISO profile in ArcCatalog
- FGDC engagement with Data.gov team is critical, need to ensure the expertise of the federal geospatial metadata community included in Data.gov decisions and aligned with ISO implementation



Acknowledgements

- Fearless leaders:
 - David Parrish and Ana Greene, EPA
- EDG team:
 - Suzanne Pierson, Catherine Harness, Baohong Ji, Steven Jett, Dan Spinosa, and Mettja Kuna, Innovate!, Inc.