Open Source Metadata Toolkit

May 14, 2014
USGS Community for Data Integration

Stan Smith, USGS
Josh Bradley, Arctic LCC

Alaska Data Integration working group
Mission: “The Alaska Data Integration Working Group (ADIwg) was formed to examine and address the technical barriers to efficiently integrate and share data within and among participating organizations.”

- Federal
  - DOI: BLM, BOEM, NPS, USFS, USFWS, USGS
- State of Alaska
  - University of Alaska (UAF, UAS)
  - Geographic Information Network of Alaska (GINA)
  - International Arctic Research Center (IARC)
- Non-Governmental Organizations (NGOs)
  - Arctic Ocean Observing System (AOOS)
  - Arctic Research Mapping Application (ARMAP) - Nunatech Consulting
  - North Pacific Research Board (NPRB)
  - North Slope Science Initiative (NSSI)
- Cooperatives/Joint-Ventures
  - Arctic LCC

www.adiwg.org
Project Background

• Project originally prompted by ADIwg’s adoption of ISO
  o Develop data metadata standard
  o Transition existing project standard from FGDC

• Work accelerated by grants
  o FY13 funds from NCCWSC accelerated work on mdTranslator prototype
  o FY14 funds from CDI
Adopting ISO

• Pros
  o Consistency with our international partners
  o Develop for the future – ISO is planned replacement for CSDGM
  o Shared implementation brings consistency to ADIwg organizations
  o NOAA is contracted to provide world-wide support for ISO (XSDs)

• Cons
  o Very complex, strict rule set for writing metadata
  o Few existing tools
  o Few US adopters
  o Significant investment in learning and tool development
ADIwg Supported Standards

Standards Setting Organizations

FGDC CSDGM
(US standard)

ISO 19115-2
(int’l standard)

ISO 19115-1
(int’l standard)

ADIwg

Project FGDC
(published/implemented)

Project ISO 19115-2
(written/documenting)

Data ISO 19115-2
(written/documenting)

Data ISO 19115-1

Project ISO 19115-1

subset

subset

subset

subset

map

map

subset

map

www.adiwg.org
Project Objectives

• Support creation of original metadata records
• Portable, open source code
• Create a robust, more user-friendly metadata standard
• Accommodate varied requirements and technical abilities
• Extensible: Read once, write many
Project Objectives (ISO)

- Eliminate necessity to learn ISO 19115/19139 standards
- Make it easier for organizations to:
  - more easily achieve ISO compliance
  - Integrate ISO metadata support into local systems
  - Implement custom web services with ISO metadata capability
  - Create or populate local or remotely hosted databases of metadata records
What the Toolkit Doesn’t Do (Yet)

• Possible, but implementation TBD
  - Translation from one XML format to another, e.g. FGDC to ISO
  - Editing of existing metadata records

• The toolkit does not provide metadata clearinghouse or catalog services.
Main Toolkit Components

- ADIwg mdJSON
- mdTranslator
- mdEditor
• Developed as an alternative metadata exchange standard
• Single schema supports projects and data
• Spatial objects supported as GeoJSON
• Can be easily extended
ADIwg mdJSON Schema

- Documents JSON syntax - data types and structure
- Available as Ruby Gem 🍃
- Can be used independent of mdTranslator
- Using latest IETF draft: version 4
- Validation engines available in many languages, see [http://json-schema.org/](http://json-schema.org/)

https://github.com/adiwg/adiwg-json-schemas
• ADIwg supported fields
• 100+ classes
• 300+ attributes
• 60% of full standard
• 3 ‘E’ size diagrams
• Example of OO class model detail.
• The class model can be used to understand/create ISO compliant metadata using the attribute names, order, class type, optionality, cardinality, and domains referenced in this OO class model.
{A} \textbf{JSON vs. ISO XML Example}

\begin{verbatim}
"address": {
  "deliveryPoint": ["101 12th Avenue", "Room 110"],
  "city": "Fairbanks",
  "administrativeArea": "AK",
  "postalCode": "99701",
  "country": "USA",
  "electronicMailAddress": ["joshua_bradley@fws.gov", "jbradley@arcticlcc.org"]
}
\end{verbatim}

\begin{verbatim}
<gmd:address>
  <gmd:CI_Address>
    <gmd:deliveryPoint>
      <gco:CharacterString>101 12th Avenue</gco:CharacterString>
    </gmd:deliveryPoint>
    <gmd:deliveryPoint>
      <gco:CharacterString>Room 110</gco:CharacterString>
    </gmd:deliveryPoint>
    <gmd:city>
      <gco:CharacterString>Fairbanks</gco:CharacterString>
    </gmd:city>
    <gmd:administrativeArea>
      <gco:CharacterString>AK</gco:CharacterString>
    </gmd:administrativeArea>
    <gmd:postalCode>
      <gco:CharacterString>99701</gco:CharacterString>
    </gmd:postalCode>
    <gmd:country>
      <gco:CharacterString>USA</gco:CharacterString>
    </gmd:country>
    <gmd:electronicMailAddress>
      <gco:CharacterString>joshua_bradley@fws.gov</gco:CharacterString>
    </gmd:electronicMailAddress>
    <gmd:electronicMailAddress>
      <gco:CharacterString>jbradley@arcticlcc.org</gco:CharacterString>
    </gmd:electronicMailAddress>
  </gmd:CI_Address>
</gmd:address>
\end{verbatim}

