Metadata Best Practices Illustrated with the CEDAR Workbench

November 2018

John Graybeal

jgraybeal@stanford.edu
What is the CEDAR Workbench?
1. A Template Designer to create forms.

2. A Metadata Editor to fill out those forms.

3. A Resource Manager to manage the forms and metadata.
With semantic services (vocabularies) from BioPortal

And APIs to access metadata remotely or submit them to external repositories
Our Best Practices

1. Quickly target your team’s metadata standards
2. Make metadata entries consistent and accurate
3. Enter and verify metadata as quickly and easily as possible
4. Drive search with well-defined vocabularies and mappings
BP1: Quickly target your team’s metadata standards

• **GIVEN:** A defined standard, community practice, or external requirement for metadata content
• **GOAL:** Quickly set up a web service that lets teams enter and verify metadata that meets the specification
• **EXAMPLE:** Minimal metadata that must meet project requirements and be submitted to an external repository
• **APPROACH:** Define a metadata form satisfying your metadata content using CEDAR’s Template Builder.
• **EXTRA BENEFITS:**
  • A sharable computable specification in JSON Schema
  • Support for manual and automated metadata entry.
BP1: Target your metadata standards

Find a similar template, or create your own

---

### Workspace

- **Shared with Me**

  - ![Folder](image-url) **Drone Minimum Information**
    - **Date Created**: 7/18/18
    - **Date Modified**: 9/4/18

---

**All / Users / Sarah Swanz / ESIP Drone Minimum Metadata**

<table>
<thead>
<tr>
<th>Title</th>
<th>Created</th>
<th>Modified</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drone Minimum Information</td>
<td>7/18/18</td>
<td>9/4/18</td>
</tr>
<tr>
<td>Drone Project Information</td>
<td>7/18/18</td>
<td>8/30/18</td>
</tr>
<tr>
<td>Drone Minimum Information metadata</td>
<td>7/20/18</td>
<td>7/20/18</td>
</tr>
<tr>
<td>1. Introduction to CEDAR Talk</td>
<td>7/19/18</td>
<td>7/19/18</td>
</tr>
<tr>
<td>Drone Platform Information</td>
<td>7/18/18</td>
<td>7/19/18</td>
</tr>
<tr>
<td>Drone Site Information</td>
<td>7/18/18</td>
<td>7/18/18</td>
</tr>
<tr>
<td>Drone Fix Information</td>
<td>7/18/18</td>
<td>7/18/18</td>
</tr>
<tr>
<td>Drone Flight Information</td>
<td>7/18/18</td>
<td>7/18/18</td>
</tr>
<tr>
<td>Drone Sensor Information</td>
<td>7/18/18</td>
<td>7/18/18</td>
</tr>
</tbody>
</table>
BP1: Target your metadata standards

Create fields and elements that match your needs
BP1: Target your metadata standards

Include fields and elements in your template

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drone Minimum Information</td>
<td>Information needed to describe all aspects of drone data</td>
</tr>
</tbody>
</table>

**Project Information**
- Project Name
- Project Funder
- Research Question
- Project Investigator

**Drone Platform Information**
- Platform Type
BP1: Target your metadata standards

Start collecting your metadata

### Drone Minimum Information

**Project Information**

- **Project Name**: AgroCapture
- **Project Funder**: Division of basic sciences
- **Research Question**: How can remote sensing information be used to rotate crops and livestock at optimal times?
- **Project Investigator**: Julisco Jeffers

**Drone Platform Information**

<table>
<thead>
<tr>
<th>Platform Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>TestTetheredConstraintPlatforms</td>
</tr>
<tr>
<td>TestWithCrewPlatform</td>
</tr>
<tr>
<td>TetheredBalloon</td>
</tr>
<tr>
<td>Towfish</td>
</tr>
<tr>
<td>Unmanned Aerial Vehicle (Drone)</td>
</tr>
<tr>
<td>VerticalProfiler</td>
</tr>
</tbody>
</table>
BP2: Make metadata entries consistent and accurate

- **GIVEN**: Complicated field values that must be exactly right
- **GOAL**: Get the metadata entered correctly
- **EXAMPLE**: Data product descriptions using complex terms
- **APPROACH**: Semantic Terms from Controlled Vocabularies, Auto-completion, Field Validation, Field Tips
- **EXTRA BENEFITS**:
  - Interoperability with semantic web (JSON-LD or RDF)
  - Early confirmation of many typographic errors
  - Less experienced users more confident in their ability to enter good metadata, and more motivated to do so.
**BP3: Enter & verify metadata as quickly and easily as possible**

- **Given**: Many assets requiring a lot of metadata entries
- **Goal**: Enter metadata quickly with minimal pain
- **Example**: Describe 40 similar files using complex values
- **Approach**: Ordered Controlled Terms, Suggestions, and an ‘Instance example’ (with Hidden fields)
- **Extra Benefits**:
  - Obtain benefit from earlier work by other contributors
  - Ability to blend automated, manual metadata entry
  - Can include ‘provenance fields’ in each filled-out form
BP4: Drive search with well-defined vocabularies and mappings

• GIVEN: Metadata from varied sources using terms relatable to other terms (e.g., synonym or parent/child relations)
• GOAL: To find all applicable matches across term sets
• EXAMPLE: Data in Google’s Data Search is indexed with terms from GCMD, CF, and SWEET. Find all data that includes air temperature.
• APPROACH: Make sure that keyword and parameter description fields in CEDAR templates require selection from well-known controlled vocabularies (or vocabularies mapped to them).
• EXTRA BENEFITS:
  • Meaning of concepts in metadata sources is defined.
  • Value of vocabularies enhanced by usage and mappings.
CEDAR References

• Sign up and use CEDAR: https://cedar.metadatacenter.org

• Learn about CEDAR: https://metadatacenter.org

• More CEDAR references: https://metadatacenter.org/refs

• On GitHub (and social media) at metadatacenter