Publishing your metadata via Data.gov

John Jediny
Chief Engineer
Program Staff

Hyon Kim
Program Director
Program management, Policy
Coordination with agencies, OMB, OSTP
Public feedback

John Jediny
Chief Engineer
Leading infrastructure migration and security accreditation of Data.gov
Will work with CTO on infrastructure issues for TTS

Philip Ashlock
Chief Architect
Acting Lead, Portfolio
Technical lead
Portfolio responsibilities
Additional data initiatives
- U.S. Data Federation
- Sustainable Development Goals (SDG)
- Open Government Partnership (OGP)

Contractors:
- 5 in application development support (REI)
- 5 in cloud support services (Aqulent)
- 1 for program management support (Cascades)
Many organisations already have their data in repositories with well-defined processes and procedures for publishing and managing data. In this case, the data can be simply pulled regularly into CKAN from the existing repositories. **To facilitate this model, we’ve developed a sophisticated and customisable “harvesting” mechanisms which can fetch and import records from many different repository sources**, some examples include:

- Web Catalog Service (CSW) Servers (ex. Geonetwork, pyCSW, Geoportal)
- Customized Content Management Systems (ex. DKAN)
- Web Accessible Folders
- ArcGIS Open Data Portal*
- Geonode
- Socrata
Data.json is derived from W3C DCAT
https://project-open-data.cio.gov
https://project-open-data.cio.gov/v1.1/metadata-resources/#crosswalks-for-geospatial-metadata
Data.gov Harvest Sources

Data.gov harvests metadata from **920+ sources**

Data.gov Harvest Sources, Validation Schema, CKAN Mapping and Implementation Details

**List of Harvest Sources:**
- `/data.json`
- CSW Server
- Web Accessible Folder (WAF)
- Single spatial metadata document
- Geoportal Server
- Web Accessible Folder (WAF) Homogeneous Collection
- Z39.50:
- ArcGIS REST API

**Validation Types and References:**
- Project Open Data (Federal)
- Project Open Data (Non-Federal)
- Autodetect
- ISO 19115 Metadata (ISO 19139 XSD)
- FGDC CSDGM Minimal Validation
- FGDC CSDGM Biological Data Profile (FGDC-STD-001.1-1999)
- FGDC CSDGM Metadata Profile for Shoreline Data (FGDC-STD-001.2-2001)
Data.gov has partnered with the FGDC to develop new DRAFT Guidance on Metadata Interoperability Best Practices

*To Be Released for Comment Shortly*

- Validation Approach of Source Metadata
- FGDC to ISO 19115 Conversion/Transformation
- NOAA XSLT that is being used for Catalog.data.gov transformation
- ISO 19115 metadata fields to CKAN dataset fields mapping
- Data.json (Project Open Data Schema) metadata fields to CKAN dataset fields mapping
- ArcGIS json metadata record fields to CKAN dataset fields mapping
- CKAN data.json/ArcGIS json metadata record fields to CSW Service ISO Metadata field mapping
- Table 1 (Harvest Source vs Validation Schema crosswalk

<table>
<thead>
<tr>
<th>Harvest Source</th>
<th>data.json</th>
<th>CSW Server</th>
<th>Web Accessible Folder (WAF)</th>
<th>Single spatial metadata document</th>
<th>Geoportal Server</th>
<th>Web Accessible Folder (WAF) Homogeneous Collection</th>
<th>Z39.50</th>
<th>ArcGIS REST API</th>
</tr>
</thead>
<tbody>
<tr>
<td>FGDC Metadata Validation Schema</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FGDC Metadata Validation Schema</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ISO 19115 Metadata (ISO_19115.XSD)</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ISO 19115 Metadata Validation</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FGDC CSDGM Biological Data Profile (FGDC-STD-O 02.1-2003)</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FGDC CSDGM Metasearch Profile for Shoreline Data (FGDC-STD-O 03.2-2003)</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FGDC CSDGM Metasearch Profile for Shoreline Data (FGDC-STD-O 03.2-2003)</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Work in Progress:
Mapping data.json (GeoDCAT) to Geoplatform 19115-2 (NAP & 19115-3)
Other Needs:

.xsd to JSON Schema for metadata validation and transformation
Metadata Transformation-as-a-Service (i.e. auto-updating legacy records)
Adapt XML based metadata repositories to JSON and YAML
Single dataset entry YAML metadata

Questions?