Big Earth Data Initiative (BEDI)
Metadata Improvement: Case Studies

John Kozimor, Ted Habermann and John Farley

This work was supported by NASA/GSFC under Raytheon Co. contract number NNG10HP02C
Big Earth Data Initiative (BEDI)

The Big Earth Data Initiative (BEDI) invests in standardizing and optimizing the collection, management and delivery of U.S. Government’s civil Earth observation data to improve discovery, access use, and understanding of Earth observations by the broader user community.

Complete and consistent standard metadata helps address all three goals.

Many agencies included.
# Projects/Communities

<table>
<thead>
<tr>
<th>Projects</th>
<th>Big Earth Data Initiative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communities</td>
<td>NASA, NOAA, USGS, USDA, EPA</td>
</tr>
<tr>
<td>Dialects</td>
<td>DIF, ECHO, ISO, FGDC, EML</td>
</tr>
</tbody>
</table>
Overview

• Tools and techniques has been developed to:
  – Compare recommendations and dialects
  – Identify the structure of metadata collections
  – Compare the structure of metadata collections
  – Evaluate and measure metadata completeness with respect to recommendations
  – Evaluate and measure metadata completeness with respect to specific organization goals

• These tools and techniques were applied to NASA metadata collections.

• The results of this analysis identify how well CMR metadata collections comply with UMM-Collection Profile.
Case Study 1: Do my dialects support my requirements?

Overview: Metadata recommendations change as new communities and needs emerge. Metadata management tools are driven by dialects. Changing those tools and training people are difficult, so adoption of new dialects is relatively slow.

Purpose: Identify gaps between existing organizational capabilities (dialects) and new recommendations (requirements).

Scope:
• ISO 19115, ECHO, DIF10 and DIF dialect support with respect to the UMM-Collection Profile
Case Study 1: Results
UMM-Collection Recommendation / NASA Dialects

Maximum Possible Score for Dialects and Recommendation

Number of Concepts

Required  Recommended  Optional

DIF10  
ECHO10  
ISO 19115  
--- Recommendation

GAP
Case Study 2: How complete is CMR metadata with respect to the UMM-Collection Profile

**Overview:** The UMM-Collection Profile specifies metadata concepts (required, recommended and optional) for use with CMR collection metadata. Required concepts are mandatory in CMR metadata.

**Purpose:** Evaluate and measure the completeness of NASA metadata collections with respect to the UMM-Collection profile.

**Scope:**
- 4,000 ISO, ECHO, DIF10 and DIF records from 17 CMR metadata collections
Case Study 2a: Results
UMM-C Required

Required UMM-Collection Concepts Missing in CMR Metadata (ECHO and DIF10)

- Project Name
- Instrument Short Name
- Platform Short Name
- SpatialExtent
- Related URL
- Keyword
- Processing Level
- Metadata Dates

Legend:
- ECHO10
- DIF10

Percent Missing
Case Study 2b: Results
UMM-C Recommended

Recommended UMM-Collection Concepts Missing in CMR Metadata (ECHO and DIF10)

<table>
<thead>
<tr>
<th>UMM-Collection Recommended Concepts</th>
<th>ECHO</th>
<th>DIF10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resource Cost or Fees</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transfer Size</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resource Format</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Media</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Topic Category</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resource Access Constraints</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resource Use Constraints</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quality Statement</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resource Status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resource Citation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Collection Data Type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Responsibility</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resource Contact</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resource Language</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Purpose</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Case Study 2c: Results
UMM-C Optional

Optional UMM-Collection Concepts Missing in CMR Metadata
(ECHO and DIF10)

- Sensor Short Name
- TwoDCoordinateSystem
- Paleo Temporal Coverage
- Additional Attributes
- Ancillary Keyword
- Publication Information
- Metadata Association

Percent Missing

ECHO10
DIF10
Case Study 3: What metadata records require updating?

**Overview:** Metadata completeness (signature) scores can be used to identify groups of metadata records that are often missing the same concepts.

**Purpose:** Utilize signature scores to identify and prioritize metadata improvement work.

**Scope:** 16,000 CMR metadata records.
Case Study 3: Results
DIF Signatures Groups

<table>
<thead>
<tr>
<th>Resource</th>
<th>Resource Id</th>
<th>Version</th>
<th>Keyword</th>
<th>Progress</th>
</tr>
</thead>
<tbody>
<tr>
<td>LAADS</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>78</td>
</tr>
<tr>
<td>LANCEMODIS</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>76</td>
</tr>
<tr>
<td>LARC</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>186</td>
</tr>
<tr>
<td>GSFC4PA</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>148</td>
</tr>
<tr>
<td>GESDISC</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>7</td>
</tr>
</tbody>
</table>
Acknowledgements
This work was supported by NASA/GSFC under Raytheon Co. contract number NNG10HP02C