www.adiwg.org
"resourceSpecificUsage": [
  {
    "specificUsage": "Developed for land use planning.",
    "userDeterminedLimitations": "Not suitable for use for navigation.",
    "userContactInfo": [
      {
        "contactId": "3",
        "role": "publisher"
      },
      {
        "contactId": "3",
        "role": "author"
      }
    ]
  },
  {
    "specificUsage": "Bathymetry datasets developed for modeling physical oceanographic conditions in the Barrow Straight.",
    "userContactInfo": [
      {
        "contactId": "1",
        "role": "custodian"
      }
    ]
  }
],
## JSON vs. ISO XML Savings

<table>
<thead>
<tr>
<th></th>
<th>Data Char.</th>
<th>Syntax Char.</th>
<th>Lines</th>
<th>% Syntax</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Early Example</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JSON</td>
<td>1682</td>
<td>2029</td>
<td>127</td>
<td>54.7%</td>
</tr>
<tr>
<td>ISO</td>
<td>1682</td>
<td>20110</td>
<td>403</td>
<td>92.3%</td>
</tr>
</tbody>
</table>

| ISO Expansion     | 9.91       | 3.17         |

| **Final Example** |           |              |       |          |
| ISO Expansion     | 1549       | 3234         | 2.09  |          |

[www.adiwg.org](http://www.adiwg.org)
• ADIwg examples show every supported field and option.
• ISO 19139 structure is precise, but verbose.
• About 4 times expansion over FGDC metadata.
• Partial ISO 19115-2 metadata record (right).
• Data dictionary must be in separate ISO 19110 file.
• Example ISO 19110 averaged 130 lines of metadata per attribute.
ADIwg mdTranslator

- Primary purpose is to transform ADIwg mdJSON into other metadata formats
- Initial target is ISO 19115-2
- "Loosely coupled" architecture allows components to be developed independently
- Will be available as Ruby gem, Rails optional
- Command-line interface

http://www.adiwg.org/mdTranslator/
What you need to know to use mdTranslator

• ADIwg JSON schema
• Metadata content
• ISO code list domains

What you DON’T need to know to use mdTranslator

www.adiwg.org
mdTranslator Architecture

ADIwg mdTranslator

- JSON Reader
- ISO XSL Translators
  - Dublin
  - FGDC
  - Other
- Internal Format
- ISO Metadata Writer
- Metadata Reader
- ISO
- Other

ADIwg JSON

Output (ISO XML)
Hosting

- Hosts the demo
- Updated via Git push
- Probably will host metadata tool here
- No cost for basic tier

For links to demo and repos visit:
www.adiwg.org/projects

- Code repositories
  - mdTranslator
  - mdTranslator-rails
  - adiwg-json-schemas

- Websites
  - GitHub Pages
  - Jekyll publishing
• On-line editor for ADIwg mdJSON
• JavaScript web-client with minimal server-side requirements
  o HTML5: Local Storage & File API
  o Optional persistence via webservice
  o Basic configuration via JSON
• Use stand-alone or embedded
• Interfaces with mdTranslator
**Toolkit in Practice**

1. Hosted Web Service
   
   Demo: [http://mdtranslator.herokuapp.com/iso19115](http://mdtranslator.herokuapp.com/iso19115)
2. Integrated System Component

- Database
- ADIwg JSON
- Agency System
- mdTranslator
  - JSON Reader
  - Metadata Writer
  - ISO Reader
  - FGDC Reader
  - Other Reader
  - ISO Writer
  - FGDC Writer
  - Other Writer
- Ruby Gem
- Agency Server
- ISO XML
3. Translator as: Hosted Metadata Editor/Translator
Development Milestones

- Support ADIwg Project Metadata standard
- Refactor and Package mdTranslator as Ruby Gem
- Complete Developer and User Documentation
- Open REST endpoint to public
- Provide web form access and code to public
- Support ISO 19115-1

Milestones Reached

- Build Reader and writer components (JSON -> ISO XML)
- Build JSON validator components
- Draft ISO JSON Standard
- Design Translator Architecture
- Verify Supported Fields List support for other standards
- Build supported ISO XML
- Determine Supported ISO Classes & Attributes
- Host Demo for mdTranslator

CDI funding to cover these milestones – 2014/2015

Funded by NCCWSC 2013

www.adiwg.org
Interested Parties

- SAON – Sustaining Arctic Observing Networks
- ADCN – Arctic Data Coordination Network
- PGC – Polar Geospatial Center
- HDF Group
- ACADIS – Advanced Cooperative Arctic Data and Information Services
- IARC – International Arctic Research Center
- BAID – Barrow Area Information Database
- IASOA – International Arctic Systems for Observing the Atmosphere
- IARPC – Interagency Arctic Research Policy Council
  - Terrestrial Ecosystems Implementation Team (TEIT)
  - Arctic Data Implementation Team (ADIT)
- SEARCH – Study of Environmental Arctic Change
- NCCWSC – National Climate Change & Wildlife Science Centers
Questions?

Discussion?